

# Improving specificity of DNA hybridization-based methods

Tatyana Chalaya, Elena Gogvadze, Anton Buzdin\*, Elena Kovalskaya and Eugene D. Sverdlov

Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Moscow 117871, Russia

Received July 21, 2004; Revised and Accepted August 19, 2004

DDBJ/EMBL/GenBank accession nos<sup>†</sup>

## ABSTRACT

**Methods based on DNA reassociation in solution with the subsequent PCR amplification of certain hybrid molecules, such as coincidence cloning and subtractive hybridization, all suffer from a common imperfection: cross-hybridization between various types of paralogous repetitive DNA fragments. Although the situation can be slightly improved by the addition of repeat-specific competitor DNA into the hybridization mixture, the cross-hybridization outcome is a significant number of background chimeric clones in resulting DNA libraries. In order to overcome this challenge, we developed a technique called mispaired DNA rejection (MDR), which utilizes a treatment of resulting reassociated DNA with mismatch-specific nucleases. We examined the MDR efficiency using cross-hybridization of complex, whole genomic mixtures derived from human and chimpanzee genomes, digested with frequent-cutter restriction enzyme. We show here that both single-stranded DNA-specific and mismatched double-stranded DNA-specific nucleases can be used for MDR separately or in combination, reducing the background level from 60 to 4% or lower. The technique presented here is of universal usefulness and can be applied to both cDNA and genomic DNA subtractions of very complex DNA mixtures. MDR is also useful for the genome-wide recovery of highly conserved DNA sequences, as we demonstrate by comparing human and pygmy marmoset genomes.**

## INTRODUCTION

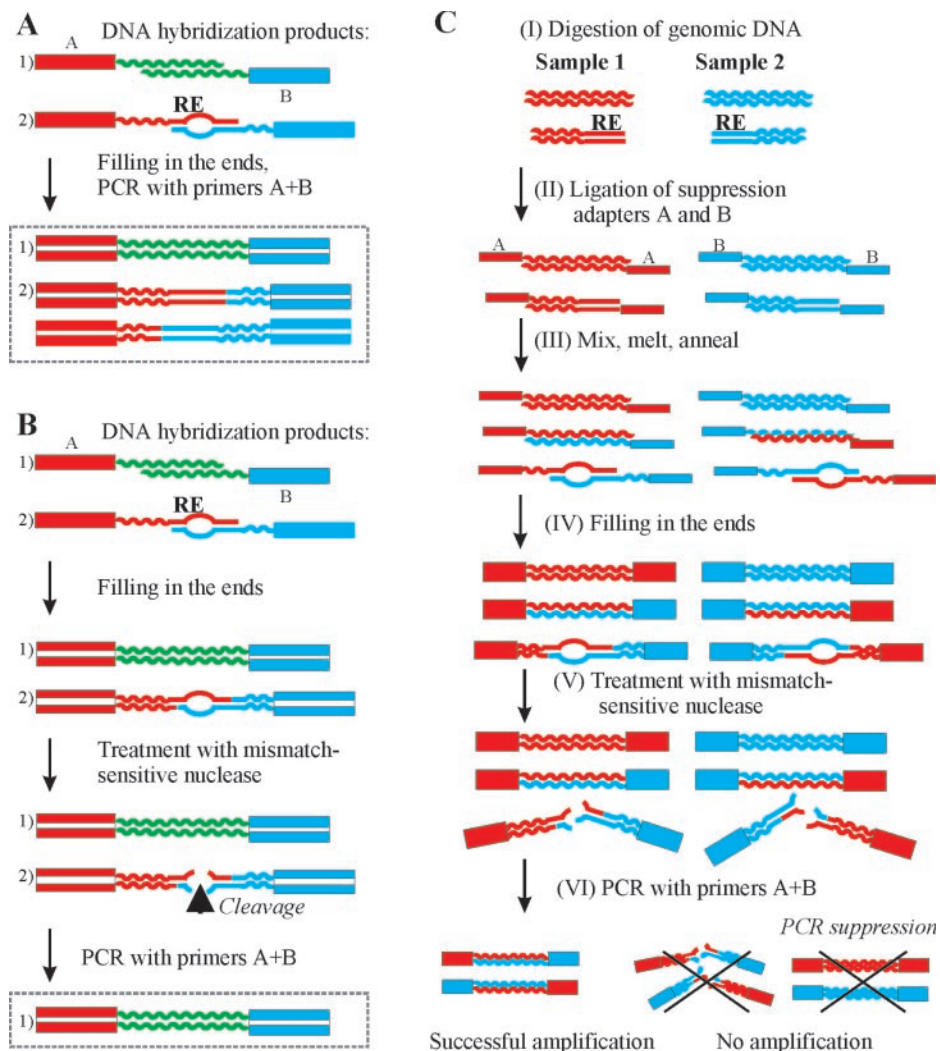
Many popular (approximately 200 PubMed citations per year) experimental techniques for genome and transcriptome analysis, such as coincidence cloning and subtractive hybridization, including representative differential analysis (RDA) (1) and suppressive subtractive hybridization (SSH) (2), are based on DNA hybridization in solution, followed by PCR

