

Wide distribution of short interspersed elements among eukaryotic genomes

Olga R. Borodulina^{a,b,*}, Dmitri A. Kramerov^a

^aEngelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov St., Moscow 117984, Russia

^bInstitute of Gene Biology, Russian Academy of Sciences, 34/5 Vavilov St., Moscow 117334, Russia

Received 14 July 1999

Abstract Most short interspersed elements (SINEs) in eukaryotic genomes originate from tRNA and have internal promoters for RNA polymerase III. The promoter contains two boxes (A and B) spaced by ~33 bp. We used oligonucleotide primers specific to these boxes to detect SINEs in the genomic DNA by polymerase chain reaction (PCR). Appropriate DNA fragments were revealed by PCR in 30 out of 35 eukaryotic species suggesting the wide distribution of SINEs. The PCR products were used for hybridization screening of genomic libraries which resulted in identification of four novel SINE families. The application of this approach is illustrated by discovery of a SINE family in the genome of the bat *Myotis daubentoni*. Members of this SINE family termed VES have an additional B-like box, a putative polyadenylation signal and RNA polymerase III terminator.

© 1999 Federation of European Biochemical Societies.

Key words: Repetitive DNA sequence; Short interspersed element; Retroposon; Polymerase chain reaction; Eukaryotic genome; *Myotis daubentoni*

1. Introduction

Short interspersed elements (SINEs) or short retroposons are 80–400 bp repetitive DNA sequences that proliferate in eukaryotic genomes via transcription followed by reverse transcription [1]. Usually the number of a SINE copies ranges from 10^4 to 10^5 ; such copies feature significant sequence variability (5–25%). Most SINEs are transcribed by RNA polymerase III and their 5'-regions demonstrate reasonable sequence similarity to certain tRNAs (commonly, lysine and arginine) [2,3]. This suggests that SINE families originated from tRNAs or respective genes.

About 25 SINE families have been found in mammals, fishes, molluscs and plants [4]. However, it is yet unclear whether SINEs are present in the genomes of most eukaryotic species.

Recently SINEs were used as phylogenetic markers and proved to be quite reliable for genotyping of middle rank taxa – families and orders [5–7]. This approach as well as general studies of SINEs require identification of novel

SINE families. Three methods to clone novel SINE families based on their functional properties are currently known.

Complementary SINE copies are present in many nuclear pre-mRNA molecules providing for long hairpin structures. Such double-stranded structures can be isolated from total pre-mRNA and used for screening of genomic libraries [8,9]. However, this method has low efficiency and pre-mRNA isolation requires large amounts of native tissues which are not always available.

The second method relies on SINE redundancy in the genome relative to single-copy sequences [10]. Using this approach the authors selected clones with repetitive sequences from a genomic library by hybridization with labeled total genomic DNA. This simple procedure, however, is followed by laborious stages of cross-hybridization, mapping and extensive sequencing of positive clones in order to filter out non-SINE repetitive sequences (LINEs and their truncated variants, retrotransposons, other mobile elements and various satellite DNAs).

The third approach utilizes two functional properties of SINEs: redundancy in the genome and transcription by RNA polymerase III. Accordingly, the in vitro transcript of total genomic DNA by RNA polymerase III was used as a probe for screening of a genomic library [11]. Although this method was used successfully by the authors [12–15], it did not work in our hands.

Here we propose a reliable and simple way to detect and clone novel tRNA-derived SINEs. Like tRNA genes, most SINEs have an internal promoter for RNA polymerase III composed of two boxes (A and B) spaced by 30–35 bp [9,16]. For this purpose we used PCR-amplified total genomic DNA as a template and primers specific to boxes A and B for PCR amplification. This reaction, designated A-B PCR, amplifies the region between boxes A and B. The resulting DNA fragments were used for screening of a genomic library to isolate full-size SINE copies.

2. Materials and methods

2.1. Tissue and DNA samples

Bat tissues were provided by V. Matveev and A. Borissenko (Zoological Museum, Moscow State University, Moscow, Russia); tissue samples from other mammalian species were provided by Drs. A. Bannikova (Biological Department, Moscow State University), E. Lyapunova (Institute of Developmental Biology, Moscow, Russia), O. Likhnova (Institute of Ecology and Evolution, Moscow, Russia), and S. Popov (Moscow Zoo, Moscow, Russia). Non-mammalian DNA samples were provided by Drs. V. Grechko, M. Evgen'ev (Institute of Molecular Biology, Moscow, Russia), A. Lomov, and T. Samigullin (Belozersky Institute, Moscow State University). DNA was isolated from frozen or ethanol-fixed tissues (liver, kidney or muscle) by incubation with proteinase K followed by phenol/chloroform extraction.

