Distinct subfamilies of primate L1Gg retroposons, with some elements carrying tandem repeats in the 5' region

Joyce A.Lloyd⁺ and S.Steven Potter*

Children's Hospital Research Foundation, Elland and Bethesda Avenues, Cincinnati, OH 45229, USA

Received April 8, 1988; Revised and Accepted June 6, 1988

ABSTRACT

Two subfamilies of L1 elements, differing dramatically in the first 1.2 kb of sequence at their 5' ends, were identified in the prosimian primate, *Galago garnettii*. Interesting patterns of sequence similarity were observed between the galago subfamilies, and with the L1s from human and from another prosimian, the slow loris. Furthermore, members of one of the subfamilies have six to eight tandemly repeated units of 73 bp, starting about 730 bp from their 5' ends. Such tandem repeats have not been reported in other primate L1s, but a striking sequence similarity was found between the galago tandem repeats and those previously described at the 5' termini of some mouse L1s [Loeb, D. D. et al. Mol. Cell. Biol. 6, 168-182, 1986]. Although the similar sequence indicates a shared, conserved function, the galago repeats are sub-terminal and therefore cannot serve as portable RNA polymerase II promoters, as has been suggested for the mouse tandem repeats.

INTRODUCTION

The L1 family of interspersed repetitive elements (LINE-1) is found in a variety of mammalian genomes (1), including those of the mouse, rat, rabbit, dog, monkey and human (2-7). Elements are more similar within than between species, due to recent amplification and/or homogenization of sequences within each mammalian lineage (8,9). It is clear that at least some of the more than 10⁴ L1 elements in the human genome are currently mobile, since de novo insertions of L1 elements into the factor VIII gene are associated with two cases of hemophilia A (10).

Full length elements are 6-7 kb, but many elements are truncated at the 5' end (2,11). L1 transcripts of various sizes have been identified in human cells (6). There appear to be RNA polymerase II L1 transcripts (12), and in at least one case some transcripts are approximately full-length (13). dA-rich tails are a characteristic feature of L1 elements (14), suggesting that they transpose through a polyadenylated RNA intermediate in a process employing a reverse transcriptase, as processed pseudogenes do. However, in contrast to processed pseudogenes, some L1s encode long open reading frames (7,15,16) with similarity at the protein level to regions of retroviral pol and other proteins with reverse transcriptase activity. Such reverse transcriptase similarity has been reported in a consensus human/African green monkey L1 sequence (17), as well as in mouse and loris (prosimian) L1s (18,19).

A possible model for L1 element transposition begins with RNA polymerase II transcription of an element residing in the genome. This transcript is then reverse transcribed into DNA employing the self-encoded

reverse transcriptase and this copy re-inserts into the genome. It is unclear how the newly generated Ll elements could retain the upstream regulatory sequences presumably necessary for subsequent RNA polymerase II transcription and continued movement. For murine Ll elements Loeb et al. (18) have proposed a possible remedy to this problem, termed in this paper "the portable promoter hypothesis". They suggest that a 208 bp sequence found tandemly repeated at the 5' ends of some mouse Lls could serve as a pseudo-internal promoter. Transcripts beginning at an upstream copy of the repeat would still carry downstream units of the 208 bp sequence which would then drive transcription of the new element. Although this portable promoter hypothesis is attractive, it fails to explain why Ll elements with 5' terminal tandem repeats have not yet been described in other mammalian species. Perhaps only a small fraction of Lls in these species remain active, with tandem repeats, and these have not been found yet. Alternatively, the 5' terminal tandem repeats may not serve as portable promoters and/or may not be necessary for transposition.

The Drosophila I, F and G elements are similar to the mammalian Ll elements. Structurally they lack the long direct terminal repeats characteristic of retrovirus like elements, and carry dA rich tracts at their 3' ends (20-23). The I and F elements encode amino acid sequence even more similar to Lls than to any other reverse transcriptase (20,21). Of particular interest are the I elements, which move rapidly following certain genetic crosses, resulting in hybrid dysgenesis (20). These elements are therefore still active, and not defective evolutionary relics. It is important to note that these Drosophila elements do not carry 5' terminal repeats (20). That is, the portable promoter hypothesis does not seem capable of explaining their mobility.