amplification of certain hybridized fractions (3,4). Although very useful and informative, these techniques are not free from some imperfections. The well-known disadvantage of complex DNA mixture hybridization is the cross-annealing of repetitive DNA, presenting in reassociating samples (5). This ‘wrong’ annealing causes ‘non-specific’ hybridization of non-orthologous DNA fragments, thus producing chimeric sequences and at the final stage significantly hampering the analysis of the resulting cDNA or genomic libraries (see Figure 1A). Such chimeras may constitute 40–60% of DNA libraries. Although this situation can be slightly improved by the addition of competitor DNA fractions containing genomic repeats into the hybridization mixture (6), the number of background clones in the libraries still remains high (see below). Here we present a new method called mispaired DNA rejection (MDR), which makes it possible to almost completely exclude the chimeric sequences from analyzing DNA subsets (Figure 1B). The technique is based on the observation that an overwhelming majority of cross-hybridizing repetitive elements, although sharing considerable sequence similarity, are not entirely identical to each other. Their DNA heteroduplexes are therefore imperfectly matched, having quite a number of mispaired bases. The latter can form single nucleotide mismatches or even extended single-stranded DNA loop regions. All such structural deviations from the normal, properly paired DNA duplexes can be recognized and cut by certain enzymes, termed here as mismatch-specific nucleases. Mispaired DNA-sensitive nucleases, serving *in vivo* as reparation or viral life cycle machinery units, are now successfully employed by investigators for mutation detection. Such approaches are both simple and rather efficient, such as the TILLING technique for large-scale mutation screening (7), Surveyor mutation detection system (8) and the elegant high-fidelity technique for endonuclease/ligase-based mutation scanning by Huang and others (9). The most commonly used mismatch-specific nucleases are phage T7 endonuclease I (10), T4 endonuclease VII (11), modified bacterial endonuclease V (9), plant CelI and Surveyor nucleases (8,12). Here we demonstrate that these enzymes, cleaving DNA at mispaired base positions, can be used for eliminating chimeric hybrids from DNA hybridization mixtures, thus strongly reducing the number of background chimeric clones from 44–60 to 0–4%.

\*To whom correspondence should be addressed. Tel: +7 095 3306329; Fax: +7 095 3306538; Email: anton@humgen.siocb.ras.ru

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

<sup>†</sup>AY688954–AY688956



**Figure 1.** Schematic representation of MDR rationale, principle and testing system. (A) DNA hybridization and PCR amplification of hybrid duplexes using standard techniques, such as subtractive hybridization and coincidence cloning. Not only exactly matched identical sequences (shown as type 1 fragments), but also a number of background chimeric duplexes (depicted as type 2 DNA), which are usually products of hybridization between REs, are PCR amplified and appear in DNA libraries. (B) The addition of mismatch-sensitive nucleases makes it possible to selectively cleave background duplexes containing imperfectly matched regions and, therefore, to enrich the resulting library in target sequences. (C) The testing system, used by us to investigate the MDR efficiency (see text). The use of MDR reduced the background chimeric clone proportion from 44–60% to 0–4%.

