RESEARCH

Characterization of *Alu* Repeats That Are Associated with Trinucleotide and Tetranucleotide Repeat Microsatellites

Chandri N. Yandava,^{1,3,6,8} Julie M. Gastier^{1,3} Jacqueline C. Pulido,^{1,2,7}
Tom Brody,^{1,3,7} Val Sheffield,^{3,4} Jeffrey Murray,^{3,4} Kenneth Buetow,^{3,5}
and Geoffrey M. Duyk^{1,2,7}

¹Department of Genetics, Harvard Medical School, and ²Howard Hughes Medical Institute, Boston, Massachusetts 02115; ³Cooperative Human Linkage Center and ⁴Department of Pediatrics, University of Iowa, Iowa City, Iowa 52245; ⁵Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

The association of subclasses of *Alu* repetitive elements with various classes of trinucleotide and tetranucleotide microsatellites was characterized as a first step toward advancing our understanding of the evolution of microsatellite repeats. In addition, information regarding the association of specific classes of microsatellites with families of *Alu* elements was used to facilitate the development of genetic markers. Sequences containing *Alu* repeats were eliminated because unique primers could not be designed. Various classes of microsatellites are associated with different classes of *Alu* repeats. Very abundant and poly(A)-rich microsatellite classes (ATA, AATA) are frequently associated with an evolutionarily older subclass of *Alu* repeats, *AluSx*, whereas most of GATA and CA microsatellites are associated with a recent *Alu* subfamily, *AluY*. Our observations support all three possible mechanisms for the association of *Alu* repeats to microsatellites. Primers designed using a set of sequences from a particular microsatellite class showed higher homology with more sequences of that class than probes designed for other classes. We developed an efficient method of prescreening GGAA and ATA microsatellite clones for *Alu* repeats with probes designed in this study. We also showed that *Alu* probes labeled in a single reaction (multiplex labeling) could be used efficiently for prescreening of GGAA clones. Sequencing of these prescreened GGAA microsatellites revealed only 5% *Alu* repeats. Prescreening with primers designed for ATA microsatellite class resulted in the reduction of the loss of markers from ~50% to 10%. The new *Alu* probes that were designed have also proved to be useful in *Alu*-*Alu* fingerprinting.

Genetic maps of mouse and human have been constructed based on highly polymorphic microsatellite markers (Dietrich et al. 1994; Gyapay et al. 1994; Murray et al. 1994). Microsatellite markers are based on the amplification of sequences containing short tandem repeats of 2–7 nucleotides. These types of repeats were found to vary in length in different individuals, making them highly polymorphic (Weber 1990; Hearne et al. 1992). Dinucleotide repeat class, $[CA]_n$ is present abundantly in mammalian genomes (Stallings et al. 1991). Thousands of markers were developed from known sequences containing these repeats from databases and from screening

small insert human genomic DNA libraries (Litt and Luty 1989; Tautz 1989; Weber and May 1989; Weissenbach et al. 1992). It was reported that amplified products of dinucleotide repeat markers showed additional bands, when these products were run on polyacrylamide gels (Economou et al. 1990; Beckman and Weber 1992). Although trinucleotide and tetranucleotide repeat microsatellites were observed to be less frequent than $[CA]_n$ repeats, their amplification products resolve into readily interpretable products. (Edwards et al. 1991, 1992; Puers et al. 1993; LeBlank-Straceski et al. 1994).

The goal of Cooperative Human Linkage Center (CHLC) is to develop high-resolution integrated human genetic maps based on tri- and tetranucleotide repeat microsatellite markers. We have been developing markers based on the most abundant tri- and tetranucleotide repeats of these microsatellites (Gastier et al. 1995; Sheffield et al. 1995). Markers

Present addresses: ⁶Department of Medicine, Harvard Medical School and Pulmonary Division, Brigham and Women's Hospital, Boston, Massachusetts 02115; ⁷Millenium Pharmaceuticals, Cambridge, Massachusetts 02139.

⁸Corresponding author.

E-MAIL Yandava@rascal.med.harvard.edu; FAX (617) 232-4623.

have been developed from microsatellite-enriched small insert genomic libraries (Pulido and Duyk 1994). It was shown that some of the highly frequent tri- and tetranucleotide repeat microsatellites (ATA, AATA, GGAA, and GAAA; see Table 2 for alphabetically minimal nomenclature) were generally associated with *Alu* repeats (Economou et al. 1990; Sinnett et al. 1990; Zuliani and Hobbs 1990; Beckmann and Weber 1992; Jurka and Pethiyagoda 1995).