This report describes the results of partial sequence analysis of 18 Ll elements from the prosimian primate Galago garnettii, the African bushbaby, with particular emphasis on their 5' ends, which might play an important role in transposition. Surprisingly, two distinct subfamilies of Ll elements were found. There are dramatic differences in the first 1.2 Kb of their 5' portions. The 5' terminus of one subfamily was defined, and the relationships between the galago subfamilies and Lls of other species was examined. The scrambled blocks of similar sequence revealed something of the complex pattern of evolution of Ll elements. Furthermore, one subfamily was found to carry in its 5' region a series of tandem repeats with significant homology to the murine 5' terminal repeats. Although this sequence similarity suggests a shared function, the galago repeats are not terminal and therefore cannot serve as portable promoters.

MATERIALS AND METHODS

DNA for the Galago garnettii library was prepared from cultured galago cells and was partially digested with MboI and cloned into EMBL3 phage. DNA sequence was determined using the Sanger enzymatic di-deoxy chain termination method (24,25), with reverse transcriptase. More than 98% of the sequence shown in Figs. 2 and 3 has been determined on both strands. The Beckman Microgenie program was used to aid in the alignment and comparison of sequence.

The Southern blot was performed using an alkaline transfer method (26) and GeneScreenPlus. The tandem repeat probe was radiolabeled by the oligonucleotide random primer method (27), and was prepared from LlGg-2. The hybridization was performed at 65° C in a solution containing 1M NaCl, 50 mM Tris pH 7.4, 1% SDS, 10% dextran sulfate, and 150 ug yeast tRNA and salmon sperm DNA/ml. Washes were in 2X SSC, 1% SDS at the same temperature.

RESULTS

Galago L1 elements were isolated by screening a galago EMBL3 library with a probe from the 5' half of a human L1, which contained some sequence similar to galago DNA, and some sequence at the 5' L1 end which did not cross-hybridize (28). Identification of galago L1 5' regions and definition of 5' ends was complicated by their variable 5' truncation. Several elements were sequenced from downstream, where human and galago elements do cross-hybridize, towards their 5' ends, and continued similarity was interpreted to indicate continued presence of L1 sequence. Five non-truncated L1s were identified, from a total of eighteen elements partially sequenced.

Two subfamilies of L1s in galago

The DNA sequence analysis revealed that (at least) two subfamilies of L1s extensively different from each other in their 5' portions are present in the galago. Restriction maps of the 5' regions of three clones from one subfamily, L1Gg-1, L1Gg-4 and L1Gg-13 are shown in Fig.1A, and two clones from the second subfamily, L1Gg-6 and L1Gg-B, are shown in Fig.1B. The L1Gg-1 and L1Gg-4 maps contain only two shared restriction sites, both for PstI, and L1Gg-13 contains no sites in common with these clones. Nevertheless, the L1Gg-1, L1Gg-4 and L1Gg-13 elements are quite similar throughout, as is shown in their sequence comparisons in Fig. 2. The 5' endpoint of the alignment represents the first base identical between the three elements, and indeed the first 16 bases are the same for all the elements, while the upstream sequence appears totally random. This implies that the 5' terminus of non-truncated members of the L1 subfamily has been defined. The most 3' sequence shown, beginning at base number 1604 in L1Gg-1 and L1Gg-4, and continuing to the last base in both, is similar to the prosimian slow loris (*Nycticebus coucang*) L1s sequenced by Hattori, et al. (19). However, the first 1.6 kb of the sequence appears to be unique to galago.

One galago subfamily contains tandem repeats

The most striking feature of the sequence shown in Fig. 2 is the presence of a number of tandemly repeated units, which begin and end at base numbers 726 and 1178 in L1Gg-1 and at base numbers 736 and 1167 in L1Gg-4. The individual galago tandem repeat units are aligned with a 73 bp consensus sequence, derived from L1Gg-1 and L1Gg-4, in Fig. 3. L1Gg-1 and L1Gg-4 contain 6 1/3 and 6 unit repeats, respectively. Some sequence near the start of the tandem repeats, which is present in L1Gg-1 and L1Gg-13, is absent in L1Gg-4. This is evident in Fig. 2, and as a result L1Gg-4 does not contain the most 5', incomplete, tandem repeat unit shown for L1Gg-1 in Fig. 3. The individual repeats are between 67% and 85% similar to the consensus derived from all 12 1/3 copies. There are seven bases which are invariant in all copies and these are marked with asterisks.