## MATERIALS AND METHODS

### DNA samples

We extracted DNA from four mixed human blood samples and blood samples of chimpanzee *Pan paniscus* and marmoset *Callithrix pigmaea* using a Genomic DNA Purification Kit (Promega).

### DNA preparation for hybridization

The digestion of human, chimpanzee and marmoset genomic DNAs was carried out as follows: 1 µg of genomic DNA was digested with 10 U of frequent-cutter blunt-end-producing restriction endonuclease AluI (Fermentas) at 37°C, for 2 h. DNA was phenol–chloroform extracted, ethanol precipitated and dissolved in 25 µl of sterile water. The suppression adapter ligation was done as described previously (13). We used T4 DNA ligase

(Promega) and standard suppression adapters, A1A2 (5'-gtaatagcactcactatagggcagcgtggtcggccgaggt-3') and B1B2 (5'-cgacgtggactatccatgaacgcacgcagcggcccccgggcaggt-3'). Ligated DNA was purified using the Quiagen PCR product purification kit, ethanol precipitated and dissolved in 5 µl of hybridization buffer (0.5 M NaCl, 50 mM HEPES, pH 8.3, 0.2 mM EDTA).

### DNA hybridization

We mixed 800 ng of each of both the DNA samples assigned for hybridization with of 8 µl of 1× hybridization buffer, denatured at 95°C for 10 min, and hybridized at 65°C or 85°C for 50 h. The final 8 µl mixture was diluted with 72 µl of dilution buffer (50 mM NaCl, 5 mM HEPES, pH 8.3, 0.2 mM EDTA). In some experiments, C<sub>0</sub>tA fraction competitor DNA (Gibco BRL) was added in 100× weight excess to the hybridization mixture.

### Filling in the termini of hybridized DNA

We used AmpliTaq DNA polymerase (1 U per microgram of hybridized DNA) to fill in the ends of DNA duplexes at 72°C for 20 min.

### Hybridized DNA treatment with mismatch-sensitive nucleases

Aliquots of 100 ng of hybridized DNA were digested with 1  $\mu$ l of Surveyor nuclease (Transgenomic) in 20  $\mu$ l of 1 $\times$  buffer supplied by the manufacturer, incubated overnight at 42°C, or treated with 0.1 U of Mung bean nuclease (Promega) at 37°C for 15 min. DNA samples were phenol–chloroform extracted and ethanol precipitated.

### PCR amplification of hybridization products and library construction

DNA samples were dissolved in 100  $\mu$ l of water and 1  $\mu$ l was PCR amplified with 0.2  $\mu$ M primers specific for the suppression adapter set used: A1, 5'-gtaatcagactcactataggc-3', and B1, 5'-cgactggactatccatgaacgca-3'. The PCR conditions were as follows: 95°C for 15 s, 65°C for 10 s, 72°C for 90 s, 15 cycles. To increase the amplification specificity, we used an additional round of nested PCR for 500-fold dissolved products of the latter PCR with 0.2  $\mu$ M primers A2, 5'-agcgtggtcgcggccgaggt-3', and B2, 5'-tcgagcggcccccggg-caggt-3' under the same cycling conditions. The number of nested PCR cycles varied substantially depending on the particular hybridization. The PCR products obtained were cloned in *Escherichia coli* DH5 $\alpha$  using a TA-cloning system (Promega). We sequenced positive clones by the dye termination method using an Applied Biosystems 373 automatic DNA sequencer.

### DNA sequence analysis

We used BLAT search (<http://genome.ucsc.edu/cgi-bin/hgBLAT>) to map clone inserts within human and chimpanzee genomes. Homology searches against GenBank were done using the BLAST Web-server at NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>) (14). The ClustalW program (15) was used for multiple alignments.

### Oligonucleotide primers

Oligonucleotide primers for PCR amplification were synthesized using an ASM-102U DNA synthesizer (Biosan, Novosibirsk, Russia).