Alu repeats can be classified into subfamilies based on DNA sequence divergence (Jurka and Smith 1988; Batzer et al. 1990; Jurka and Milosavljevic 1991; Jurka 1993). Each subfamily was thought to arise from a distinct founder sequence. It was suggested from the divergence of their sequences that these subfamilies appeared at different evolutionary times. *Alu* repeats were classified into nine subfamilies, namely *AluJ*, *AluSx*, *AluSq*, *AluSp*, *AluSc*, *AluY*, *AluYa5*, *AluYa8*, and *AluYb8*, according to the standardized nomenclature (Batzer et al. 1996). Some of the microsatellite markers were reported to be associated with *Alu* repeats (Zuliani and Hobbs 1990; Jurka and Pethiyagoda 1995). Because microsatellite repeats are present close to *Alu* repeats, one of the designed primers may lie within the *Alu* repeat. Primers developed from such sequences could result in weak amplification and/or high background. In our strategy of primer design, we generally avoid microsatellites that are associated with *Alu* repeats. This can be done only after sequencing the clones. To reduce sequencing costs, we show here that we can prescreen the microsatellite clones with *Alu* probes. The knowledge of the *Alu* families that are associated with these microsatellites will help in understanding their evolution and will also be useful in designing better prescreening probes. In the present study we report the subclass distribution of the microsatellite-associated *Alu* repeats. We report differences and characteristics of microsatellites in their association with *Alu* repeats. We have used the knowledge of this association of *Alu* repeats to design microsatellite class-specific *Alu* probes to prescreen the ATA and GGAA classes of repeats.

RESULTS

Classification of *Alu* Repeats That Are Associated with Single Tandem Repeats

Many of the simple tandem repeat (STR) classes were found to be associated with *Alu* repeats. We

have analyzed *Alu*-associated CA repeats as well as tri- and tetranucleotide repeat classes for the type of *Alu* subclasses with which they are associated. *Alu* repeats were classified using the Pythia server (Jurka and Milosavljevic 1991). When there was not enough information for classification of an *Alu*, it was not grouped into any subclass. For example, some of the sequences could not be classified either as *AluJ* or *AluS* class and therefore were not included. An evolutionarily recent subclass, *AluYb8*, is very rare among all of the STRs (~2%–3%; data not shown). Similarly, the *AluSp* subclass is also very rare. As the number of sequences studied for some of the repeat classes (GGT, GAA, GTAT) was small, these classes were not further analyzed. For further comparisons, subclasses were grouped into major classes. The distribution of these classes is shown in Table 1, where different classes of microsatellite repeat differ in the type of *Alu* class with which they are associated. This difference is very significant ($\chi^2 = 71.271$, $df = 24$, $P < 0.0001$). It is interesting to note that *AluSx* is increased in poly(A)-containing classes and the most frequently *Alu*-associated classes such as AATA (cell $\chi^2 = 6.81$), ATA (cell $\chi^2 = 11.03$), whereas this *Alu* subclass is decreased in less *Alu*-associated CA (cell $\chi^2 = 10.19$) and GATA (cell $\chi^2 = 4.62$), particularly in CA. In contrast, *AluY* is increased in CA (cell $\chi^2 = 7.69$) and GATA (cell $\chi^2 = 7.18$), decreased in ATA (cell $\chi^2 = 5.09$), AATA (cell $\chi^2 = 4.05$), and GAAA (cell $\chi^2 = 3.61$), and is not changed in G-rich classes (GGA and GGAA). GAAA differs in its association with the oldest *Alu* class, *AluJ* (cell $\chi^2 = 4.53$), which is increased. This analysis shows that repeat classes share certain characteristics in their association with *Alu* classes.

Primer Design

To evaluate the efficacy of the designed primers, sequences of these primers were compared with *Alu*-associated GGAA repeat microsatellite sequences. From 1300 GGAA repeat sequences, 103 *Alu*-containing sequences were used in the construction of a local database (library) using DATASET program of the Genetics Computer Group (GCG) package. This library was searched by FASTA with *Alu* primer sequences. The homology of these primers with the sequences from the local library is presented in Table 2. We used a criterion of >85% homology for the selection of our primers. As *AluATA2* overlaps with 22- of 23-bp complementary strand of *AluGAAA3*, it was not analyzed for the homology. *AluGGAA4* primer, which was designed by using se-