It is interesting to note that there is some variability in the number of copies of the tandem repeats, and that there is a detectable pattern of sequence differences between individual units. One sequenced L1 (L1Gg-2) contained a total of eight unit tandem repeats, instead of six (data not shown). As the differences from the consensus in Fig. 3 reveal, the tandem repeats in L1Gg-1 and L1Gg-4 have a structured order, with the first full repeat in L1Gg-1 most like the first in L1Gg-4, and the second like the second, etc.

Fig. 1C is an autoradiogram of a galago genomic Southern blot hybridized to a probe from L1Gg-2 containing only tandem repeat DNA. The band at about 1 kb in the PstI digest could be predicted from the L1Gg-1 restriction map. The diffuse appearance of this and other bands may be caused by differences in size due to the presence of different numbers of

tandem repeat units, or fractions of units, in individual elements. Use of the L1Gg-2 probe in screening the galago EMBL3 library allowed us to estimate that the tandem repeats are present in about 1.5×10^3 galago L1 elements/haploid genome.

Primate L1 5' end sequence comparisons

A schematic sequence comparison between 5' L1 regions from another prosimian, the loris, and the tandem and non-tandem subfamilies in galago,

Figure 1

Restriction maps of the 5' ends of members of two galago L1 subfamilies, and an autoradiogram of a genomic galago Southern blot hybridized to a probe containing the galago tandem repeats. A) The restriction maps of the sequenced regions of the clones containing L1Gg-1, L1Gg-4 and L1Gg-13, which are members of the tandem subfamily. The enzymes used are: B=BglII, RI=EcoRI, RV=EcoRV, K=KpnI, N=NheI and P=PstI. The EcoRI sites located most 5' are part of the flanking sequence, and not within the L1, in all cases. B) Restriction maps of the 5' ends of two members of the non-tandem subfamily, L1Gg-6 and L1Gg-B. The enzymes used here are the same as in Fig.1A with the addition of: H=HindIII, Pv=PvuII, and X=XhoI. There are no EcoRV or NheI sites in L1Gg-6 or L1Gg-B. A few of the sites in Fig.1B have been determined from sequence on one strand only. The most 5' KpnI site in L1Gg-B is in the sequence flanking the element. C) An autoradiogram of a Southern blot of 0.1 µg of galago DNA digested with PstI, EcoRI or EcoRI plus PstI. The probe contained tandem repeat DNA only. The sizes marked are in kb.

Figure 2

The DNA sequence and comparison of the 5' ends of three members of the galago tandem subfamily, L1Gg-1, L1Gg-4 and L1Gg-13. The sequence of L1Gg-1 is shown on the top line, beneath it is the sequence of L1Gg-4, with only differences shown, and on the bottom line is L1Gg-13. Blank spaces indicate the same base is present in L1Gg-4 or L1Gg-13 as in L1Gg-1. The presence of another base is indicated, and dashes show that no base is present and have been introduced to improve the alignment. The numbers at the right indicate base pairs from the 5' end. Less sequence is shown for L1Gg-13 than for the other L1s, only a portion of the tandem repeat region is included. The bold lower case letters at 726 bp in L1Gg-1 and 736 bp in L1Gg-4 mark the beginning of the tandemly repeated sequence, and those at 1178 bp in L1Gg-1 and 1167 bp in L1Gg-4 represent the end of the tandem repeats.

	10	20	30	40	50	60	70	
	* * *		* *		*		*	
CTGCGAGCAGAGGCTT-GCTA--GGTGTGGACTGAGACCTGTGGGTGGGGGCATGGTCCCTGGCTCCAACCTGGGC								
				LlGg-1	A..TC...A...-.C..TG....			750
				LlGg-4	-----			735
...A.....-A.---.T.ACAG...A.A.CAAT.....AT.....A.AA.								822
...T.....A...A.---CA.....CA.....A.A.CAAA...G.A.A.TG...AT.....AAAA.								808
..T.A.....T.A...-A..T---A...CA.C....A...CTA.....A...T....GTG..A.								895
..CT.....T.....A..G---A...A..CA.....A...T.A..TG.....A...T....G.G....								881
.A.A.....TA.C--..A...A.....CAAA...A.-TG..A.G.T...A.A.T....A..								968
TG.T.....TA.....A.C--..A.....C..A.-TGCA..G....A..A..A.A..								953
...T.A.....TC.AG--CA.....TGG.CA...AAA.....T.....A..A..AC..T.								1042
...A.A.C.....-C.-G.CA....T...A.TTG.CA.ACA.....T.....T.								1027
TG..A.A.....A..A..G--A.A....CA-----G.....G.....								1108
TG..A.....A--G--A.....TG.....G.....A.								1099
..AT.....AAGAG.TC.ACA..T.C.....T.TG.....A.TG...AT.....A..								1178
...T.....AAG.A.AC.AC.....T.TG.....A.TG...ATT.T.....								1167