### PCR amplification of sequences conserved among human and marmoset genomes

Forty nanograms of blood DNA sample from Old World monkey *C. pigmaea* was PCR amplified using three pairs of 0.2  $\mu$ M unique genomic primers flanking the presumable conserved genomic loci (set 1, forward 5'-cacagcagctgca-taaca-3', reverse 5'-aatgtcctctgtaaggtgg-3'; set 2, forward 5'-cattcattctcagctccacc-3', reverse 5'-cctgcgtcaccttgacca-3'; set 3, forward 5'-agcctgctctgaaccagaatc-3', reverse 5'-cagaag-tctctcagcttagcc-3'). The PCR was conducted at 95°C for 15 s, 56°C for 10 s, 72°C for 1.5 min, 30 cycles. The resulting 193, 283 and 152 bp long PCR products, accordingly, were analyzed on 1.2% agarose gels and sequenced. Marmoset

genome sequences were deposited in GenBank under the accession numbers AY688954–AY688956.

## RESULTS AND DISCUSSION

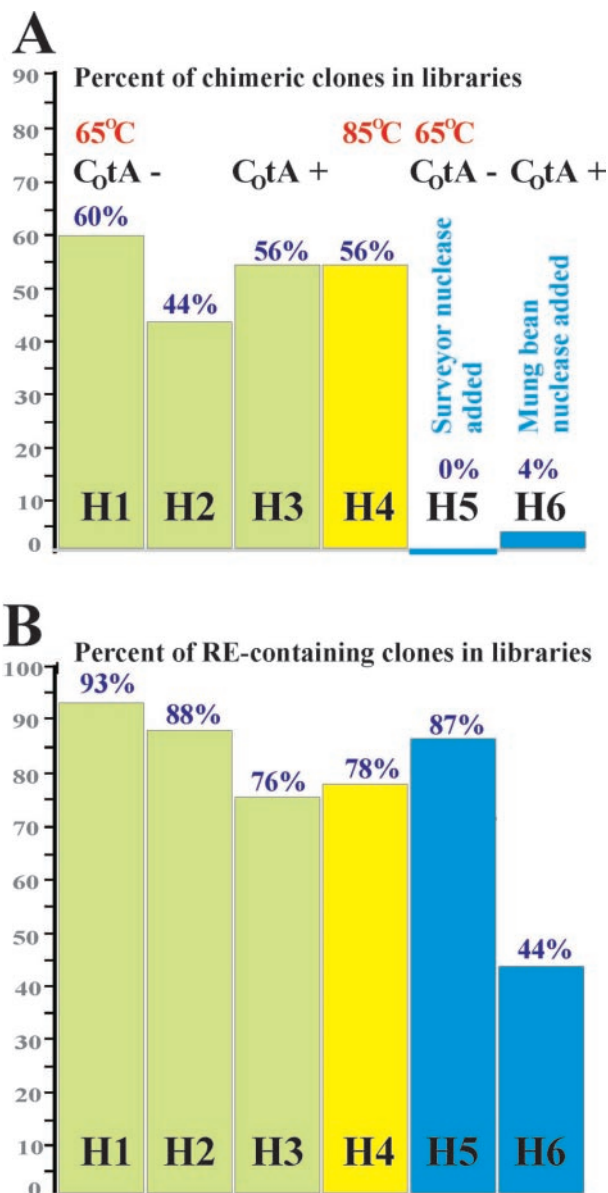
In order to investigate the MDR efficiency, we used a testing system (see Figure 1C) comprising (i) digestion of mammalian genomic DNA with frequent-cutter enzyme, (ii) ligation of different oligonucleotide suppression adapters (required for the so-called 'PCR suppression' effect described below) to the digested DNA, (iii) melting and annealing of two DNA portions harboring different adapters, (iv) filling-in the ends of DNA duplexes with DNA polymerase, (v) treatment with mismatch-sensitive nuclease and (vi) PCR amplification of heteroduplexes, that were not cleaved at the previous stage, with primers specific to both adapters using PCR-suppression effect, which have been described in detail previously (16). Briefly, it includes the ligation of restriction fragments to a panhandle-like structure forming adapter. We used standard adapters (13) that forming after ligation to restriction fragments  $\sim$ 40 bp long GC-rich inverted repeats at their termini. Therefore, such single-stranded DNA fragments contained self-complementary termini capable of forming strong intramolecular stem-loop structures. PCR of the DNA fragments with such termini is therefore suppressed in homoduplexes when primers targeted at the 5' ends of the ligated adapters are used. In contrast, heteroduplex molecules have different termini unable to form stem-loop structures, and can be efficiently PCR amplified further in this system. Nested PCR with primers A2 and B2 increases the specificity of the amplification. This procedure thus ensures exclusive amplification of only the heteroduplex DNA. The control experiments had all of the stages mentioned above, except the step (v), i.e. treatment of hybridized DNA with nucleases. To examine the effect of the competitor DNA, containing genomic repeats, on the hybridization process and on the quality of the resulting libraries, in some experiments we added the quickly reassociating human genomic DNA fraction C<sub>0</sub>tA, which is highly enriched in abundant repetitive sequences and commercially available, to the hybridization mixture, taken in a 100-fold weight excess. We used two mismatched DNA-sensitive nucleases: Surveyor nuclease, which recognizes and cleaves mispaired DNA structures within DNA duplexes, and Mung Bean nuclease, which degrades single-stranded DNA and, therefore, is able to attack loop structures in chimeric hybrids. Mammalian DNAs were chosen for model experiments because they stand among most complex eukaryotic genomes, thus producing very complex hybridization mixtures, far more complex than those of cDNAs. Thus, by solving the challenge of unwanted chimera formation for complex mammalian genome libraries, one may be assured that this obstacle will be surmounted for lower complexity libraries too (such as those of cDNA or of less complex genomes). We carried out six separate hybridizations under the following conditions: (H1), two portions of the fragmented human DNA with different ligated adapters were hybridized at 65°C (T65) without the addition of C<sub>0</sub>tA fraction (C<sub>0</sub>tA–) and without digestion with mismatch-sensitive nucleases (N–); (H2), human and chimpanzee DNAs hybridized at T65, C<sub>0</sub>tA–, N–; (H3), human–human DNA, T65, C<sub>0</sub>tA+, N–; (H4), human–human DNA,