Table 1. Major **Alu** Family Classification of Microsatellite Repeats

Repeat class ^a	<i>AluI</i>		<i>AluSb</i>		<i>AluSc</i>		<i>AluSpq^b</i>		<i>AluSx</i>		Total
	obs.	exp.	obs.	exp.	obs.	exp.	obs.	exp.	obs.	exp.	
ATA											
(AAT)	9	9.9	4	11.7	4	2.9	8	8.6	18	9.8	43
AATA											
(AAAT)	11	12.5	7	14.7	2	3.7	10	10.8	24	12.3	54
CA											
(AC)	11	12.2	25	14.5	5	3.6	11	10.6	1	12.1	53
GAAA											
(AAAG)	25	16.4	11	19.4	4	4.8	16	14.2	15	16.2	71
GATA											
(AGAT)	25	24.9	44	29.5	6	7.3	19	21.7	14	24.7	108
GGA											
(AGG)	4	5.1	6	6.0	2	1.5	6	4.4	4	5.0	22
GGAA											
(AAGG)	14	18.0	20	21.3	6	5.3	16	15.6	22	17.8	78
Total	99		117		29		86		98		429

(obs.) Observed number of sequences; (Exp.) Expected number of sequences. $\chi^2 = 71.271$; df = 24; $P < 0.001$.^aRepeat names in parentheses correspond to the alphabetically minimal nomenclature proposed by Jin et al. (1994).^bIncludes *AluSp* and *AluSq*.

quences from the GGAA repeat class, shows >90% homology with 31.1% of the sequences. Primers from ATCC, Aluconatcc1a and Aluconatcc2, do not show >85% homology with any of the sequences. Another commercially available probe from Lifecodes, Inc., Alucon has >85% homology with only ~9% of the sequences. Other *Alu* primers designed for the prescreening of *Alu* repeats from other classes of microsatellites, *AluGAAA3* and *AluATA1*, show considerable homology with sequences

from the local library. Therefore, *AluGGGAA4*, *AluGAAA3*, and *AluATA1* primers were used as probes in our further *Alu* prescreening efforts.

Screening with a Multiplex-Labeled Probe

Screening with a single mixed probe will reduce the amount of work considerably when compared to screening with each probe separately. To test the feasibility of screening with a mixed probe, four duplicate filters of 450 GGAA repeat positive colonies were replicated. Each filter was probed with *AluGGGAA4*, *AluATA1*, *AluGAAA3*, or a mixed probe, which was labeled with all of the primers together. Comparative results for a single plate are shown in Figure 1. Some of the colonies were positive for more than one probe. The number of positive colonies is similar when screening was carried with either a mixed probe or individual probes. Therefore, further prescreenings were carried out with mixed probe.

The aim in designing new *Alu* primers was to reduce the number of *Alu*-containing clones to be se-

Table 2. Homology of **Alu** probes to GGAA microsatellite sequences with **Alu** repeats

<i>Alu</i> probe	>90% Homology		>85% Homology	
	number	%	number	%
<i>AluGGGAA4</i>	32	31.1	38	36.9
<i>AluGAAA3</i>	22	21.4	34	33.0
<i>AluATA1</i>	21	20.4	24	23.3
<i>AluGAAA1</i>	5	4.9	20	19.4
<i>AluGAAA4</i>	9	8.7	16	15.5
<i>AluCon</i>	5	4.9	9	8.7
<i>AluGAAA2</i>	2	1.9	5	4.9
<i>AluConatcc1a</i>	0	0.0	0	0.0
<i>AluConatcc2</i>	0	0.0	0	0.0

MICROSATELLITE ASSOCIATION WITH *ALU* REPEATS

Figure 1 Prescreening of GGAA microsatellite-positive clones for *Alu* repeats. Four duplicate filters were replicated and probed with *Aluaat1* (A), *Alugaaa3* (B), *Aluggaa4* (C), and a multiplex probe containing three *Alu* oligonucleotides (*Aluaat1*, *Alugaaa3*, and *Aluggaa4*) (D).

quenced. A total of 21,510 colonies was probed with [GGAA]₁₀ probe. About 10.1% of these colonies is positive for GGAA repeats. Positive colonies were picked into 96-well microtiter plates and screened for GGAA and *Alu* repeats. From the secondary screening 1984 clones were found to be positive with [GGAA]₁₀. Of the 1984 positive clones, ~40.2% (799 clones) were *Alu* positive when they were probed with the mixed probe. As ~50% of GGAA microsatellites are generally associated with *Alu* repeats, >80% of *Alu*-positive GGAA microsatellite clones could be prescreened by this method. When 97 *Alu* prescreened clones were sequenced, only 5 of them contained *Alu* repeats in them.