*Corresponding author. Fax: (7) (095) 135-1405.
E-mail: kramerov@genome.eimb.relarn.ru

Abbreviations: SINE, short interspersed element; LINE, long interspersed element; Y, pyrimidines; R, purines; 1×SSC, 0.15 M NaCl and 0.015 M Na citrate; TE, 10 mM Tris-HCl, pH 8.0 and 1 mM EDTA

2.2. A-B PCR

The reaction was carried out in 100 μ l mix containing 0.1 ng of genomic DNA, 0.3 μ M primers A and B (Fig. 1), 200 μ M dNTPs, 2 U Taq polymerase and the reaction buffer (67 mM Tris-HCl pH 8.6, 2.5 mM MgCl₂, 16.6 mM (NH₄)₂SO₄ and 0.001% Triton X-100) and included 27 amplification cycles (95°C, 1 min; 34°C, 1 min and 72°C, 1 min). The PCR mix was phenol/chloroform extracted; DNA was ethanol precipitated, dissolved in 10 μ l 10 mM Tris-HCl, pH 8.0 and 1 mM EDTA (TE) and analyzed by electrophoresis in 5% NuSieve (FMC) agarose gel.

2.3. Labeling PCR products

The gel slices (70 mg on average) containing DNA bands were cut off and melted in the presence of three volumes TE. 2 μ l of this solution was added to a PCR reaction mixture (25 μ l) containing 30 μ Ci [α ³²P]dATP and 17 PCR cycles were carried out as described above.

2.4. Library construction and screening

Libraries were prepared by cloning genomic DNA digested with *Hind*III and *Eco*RI into pBluescript SK⁺. Colony hybridization used for library screening was carried out at 65°C in a mixture containing 4 \times SSC (1 \times SSC=0.15 M NaCl and 0.015 M Na citrate), 0.5% SDS, 5 \times Denhardt's solution, 0.1 mg/ml boiled herring sperm DNA and a labeled PCR product. Filters were washed for 30 min at 42°C in 0.1 \times SSC and 0.1% SDS.

2.5. Cloning PCR products

Some PCR products were isolated from agarose gel by electrotransfer to a DEAE filter (NA 45, Schleicher and Schuell) placed in a slit ahead of the band. DNA was eluted from the membrane by incubation for 30 min at 65°C in 1 M NaCl, Tris-HCl, pH 7.5 and 1 mM EDTA. Thus isolated DNA fragments were cloned into a *Sma*I site of pGEM-7Zf⁺.

Double-stranded plasmid templates were sequenced using the dideoxynucleotide method with Sequenase 2.0 (Amersham Life Science) according to the manufacturer's instructions with modifications [17].

3. Results and discussion

We deduced consensus sequences of boxes A and B from nucleotide sequence alignment of 11 families of tRNA-derived SINEs from the genomes of mammals [9], [6,18,19], reptiles [12], fishes [13], and mollusks [14,15]. Two box A consensus sequences (A and A') deduced differed by a single pyrimidine insertion in the A' (Fig. 1a). Oligonucleotides corresponding to these consensus sequences were used in PCR (Fig. 1b).

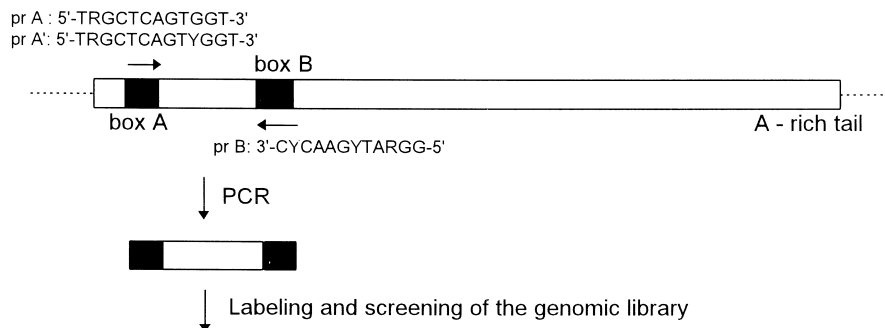
We used small amounts of genomic DNA (0.1 ng) by far insufficient for amplification of tRNA genes but enough to amplify the much more abundant SINEs. A-B PCR analysis of 18 mammalian species from 11 orders is presented in Fig. 2a. The expected 55 bp PCR product was observed in 10 species (lanes 1, 4, 5, 7, 8, 11–13, 16 and 17), while in four (lanes 9, 10, 14 and 15) the band was faint. No product was observed in the remaining four cases (lanes 2, 3, 6 and 19). We tried A-B PCR with DNA of non-mammalian species as well (birds, reptiles, amphibians, fishes, invertebrates and plants) and the 55 bp PCR product was observed in 16 out of 17 species (Fig. 2b). Since the results of A-B PCR were positive for most species tested, it can further be used to screen a wide range of eukaryotic species.