T85, C<sub>0</sub>tA+, N-; (H5), human–chimpanzee DNAs, T65, C<sub>0</sub>tA-, Surveyor nuclease added; (H6), human–human DNA, T65, C<sub>0</sub>tA+, Mung bean nuclease added.

The resulting DNA libraries were cloned into *E.coli*, and 300 inserts (50 random transformants from each library) were sequenced. After removal of low-quality and vector sequences, approximately 40 randomly picked up inserts from each library were analyzed further. We applied the following criteria for the chimera detection: such sequences did not match genomic databases entirely, but their separate 5'- and 3'- terminal fragments matched the databases. Figure 2 depicts the results of the analysis of six of our DNA libraries. It is clear that the addition of C<sub>0</sub>tA fraction and the hybridization temperature increase from 65 to 85°C has essentially no effect on the number of chimeric clones, in contrast to the addition of mismatch sensitive nucleases. Both Mung bean and Surveyor nucleases display the strong effect on the chimera formation, greatly reducing their number from 44–60% to 0–4% of clones. Many of the inserts sequenced contained genomic repetitive elements (REs), which is not surprising, as they constitute ~50% of mammalian DNA (17). Such RE sequences even if they correspond to correct genomic loci may match different positions on the genomic DNA, thus making their exact mapping problematic. Therefore it is desirable to minimize the portion of such kind of sequences in the libraries. We found that the proportion of repetitive elements containing inserts differed considerably among the libraries: C<sub>0</sub>tA- libraries contained high number of REs independently on the addition of the nuclease (87–93% of the sequenced clones), C<sub>0</sub>tA+/N- libraries—slightly smaller proportion of REs (76–78%), and finally C<sub>0</sub>tA+/N+ library (H6, Mung bean nuclease added) had only 44% of RE-containing inserts. These data show that the best results in the library construction can be achieved with both (i) addition of RE-containing competitor DNA into hybridization mixture and (ii) treatment of hybridized DNA with mismatch-sensitive nucleases.