Prescreening ATA Microsatellites for *Alu* Repeats

Similarly, to estimate the success of prescreening with ATA microsatellites, we studied the number of sequences where no primers could be designed because of a flanking *Alu* repeat. The results are shown in Figure 2. The loss of microsatellite clones is expressed as the percent of clones lost because primers could not be designed for genotyping by PCR. The loss was remarkably reduced when the screening was carried out with the primers specifically designed for this class used in addition to the commercial probe.

Alu-*Alu* PCR

Primers designed for *Alu* sequences are often useful for identifying overlapping sequences within yeast artificial chromosomes (YACs) or other large-insert clones. In addition, *Alu*-*Alu* PCR is used frequently for generating probes from YACs for fluorescence in situ hybridization (FISH). To determine whether the *Alu* probes that were designed for prescreening clones could be useful for *Alu*-*Alu* PCR or fingerprinting YACs, we used the primers to amplify several YACs from the chromosome 6q23 region. Ideal primers for *Alu*-*Alu* PCR are located near the end of the *Alu* consensus sequence and are directed away from the *Alu* element to amplify as much non-*Alu* DNA as possible. Of the seven

new *Alu* probes that were designed (*AluATA1*, *AluATA2*, *AluGAAA1*, *AluGAAA2*, *AluGAAA3*, *AluGAAA4*, and *AluGGAA4*), only *AluATA1*, *AluGAAA3*, and *AluGGAA4* were located at the end of the *Alu* consensus and point outward (data not shown). However, all eight of the primers gave distinct bands when used to amplify YAC DNA, suggesting that all of them could be used for fingerprinting YACs. An example is shown in Figure 3.

DISCUSSION

We have been constructing high-resolution human genetic maps at CHLC based on tri- and tetranucleotide repeat markers from small insert genomic libraries. We have optimized our marker selection based on a highly efficient, positive selection (Pulido and Duyk 1994), and we have concentrated on efficient screening of these markers to minimize our sequencing efforts. We have observed from our sequencing data that some of the tri- and tetranucleotide classes are closely associated with *Alu* repeats. For example, the association with *Alu* repeats for GGAA, ATA, and GAAA classes ranged from 50%–85%, whereas the association is small for GATA, <20%. Similar frequencies were observed from our databank search (C.N. Yandava and G.M.

Figure 2 Effectiveness of prescreening of ATA microsatellite-positive clones. Prescreening was carried out with Quick-Lite (*Alucon*) or three *Alu* probes (*Alucon*, *AluATA1*, and *AluATA2*). Numbers (n) of sequences tested at each point are indicated for each prescreening point.

Duyk, unpubl.) and previous studies (Beckmann and Weber 1992; Jurka and Pethiyagoda 1995).

Alu repeats are classified into subfamilies based on DNA sequence divergence and were thought to arise from a distinct founder sequence. It was suggested from the divergence of their sequences that these subfamilies appeared at different evolutionary times (Britten et al. 1988; Jurka and Smith 1988; Quentin 1992). As a variation in the degree of association with *Alu* repeats was observed with different microsatellite repeats, we classified these microsatellites into different *Alu* subfamilies to gain a better understanding of the evolution of the microsatellites. It was observed that microsatellite repeats showed statistically significant differences in their association with the type of *Alu* subfamilies.

According to Jurka and Milosavljevic (1991), *AluY* (Batzer et al. 1996) was the youngest and it was divided further into *AluYa5* (Batzer et al. 1996), which was also known as PV (predicted variant) (Matera et al. 1990) or HS (human specific) (Batzer et al. 1990), and *AluYb8* (Batzer et al. 1996), which was previously known as Sb2 (Jurka 1993). In a recent study, Kapitonov and Jurka (1996) estimated the ages of *Alu* subfamilies. *AluJ* was further divided into *AluJ0* and *AluJb*. The age of these subfamilies was estimated to be ~81 million years (Myr). The estimated ages of *AluSx* and *AluY* were 37 and 19 Myr, respectively. In the present study, we demonstrated a difference in *Alu* subfamily association among microsatellite families. This association indicates the differences in age association with *Alu* subfamilies. GAAA repeat microsatellites showed an association with the oldest class, *AluJ*. We have ob-