Absence of the A-B PCR product in some cases is most likely due to significant mismatch between primers and the template DNA, as is the case with the rabbit genome where the promoter of C SINE [20] differs from the primer sequences by five nucleotides (Fig. 2a, lane 6).

We further tested the specificity of the A-B amplification on the example of water bat *Myotis daubentoni*. In this case two

| | box A | box B | Ref |
|---------------|------------------------|-----------------------------|-----|
| polIII cons | TGGC N AGT-GG | GGTTCGANNCC | 16 |
| Fok | TGGCTCAGTTGGT | GGGTTC-ATTCC | 13 |
| Sma | TAGCTCAGTTGGT | GGGTTCGATTCC | 13 |
| SK | TAGCTCAGT C GGT | GGGTTCGATaCC | 14 |
| OR | TGGCTtAGT-GGT | GGtTTCGATTCC | 15 |
| OK | TGGCTgAGa-GGT | GGGTTCAgTTCC | 15 |
| TORT | TAGCTCAGT-GGT | GAGTTCAAT-CC | 12 |
| SSPRE | TGGCTCAGT-GGT | aGGTTCGATCCC | 18 |
| B2 | TGGCTCAGT-GGT | GAGTTCAATTCC | 9 |
| DIP | TGGCTCAGY-RGT | GRGTTCGATTCC | 6 |
| MEN | TGGCTCAGcCGGT | GGGTTCaATTCC | 6 |
| ID | TRGCTCAGT-GGT | GGGTTCGgTCCC | 19 |
| SINE cons A: | TRGCTCAGT-GGT | cons B: GRGTTCRATYCC | |
| SINE cons A': | TRGCTCAGTYGGT | | |

a



b

Fig. 1. a: Boxes A and B of 11 tRNA-derived SINEs and their consensus (cons. A, cons. A' and cons. B). The nucleotides different from the consensus are indicated in lower case. The general consensus sequences of boxes A and B of various RNA polymerase III transcribed genes are shown to allow comparison (upper line). b: A scheme of A-B PCR.