We tried to address the question whether MDR technique can be applied to interspecies DNA hybridizations. To this end, in two hybridization experiments (H2 and H5) we hybridized human and chimpanzee DNA. Human and chimpanzee genomes are closely related, displaying ~98% sequence identity (17). The results suggest that MDR reduces the number of chimeric sequences from 44% even in the absence of detected chimeras. All sequenced inserts from the Surveyor-nuclease-treated library (H5) did contain sequences highly conservative between the two genomes (average identity of 98.3%). Some inserts contained regions, evolutionarily conserved among the sequenced mammalian genomes—those of human, chimpanzee, mouse and rat. This observation suggests that MDR could be also applied for the recovery of evolutionary conserved sequences between different genomes. To investigate this, we performed another interspecies hybridization (H7), between human and new world monkey *C.pygmaea* genomes, at 65°C, followed by the subsequent digestion with Surveyor nuclease. The pygmy marmoset *C.pygmaea* genome is more divergent from human than chimpanzee DNA [human and new world monkey ancestor lineages separated roughly 45 million years ago (18)], thus showing ~20% DNA sequence divergence [assuming the average nucleotide substitution rate in primate genomes to be  $2.2 \times 10^{-9}$  bases per million years (19)]. Fifty



**Figure 2.** Comparison of six DNA libraries, created under different hybridization conditions with or without the use of MDR. H1, human–human DNA hybridization at 65°C (T65), without competitor C<sub>0</sub>tA DNA (C<sub>0</sub>tA-), no mismatch-sensitive nucleases added (N-); H2, human–chimpanzee DNA, T65, C<sub>0</sub>tA-, N-; H3, human–human DNA, T65, C<sub>0</sub>tA added (C<sub>0</sub>tA+), N-; H4, human–human DNA, T85, C<sub>0</sub>tA+, N-; H5, human–chimpanzee DNA, T65, C<sub>0</sub>tA-, Surveyor nuclease added; H6, human–human DNA, T65, C<sub>0</sub>tA+, Mung bean nuclease added. (A) Colored column height reflects the proportion of chimeric clones in analyzed libraries. The number of chimeric sequences is dramatically decreased in libraries, treated with mismatch sensitive nucleases. (B) Column height reflects the proportion of clone inserts, containing RE sequences. It can be seen that the addition of C<sub>0</sub>tA competitor DNA alone slightly decreases the number of RE-containing clones, but the combination of both C<sub>0</sub>tA addition and nuclease digestion yields the best result in library construction.

clones from the resulting library were sequenced, and 45 good-quality insert sequences were further analyzed. Seventy-one percent of inserts represented moderately (~14%) divergent genomic repeats, which are believed to be present in both human and marmoset genomes, and the remaining 29% (13 inserts) were unique sequences (see Table 1).