Figure 3

A consensus sequence for the tandem repeats in LlGg-1 and LlGg-4, with sequence comparisons of the individual repeat units to the consensus. The top line is a consensus sequence derived from all 12 1/3 copies of the tandem repeat present in LlGg-1 and LlGg-4. Asterisks indicate bases which are the same in all 12 1/3 copies. The individual repeats from LlGg-1 are shown in order starting on the next line, with each LlGg-4 repeat below the corresponding full-length LlGg-1 repeat. Dots indicate that the base present is the same as the consensus, if a different base is present it is shown, and dashes have been introduced to improve alignment where a gap occurs in one or more copies. The numbering system at the right is the same as in Fig. 2.

is found in Fig. 4. The data for the loris is from a consensus provided by Hattori, et al. (19), for the tandem subfamily is from the sequence of LlGg-1, LlGg-4 and LlGg-13, and for the non-tandem subfamily is from the sequence of LlGg-6 and LlGg-B. In the diagram, boxes labeled identically are blocks of DNA sequence with at least 59% similarity (see Fig. 4 legend). The N3 block of sequence is shared between all three groups, but only the non-tandem galago subfamily shares the N2 box with loris. The N2 and N3 boxes are not adjacent in the galago non-tandem subfamily, as they are in loris. The size of the N2 box in galago is an average, as it measured 75 bp in LlGg-6 and 240 bp in LlGg-B, with the same 3' endpoint. G4 and G5 are sequences unique to the non-tandem galago subfamily, though since only two elements in this subfamily have been sequenced, it is possible that one of the two is truncated and that the G4 region extends further 5'. The X represents a stretch of sequence about 200 bp long which is not similar between the two non-tandem clones, LlGg-6 and LlGg-B. The G2 and G3 boxes are common to both galago subfamilies, though the sequence located between them is not. The TAN and G1 boxes are found only in the tandem subfamily, the TAN box contains the tandem repeats plus some flanking 5' sequence. The point where TAN juxtaposes G3 marks the end of the last of the 73 bp repeats. The African green monkey/human consensus (16) and one partially sequenced galago Ll (LlGg-10, data not shown) show similarity in sequence starting at distances 1.2 and 1.9 kb from their respective 5' termini. The African green monkey/human consensus open reading frame begins near 1.2 kb from the 5' end (15).

Figure 4

A schematic comparison of the DNA sequence at the 5' ends of loris L1s and of the galago tandem and non-tandem L1 subfamilies. Boxes marked with the same notation represent similar sequences. The scale below measures bp. The top line represents the 5' end of a loris L1 consensus sequence (19), the second line is the galago tandem subfamily (based on L1Gg-1, L1Gg-4 and L1Gg-13) and the bottom line is the galago non-tandem subfamily. Only the 5' ends of the elements are depicted in this illustration, full length L1s are 6-7kb. The data for the non-tandem subfamily schematic comparison is from sequence analysis of L1Gg-6 and L1Gg-B (unpublished data), much of this sequence is from one strand only. N/H represents DNA which is 62% similar between a loris and an African green monkey/human consensus (19,16), the galago sequence similar to human begins at about 1.9 kb from the galago 5' end in a partially sequenced L1 (L1Gg-10, unpublished data). N3 is common to loris and to both galago subfamilies (at least 72% similar). N2 is found in the loris and in the non-tandem galago subfamily only, it is more than 72% similar. G2 and G3 are 59% and 60% similar, respectively, between the tandem and non-tandem subfamilies, and are not found in loris. G1 and TAN are exclusively found in the tandem subfamily. G4 and G5 are in the non-tandem subfamily only. The X represents a stretch of sequence about 200 bp long which is not similar between L1Gg-6 and L1Gg-B.