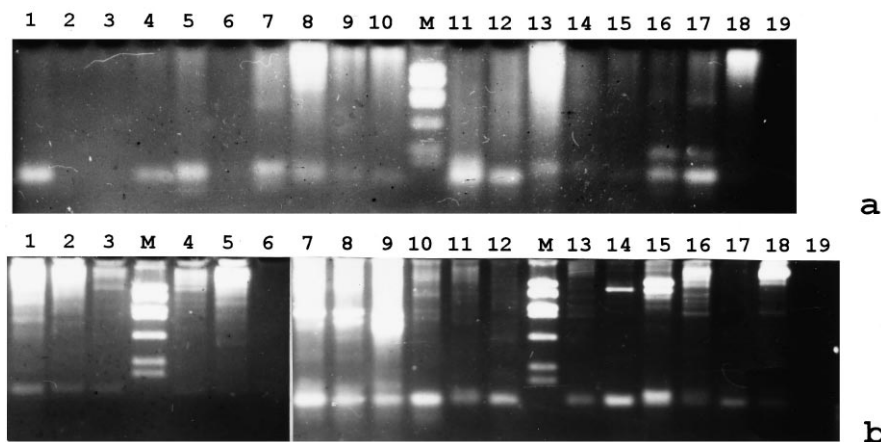


Fig. 2. The A-B PCR of genomic DNA. a: Mammalian species. 1, mouse; 2, tree shrew (*Tupaia glis*); 3, human; 4, bottle-nosed dolphin (*Tursiops truncatus*); 5, cattle; 6, rabbit; 7, dog; 8, cat; 9, walrus; 10, short-eared elephant shrew (*Macroscilides proboscideus*); 11, Daurian hedgehog (*Erinaceus dauricus*); 12, Eurasian common shrew (*Sorex araneus*); 13, large Japanese mole (*Mogera robusta*); 14, Russian desman (*Desmana moschata*); 15, greater horseshoe bat (*Rinolophus ferrumequinum*); 16, water bat (*M. daubentoni*); 17, Nathusius' pipistrelle (*Pipistrellus nathusi*); 18, short-nosed fruit bat (*Cynopterus sphinx*); 19, no DNA. b: Non-mammalian species. 1, rock lizard (*Lacerta valentini*); 2, rapid fringed lizard (*Eremias velox*); 3, common frog (*Rana ridibunda*); 4, gull (*Larus cachinans*); 5, crane (*Grus leucogeranus*); 6, no DNA; 7, brown trout (*Salmo trutta*); 8, pink salmon (*Oncorhynchus gorbusha*); 9, whitefish (*Coregonus clupeaformylus*); 10, sea urchin (*Strongylocentrotus droebachiensis*); 11, earthworm (*Lumbricus terrestris*); 12, capricorn beetle (*Acanthocinus aedilis*); 13, bark beetle (*Ips sexdentatus*); 14, ant (*Lasius niger*); 15, ant (*Mirmica ruginodis*); 16, moss (*Funkia ovata*); 17, buckwheat (*Fagopyrum tataricum*); 18, maize; 19, no DNA. M: size marker, pGEM3Z DNA, digested with *Hae*III. The two lowest bands are 102 and 80 bp fragments.

PCR products (55 and 90 bp) were observed (Fig. 2a, lane 16). Both fragments were cloned and sequenced (Fig. 3). The resulting 17 sequences were similar but not identical, suggesting that they may have been amplified from individual copies of a single SINE family.

Screening of the bat genomic library by hybridization with the 55 bp PCR product identified positive clones which constituted 2% of the library. We sequenced several of those either directly (short DNA inserts) or after subcloning and re-hybridizing with the 55 bp probe. Alignment of the sequences is presented in Fig. 4. It is noteworthy that neither of the sequences perfectly matched the primers. However, significant variability and abundance in the genome ensured

PCR amplification of those in this condition. Accordingly, replacement of the A' primer with A did not significantly affect A-B PCR results (data not shown).

Fig. 4 presents a novel SINE named VES (after Vespertilionidae, the family of *M. daubentoni*). The VES element is a typical SINE: its copies are divergent (6–25%), flanked by short (7–17 bp) direct repeats (underlined in Fig. 4) and end with an A-rich sequence. As in several other SINES, the VES element's A-rich tail is followed by a T_{3–6} stretch, an efficient terminator for RNA polymerase III [16,21]. In addition, there is a TC motif located immediately upstream of the A-rich tail, also found in some other SINES [6,20]. Finally, the 5'-end of VES shares similarity with mammalian (bovine) tRNA^{Tyr}

```

MdaPCR.05L: >.GA..G..T-.T...G..T...AA...TTCA...T...T.T.....C.A..TA.-----<
MdaPCR.07L: >-.GA...T-.T..C...C...GAA.A...CCA...-.....A.....CA.-----<
MdaPCR.01L: >.-A...T-.G.....A.....T.....A.....T.....T.....<
MdaPCR.02L: >.....G.C-.....A...A..TC-C.....T..C.T.....C..CCTGG.....<
MdaPCR.03L: >-G.....C-A.....C..C..C.....TT.....T.....C<
MdaPCR.04L: >.....T-.....-.....-.....C.....<
MdaPCR.08L: >.....C.....T.....T.....T.C...T.....A.A.....T.A.....<
MdaPCR.09L: >.....CG..T.....-.....C.....A.....G.....T..CA.-----<
          T-GAGCATY-GACCTATGAACCAGGAGGTCAT-GGTTTCGATTCCCGGTCAGGGCACATGCCYGG      GTTGT-
          * *****-***** * *****
          TAGAGCATC-GACCTATGAACCAARGGGTCAT
MdaPCR.04S: > .G.....T-.G...G.AC..TG.A.....G.<
MdaPCR.06S: >.....G.T-.G...A.G..TG.A.....TC<
MdaPCR.08S: >...CAG.T-.G...A.G..TG.A.....TC<
MdaPCR.10S: >.....TG..-A...GT..G.T.T.T.....<
MdaPCR.02S: > .A.....-G...G..C.....A.....<
MdaPCR.09S: >-.....-.....T.T..GA..G.<
MdaPCR.03S: >...TG..-A...G.....GGA.....<
MdaPCR.05S: >.....-.....GGA.....<
MdaPCR.07S: >.....-.....GGA.....<

```

Fig. 3. Sequences of 90 and 55 bp fragments amplified in the A-B PCR of *M. daubentoni* genomic DNA. The corresponding consensus sequences are given below and above the sequences, respectively; matching nucleotides between the consensus sequences are denoted by asterisks. A and B primers are shown as > and <, respectively.