**Table 1.** Human–marmoset hybridization library clone analysis

N <sup>a</sup>	Clone description <sup>b</sup>	Genome location <sup>c</sup>	Chimp. <sup>d</sup>	Marm. <sup>e</sup>	Rat <sup>f</sup>	Mouse <sup>g</sup>
1	Exon of the gene <i>GP5</i>	3; 195599709-860	99.4	94.0	ND	ND
2	At 4 kb downstream the gene <i>IL13</i> , 8 kb upstream the gene <i>IL4</i>	5; 132027632-918	99.4	95.2	ND	ND
3	Intron of the gene <i>BC021928</i>	18; 38162564-757	100	95.3	ND	ND
4	Intron of the gene <i>CHRM3</i>	1; 236286986-7140	97.5	ND	88.3	89.5
5	Intron of the gene <i>TMF1</i>	3; 69183331-485	95.5	ND	79.5	77.9
6	Intron of the gene <i>EXT1</i>	8; 119094950-5080	99.3	ND	85.9	89.2
7	Intron of the gene <i>BSN</i>	3; 49653232-316	98.9	ND	ND	88.7
8	Intron of the EST BG992862	14; 18572238-350	98.3	ND	89.7	90.1
9	Exon of the gene <i>RYR3</i>	15; 31927735-947	99.6	ND	89.9	92.5
10	Exon of the gene <i>DVL2</i>	17; 7071405-515	100	ND	91.8	91.8
11	Exon of the gene <i>CD22</i>	19; 40527744-847	100	ND	90.0	94.8
12	Exon of the gene <i>GALNT12</i>	9; 98673703-802	99.0	ND	89.1	90.3
13	Exon of the gene <i>C6orf85</i>	6; 3214634-781	99.4	ND	95.9	97.3
14	RE: LTR56, div. 14%	Unknown	?	?	?	?
15	RE: L1P, div. 18%	Unknown	?	?	?	?
16	RE: L1P1, div. 6%	Unknown	?	?	?	?
17	RE: L1P, div. 18%	Unknown	?	?	?	?
18	RE: FRAM, div. 19%	Unknown	?	?	?	?
19	RE: AluSx, div. 20%	Unknown	?	?	?	?
20	RE: AluSg, div. 13%	Unknown	?	?	?	?
21	RE: AluSg, div. 9%	Unknown	?	?	?	?
22	RE: AluSg, div. 9%	Unknown	?	?	?	?
23	RE: AluSg, div. 14%	Unknown	?	?	?	?
24	RE: AluSg, div. 8%	Unknown	?	?	?	?
25	RE: AluSg, div. 18%	Unknown	?	?	?	?
26	RE: AluSg, div. 13%	Unknown	?	?	?	?
27	RE: AluSg, div. 13%	Unknown	?	?	?	?
28	RE: AluSg, div. 10%	Unknown	?	?	?	?
29	RE: AluSg, div. 22%	Unknown	?	?	?	?
30	RE: AluSc, div. 8%	Unknown	?	?	?	?
31	RE: AluSc, div. 8%	Unknown	?	?	?	?
32	RE: AluSc, div. 8%	Unknown	?	?	?	?
33	RE: AluSc, div. 8%	Unknown	?	?	?	?
34	RE: AluSc, div. 8%	Unknown	?	?	?	?
35	RE: AluSp, div. 15%	Unknown	?	?	?	?
36	RE: AluSp, div. 13%	Unknown	?	?	?	?
37	RE: AluSp, div. 18%	Unknown	?	?	?	?
38	RE: AluJ, div. 14%	Unknown	?	?	?	?
39	RE: AluJo, div. 12%	Unknown	?	?	?	?
40	RE: AluJo, div. 12%	Unknown	?	?	?	?
41	RE: AluJo, div. 15%	Unknown	?	?	?	?
42	RE: Alu, div. 13%	Unknown	?	?	?	?
43	RE: Alu, div. 21%	Unknown	?	?	?	?
44	RE: Alu, div. 15%	Unknown	?	?	?	?
45	RE: Alu, div. 17%	Unknown	?	?	?	?

ND, the absence of data; ?, the impossibility of RE sequence assignment to certain positions within the genomes and, therefore, the lack of information about sequence conservation among the mammalian DNAs.

<sup>a</sup>Clone number.

<sup>b</sup>Clone description column: human gene neighborhood of unique sequences (clones 1–13), or assignment of repetitive sequences (clones 14–45) to certain RE groups.

<sup>c</sup>Sequence exact location (chromosome number; coordinates) in the human genome, found with UCSC genome browser using May 2004 human genome assembly. Sequence identity to the corresponding locus of the <sup>d</sup>chimpanzee, <sup>e</sup>marmoset (*C. pygmaea*), <sup>f</sup>rat and <sup>g</sup>mouse genomes.

Ten such unique sequences were conserved among human, chimpanzee, mouse and rat genomes (Table 1, clones 4–13), three other inserts were conserved among human and chimpanzee DNAs. In order to confirm the high conservation value of these sequences among human and marmoset, we PCR-amplified and sequenced the corresponding loci from *C. pygmaea* genome for three such individual sequences (Table 1, clones 1–3). Indeed, all sequenced marmoset loci displayed significant DNA conservation and similarity to the corresponding human loci with the average sequence identity of 95%, thus showing about 4-fold slower mutation rate for these loci than neutral base substitution rate.

The results presented above strongly suggest that the MDR technique may provide a useful tool for the refinement of various DNA libraries obtained with the use of DNA reassociation, including subtractive and normalized genomic and cDNA libraries. Although Surveyor nuclease was somewhat more efficient than Mung Bean nuclease for the particular application of MDR approach described here, for other tasks the latter may also be very useful. In particular, treatment with Mung Bean nuclease may be helpful for cDNA hybridization, where a lot of non-specific cross-annealing occurs between homologous gene transcript copies. Thus, a mixture of both nucleases can be offered for routine use. The technique may also considerably improve genome-wide recovery of

evolutionarily conserved sequences. The experimental techniques for identification of evolutionarily conserved regions are required for the comparison of sequenced and/or unsequenced genomes, thus making MDR a universal method. In all cases the technique application will hopefully diminish the confusion caused by cross hybridization of closely related but different paralogous sequences.