DISCUSSION

The results of this study show that there are at least two subfamilies of L1 elements in the prosimian primate galago; they differ dramatically in sequence composition near their 5' ends. A notable characteristic of one of

Figure 5

Sequence similarity between a portion of the galago (L1Gg) and of the mouse (L1Md) consensus tandem repeats. The numbers above L1Gg are from the 73 bp consensus in Fig. 3, and the L1Md numbers are from a consensus derived from the 208 bp mouse tandems of Loeb, et al. (18). There are two mismatches in the 17 bp sequence and they are in small case. There are asterisks above the three bases in the galago consensus which are identical in all twelve units of the tandem repeats in L1Gg-1 and L1Gg-4.

these subfamilies is the presence of between six and eight copies, in individual subfamily members, of a 73 bp tandemly repeated unit, at a distance of about 730 bp from the 5' end. These are the first tandem repeats found in primate Lls, however they may be an evolutionarily conserved structural motif of Lls, since tandem repeats have previously been identified in the mouse (18) and rat (29,30). The mouse repeats are positioned at the extreme 5' termini of Lls, unlike the galago and rat repeats which are internal. The number of units of tandem repeats varies from element to element in all three species, perhaps due to unequal cross-over recombination between elements (31) or to slipped mispairing during DNA replication (32). The individual repeated tandem units are more similar to each other in the mouse elements than in the galago. Of 5 complete mouse copies all are at least 86% alike, while 12 galago copies are only 67-85% similar to a galago consensus.

Interestingly, the galago repeats are similar in sequence to those of the mouse, although the rat repeats show no significant homology to either galago or mouse repeats. As shown in Fig. 5 a 17 bp stretch of DNA from the galago consensus matches a mouse consensus at 15 positions. Of the three positions in the 17 which were identical in all twelve tandem repeats in LlGg-1 and LlGg-4, all three also match with the mouse consensus. This degree of similarity between the galago and the mouse tandem repeats is highly significant since a sequence of only 73 bp is being compared to one of only 208 bp in length. This conservation argues that the similar sequences have some shared, essential function in Lls. The galago repeats, however, are not terminal and cannot, therefore, provide portable promoters as has been suggested for the mouse repeats (18).

The conserved region of the galago and mouse tandem repeats might serve some other essential function for Ll transcription or transposition, though it is not clear why tandemly repeated units are then necessary. Indeed it is possible that a single copy of the conserved region is sufficient for some function. Thus, elements in the mouse or in primates which do not contain tandem repeats, but do contain this essential sequence, might still be capable of transposition. Shorter portions of the 17 bp region from galago, which is conserved in the mouse, were found near the 5' ends of consensus loris [453 GAGCA-AGGATTG 464] and human [419 AGCAGAG(ag)GCT 430] Ll sequences (19,16), which have no tandem repeats. It is interesting to note that these matches include the three bases marked in Fig. 5 as common to all twelve tandem repeats in LlGg-1 and LlGg-4.

Though at least two subfamilies of Lls are present in the mouse genome (18,33), the only distinct subfamily previously identified in a primate genome is that of a human Ll with a 131 bp insertion (34). The galago genome harbors at least two very different subfamilies of Lls. Indeed, the two galago non-tandem elements sequenced, LlGg-6 and LlGg-B, may in fact be members of distinct subfamilies, since they each contain some sequence (the X box in Fig. 4) not found in the other. Also, a region of sequence (N2, in Fig. 4) from a progenitor to the loris and galago elements appears to have been incorporated into both the loris subfamily and the non-tandem galago subfamily of Lls, but is not present in the galago tandem subfamily, where it may since have been deleted. It is interesting that the G2 box and downstream G3 and N3 boxes (see Fig. 4) are common to the tandem and non-tandem subfamilies of galago, while the sequence between is unrelated. It is not surprising that the loris and galago elements share more similar sequence with each other than with the human consensus, but the substantial differences between the Lls of the closely related loris and galago, and indeed between galago subfamilies, was unexpected. It is clear that there is still much to be learned about the evolution, function and mechanism of transposition of Ll elements.

ACKNOWLEDGEMENTS

The Galago garnettii cells cultured to make DNA were provided by the Duke University Primate Center. M. Singer and M. Hattori kindly provided unpublished consensus sequences. We would like to thank A. Lamb for culturing cells, and for his involvement in the initial work showing that human L1 5' ends do not cross-hybridize with galago. We also appreciate the help of B. Eident in preparing phage DNAs and DNA sequencing. We thank C. Duncan for critically reading the manuscript. This work was supported by National Institutes of Health grant GM30816 and National Cancer Institute grant CA00852.