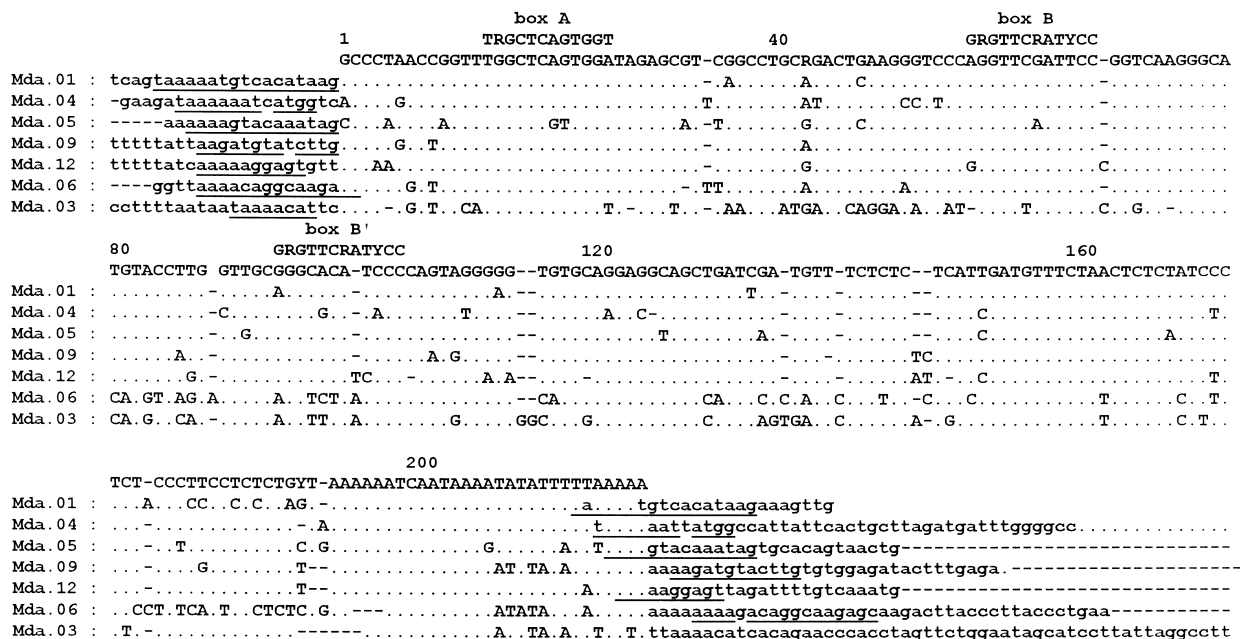


Fig. 4. The VES element. Sequences of seven cloned copies are shown below the consensus. Sequences flanking the SINEs are indicated in lower case and flanking direct repeats are underlined. A-B PCR primers are given at the top of the VES consensus sequence. EMBL Database accession numbers: Y19045–Y19051.

(data not shown) suggesting that the VES SINE family originated from tRNA of this type.

Presence of the second sequence similar to the box B of the RNA polymerase III promoter is a unique feature of the VES element. This sequence, designated box B', is located downstream of the canonical box B (Fig. 4). It is possible that the VES element appeared as a result of recombination between two tRNA genes, pseudogenes or SINEs; in this case the B' box represents a residual RNA polymerase III promoter. The role of box B' is unclear; it may be involved in VES element transcription. At least, appearance of an additional 90 bp fragment in A-B PCR on water bat DNA is clearly due to box B' (Figs. 3 and 4).

Structural analogy between VES and rodent B2 3'-ends, polyadenylation signal AATAAA followed by RNA polymerase III terminator (Fig. 4 and [22,23]), suggests post-transcriptional polyadenylation of VES similar to that of small B2 RNA, which has so far been the only example of this kind.