## ACKNOWLEDGEMENTS

The study was supported by the Physico-Chemical Biological Program of the Russian Academy of Sciences, and by the grants 02-04-48614-a and 04-04-48564-a of the Russian Foundation for Basic Research.

## REFERENCES

1. Lisitsyn, N. and Wigler, M. (1995) Representational difference analysis in detection of genetic lesions in cancer. *Methods Enzymol.*, **291**–304.
2. Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. *et al.* (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl Acad. Sci. USA*, **93**, 6025–6030.
3. Sasaki, H., Nomura, S., Akiyama, N., Takahashi, A., Sugimura, T., Oishi, M. and Terada, M. (1994) Highly efficient method for obtaining a subtracted genomic DNA library by the modified in-gel competitive reassociation method. *Cancer Res.*, **54**, 5821–5823.
4. Nagayama, K., Enomoto, N., Miyasaka, Y., Kurosaki, M., Chen, C.H., Sakamoto, N., Nakagawa, M., Sato, C., Tazawa, J., Ikeda, T. *et al.* (2001) Overexpression of interferon gamma-inducible protein 10 in the liver of patients with type I autoimmune hepatitis identified by suppression subtractive hybridization. *Am. J. Gastroenterol.*, **96**, 2211–2217.
5. Hames, B.D. and Higgins, S.J. (eds) (1985) *Nucleic Acid Hybridization. A Practical Approach*. IRL Press, Oxford, Washington, DC.
6. Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. 3rd edn. Cold Spring Harbour Laboratory Press, Cold Spring Harbor, NY.
7. Till, B.J., Burtner, C., Comai, L. and Henikoff, S. (2004) Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Res.*, **32**, 2632–2641.
8. Qiu, P., Shandilya, H., D'Alessio, J.M., O'Connor, K., Durocher, J. and Gerard, G.F. (2004) Mutation detection using Surveyor nuclease. *Biotechniques*, **36**, 702–707.
9. Huang, J., Kirk, B., Favis, R., Soussi, T., Paty, P., Cao, W. and Barany, F. (2002) An endonuclease/ligase based mutation scanning method especially suited for analysis of neoplastic tissue. *Oncogene*, **21**, 1909–1921.
10. Babon, J.J., McKenzie, M. and Cotton, R.G. (2003) The use of resolvases T4 endonuclease VII and T7 endonuclease I in mutation detection. *Mol. Biotechnol.*, **23**, 73–81.
11. Mikhailov, V.S. and Rohrmann, G.F. (2002) Binding of the baculovirus very late expression factor 1 (VLF-1) to different DNA structures. *BMC Mol. Biol.*, **3**, 14.
12. Kulinski, J., Besack, D., Oleykowski, C.A., Godwin, A.K. and Yeung, A.T. (2000) CEL I enzymatic mutation detection assay. *Biotechniques*, **29**, 44–46, 48.
13. Lavrentieva, I., Broude, N.E., Lebedev, Y., Gottesman, I.I., Lukyanov, S.A., Smith, C.L. and Sverdlov, E.D. (1999) High polymorphism level of genomic sequences flanking insertion sites of human endogenous retroviral long terminal repeats. *FEBS Lett.*, **443**, 341–347.
14. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
15. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
16. Gurskaya, N.G., Diatchenko, L., Chenchik, A., Siebert, P.D., Khaspekov, G.L., Lukyanov, K.A., Vagner, L.L., Ermolaeva, O.D., Lukyanov, S.A. and Sverdlov, E.D. (1996) Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemagglutinin and phorbol 12-myristate 13-acetate. *Anal. Biochem.*, **240**, 90–97.
17. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
18. Sverdlov, E.D. (2000) Retroviruses and primate evolution. *Bioessays*, **22**, 161–171.
19. Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.