*To whom correspondence should be addressed

+Present Address: Dept. of Microbiology and Molecular Genetics, University of Cincinnati College of Medicine, 231 Bethesda Ave., Cincinnati, OH 45267, USA

REFERENCES

1. Burton, F.H., Loeb, D.D., Voliva, C.F., Martin, S.L., Edgell, M.H. and Hutchison, C.A. III (1986) *J. Mol. Biol.* **187**, 291-304.
2. Voliva, C.F., Jahn, C.L., Comer, M.B., Hutchison III, C.A. and Edgell, M.H. (1983) *Nucleic Acids Res.* **11**, 8847-8859.
3. D'Ambrosio, E., Waitzkin, S.D., Witney, F.R., Salemm, A. and Furano, A.V. (1986) *Mol. Cell. Biol.* **6**, 411-424.
4. Demers, G.W., Brech, K. and Hardison, R.C. (1986) *Mol. Biol. Evol.* **3**, 179-190.
5. Katzir, N., Rechavi, G., Cohen, J.B., Unger, T., Simoni, F., Segal, S., Cohen, D. and Givol, D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1054-1058.
6. Lerman, M.I., Thayer, R.E. and Singer, M.F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3966-3970.
7. Potter, S.S. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1012-1016.
8. Deininger, P.L. and Daniels, G.R. (1986) *Trends in Genetics* **2**, 76-80.
9. Martin, S.L., Voliva, C.F., Hardies, S.C., Edgell, M.H. and Hutchison III, C.A. (1985) *Mol. Biol. Evol.* **2**, 127-140.
10. Kazazian, H.H. Jr., Wong, L., Youssoufian, H., Scott, A.F., Philips, D.G. and Antonarakis, S.E. (1988) *Nature* **332**, 164-166.
11. Grimaldi, G., Skowronski, J. and Singer, M.F. (1984) *EMBO J.* **3**, 1753-1759.
12. Shafit-Zagardo, B., Brown, F.L., Zavodny, P.J. and Maio, J.J. (1983) *Nature (London)* **304**, 277-280.
13. Skowronski, J. and Singer, M.F. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6050-6054.
14. DiGiovanni, L., Haynes, S.R., Misra, R. and Jelinek, W.R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6533-6537.
15. Martin, S.L., Voliva, C.F., Burton, F.H., Edgell, M.H. and Hutchison III, C.A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2308-2312.
16. Singer, M.F. and Skowronski, J. (1985) *Trends Biochem. Sci.* **10**, 119-122.
17. Fanning, T. and Singer, M. (1987) *Nucleic Acids Res.* **15**, 2251-2260.
18. Loeb, D.D., Padgett, R.W., Hardies, S.C., Shehee, W.R., Comer, M.B., Edgell, M.H. and Hutchison III, C.A. (1986) *Mol. Cell. Biol.* **6**, 168-182.
19. Hattori, M., Kuhara, S., Takenaka, O. and Sakaki, Y. (1986) *Nature (London)* **321**, 625-628.
20. Fawcett, D.H., Lister, C.K., Kellett, E. and Finnegan, D.J. (1986)

- Cell 47, 1007-1015.
21. Di Nocera, P.P. and Casari, G. (1987) Proc. Natl. Acad. Sci. USA 84, 5843-5847.
 22. Di Nocera, P.P., Digan, M.E. and Dawid, I.B. (1983) J. Mol. Biol. 168, 715-727.
 23. Di Nocera, P.D.P., Graziani, F. and Lavorgna, G. (1986) Nucleic Acids Res. 14, 675-691.
 24. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
 25. Duncan, C.H. (1985) New England Nuclear Product News 4, 6-7.
 26. Chomczynski, P. and Qasba, P.K. (1984) Biochem. Biophys. Res. Commun. 122, 340-344.
 27. Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.
 28. Lloyd, J.A., Lamb, A.N. and Potter, S.S. (1987) Mol. Biol. Evol. 4, 85-98.
 29. D'Ambrosio, E., Waitzkin, S.D., Witney, F.R., Salemme, A. and Furano, A.V. (1986) Mol. Cell. Biol. 6, 411-424.
 30. Soares, M.B., Schon, E. and Efstratiadis, A. (1985) J. Mol. Evol. 22, 117-133.
 31. Smith, G.P. (1976) Science 191, 528-535.
 32. Farabaugh, P.J., Schmeissner, U., Hofer, M. and Miller, J.H. (1978) J. Mol. Biol. 126, 847-863.
 33. Wincker, P., Jubier-Maurin, V. and Roizes, G. (1987) Nucleic Acids Res. 15, 8593-8606.
 34. Hattori, M., Hidaka, S. and Sakaki, Y. (1985) Nucleic Acids Res. 13, 7813-7827.