Dot hybridization of a VES probe to genomic DNA from various species showed the presence of this SINE family in two (Vespertilinidae and Molossididae) out of six bat families tested (data not shown). This result confirms closer relationship between these two bat families as compared to others (Rinolophidae, Hipposideridae, Emballonuridae, and Pteropodidae).

In conclusion, here we detected, isolated and studied three SINE families from insectivore (hedgehogs, shrews, and moles) genomes using A-B PCR (Fig. 2, lanes 11–13). These experimental data will be reported elsewhere.

The method described here is instrumental in searching for novel SINEs. It is simple, consumes less effort and material as compared to conventional methods, yet is specific and efficient. Two prerequisites are set to identification on the elements: RNA polymerase III promoter and abundance in the

genome. This approach can be easily adjusted to a specific target by varying primer sequences or PCR conditions.

We believe this approach will help reveal novel SINE families in various groups of animals and plants.

Acknowledgements: We are grateful to Dr. N. Vassetzky for helpful discussion. This work was supported by The Russian Foundation for Basic Research (Projects 96-04-49536 and 99-04-49177).

References

- [1] Rogers, J.H. (1985) *Int. Rev. Cytol.* 93, 187–279.
- [2] Daniels, G.R. and Deininger, P.L. (1985) *Nature* 317, 819–822.
- [3] Sakamoto, K. and Okada, N. (1985) *J. Mol. Evol.* 22, 134–140.
- [4] Okada, N. and Ohshima, K. (1995) in: *The Impact of Short Interspersed Elements (SINEs) on the Host Genome* (Maraia, R.J., Ed.), pp. 61–75, R.G. Landes, Austin, TX.
- [5] Shimamura, M. et al. (1997) *Nature* 388, 666–670.
- [6] Serdobova, I.M. and Kramerov, D.A. (1998) *J. Mol. Evol.* 46, 202–214.
- [7] Kramerov, D., Vassetzky, N. and Serdobova, I. (1999) *Mol. Biol. Evol.* 16, 715–717.
- [8] Kramerov, D.A., Grigoryan, A.A., Ryskov, A.P. and Georgiev, G.P. (1979) *Nucleic Acids Res.* 6, 697–713.
- [9] Krayev, A.S., Markusheva, T.V., Kramerov, D.A., Ryskov, A.P., Skryabin, K.G., Bayev, A.A. and Georgiev, G.P. (1982) *Nucleic Acids Res.* 10, 7461–7475.
- [10] Bennett, K.L., Hill, R.E., Pietras, D.F., Woodworth-Gutai, M., Kane-Haas, C., Houston, J.M., Heath, J.K. and Hastie, N.D. (1984) *Mol. Cell Biol.* 4, 1561–1571.
- [11] Okada, N., Endoh, H., Sakamoto, K. and Matsumoto, K. (1985) *Proc. Jpn. Acad.* 61, 363–367.
- [12] Endoh, H., Nagahashi, S. and Okada, N. (1990) *Eur. J. Biochem.* 189, 25–31.
- [13] Kido, Y., Aono, M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M. and Okada, N. (1991) *Proc. Natl. Acad. Sci. USA* 88, 2326–2330.
- [14] Ohshima, K., Koishi, R., Matsuo, M. and Okada, N. (1993) *Proc. Natl. Acad. Sci. USA* 90, 6260–6264.

- [15] Ohshima, K. and Okada, N. (1994) *J. Mol. Biol.* 243, 25–37.
- [16] Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) *Nature* 294, 626–631.
- [17] Redston, M.S. and Kern, S.E. (1994) *BioTechniques* 17, 286–288.
- [18] Takahashi, H., Awata, T. and Yasue, H. (1992) *Anim. Genet.* 23, 443–448.
- [19] Kim, J., Martignetti, J.A., Shen, M.R., Brosius, J. and Deininger, P. (1994) *Proc. Natl. Acad. Sci. USA* 91, 3607–3611.
- [20] Cheng, J.F., Printz, R., Callaghan, T., Shuey, D. and Hardison, R.C. (1984) *J. Mol. Biol.* 176, 1–20.
- [21] Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) *Annu. Rev. Biochem.* 55, 631–661.
- [22] Kramerov, D.A., Tillib, S.V., Ryskov, A.P. and Georgiev, G.P. (1985) *Nucleic Acids Res.* 13, 6423–6437.
- [23] Kramerov, D.A., Tillib, S.V., Shumyatsky, G.P. and Georgiev, G.P. (1990) *Nucleic Acids Res.* 18, 4499–4506.