served that the frequently *Alu*-associated and poly-(A)-rich ATA and AATA families (Economou et al. 1990; Sinnet et al. 1990; Zuliani and Hobbs 1990; Beckmann and Weber 1992; Jurka and Pethiyagoda 1995) showed marked decrease in the expected number of *AluY* family sequences. On the other hand, less frequently *Alu*-associated CA (Beckmann and Weber 1992; Jurka and Pethiyagoda 1995) and GATA (Jurka and Pethiyagoda 1995; C.N. Yandava et al., unpubl.) showed an increased association with this *Alu* subfamily. The reverse was observed with the oldest of *AluS* subfamily, *AluSx*. These results might suggest different origins and ages of these microsatellite classes. Recently, Arcot et al. (1995) proposed three possible mechanisms for the association of microsatellites and *Alu* elements. According to the first mechanism proposed, an *Alu* element integrates in a pre-existing microsatellite and the resulting flanking direct repeats containing the microsatellite could expand or contract over time. In the second mechanism, mutations are introduced during the process of reverse transcription and these mutations are used in further expansion of microsatellites. In the third case, the accumulation of random mutations in the *Alu* element oligo(dA) tail and expansion results in the genesis of microsatellites.

Our results suggest the possibility of all three mechanisms for the genesis of microsatellites. The major microsatellites found in plants (Lagercrantz et al. 1993), poly(A)-containing microsatellites, are more abundant and are greatly associated with *Alu* elements (Jurka and Pethiyagoda 1995). These facts suggest an insertion of *Alu* elements adjacent to

Figure 3 *Alu-Alu* PCR of YAC DNAs using *AluATA1* primer. Ten YACs (lanes are labeled with different YACs tested) from the chromosome 6q23 region were amplified using *AluATA1* and PDJ34 (a primer commonly used for *Alu-Alu* PCR). Common bands suggest that the YACs overlap (e.g., 917C6 and 957A3).

these repeats. The most recent and less *Alu*-associated microsatellites such as CA and GATA might be generated by the second and third mechanisms proposed by Arcot et al. (1995).

Although highly *Alu* associated tri- and tetranucleotide repeat families (ATA, AATA, GGAA, and GAAA) were found to be more frequent than the less *Alu*-associated repeat classes (e.g., GATA), the genotyping of such microsatellites is very hard for two reasons. First, the placement of one of the PCR primers within the highly repeated *Alu* element might cause a weak amplification and/or high background (Economou et al. 1990; Weber 1990). Secondly, if the primers were designed to exclude the *Alu* repeat the expected size of the PCR product would be larger than the usual fragment that could be amplified by designing one of the primers within the *Alu* repeat. The increase in the fragment size might cause problems in the resolution of the amplified products. Studies with tetranucleotide polymorphism at the human β -actin-related pseudogene 2 (ACTBP 2) showed differences in the information obtained with different product sizes. When primers designed to amplify a product of 291 bp were used, 21 alleles were observed (Polymeropoulos et al. 1991). In contrast, primers designed to amplify a product of 519 bp resulted in the detection of only five alleles (Warne et al. 1991). To avoid these problems, we routinely do not design primers for sequences with *Alu* repeats. As mentioned earlier, microsatellites associated with *Alu* repeats are more frequent; therefore, we prescreen our clones for *Alu* repeats in our marker development strategy.

Our prescreening of ATA-positive clones for *Alu* repeats with a commercial *Alu* probe (Alucon) was not successful in eliminating the majority of the *Alu*-positive clones. We therefore designed primers for each of the classes (ATA, GGAA, and GAAA) and checked the homology of primers with GGAA repeat sequences. The majority of the sequences have >90% homology with the primer designed for that class (*Alu*GGAA4). As GAAA and ATA repeat classes share some features in their association with *Alu* subclasses, they also showed major homology with these sequences. Probes designed for conserved regions of *Alu* repeats (*Alucon*, *Aluconatcc1a*, and *Aluconatcc2*) did not show >85% homology with any of the sequences.

To make screening cost effective, we multiplexed the screening. This screening was as effective as screening with single probes. About 40% of GGAA-positive clones could be prescreened as *Alu*-positive clones. This screening was very effective, as it was shown that only ~5% of the clones sequenced

MICROSATELLITE ASSOCIATION WITH *ALU* REPEATS

were positive for *Alu* repeats. Similarly, the loss of ATA microsatellite clones due to the presence of *Alu* repeats was reduced from 50% to 10% by prescreening these clones with *Alu* probes. Therefore, this prescreening could be cost effective. Similar strategies can be applied for prescreening of *Alu* repeats associated with other classes of microsatellites. Probes designed for *Alu* prescreening can also be used for physical mapping purposes. The new *Alu* primers designed were successfully used for *Alu*-*Alu* fingerprinting. Three of the primers (*Alu*ATA1, *Alu*GAAA3, and *Alu*GGAA4) are positioned such that they are also useful for developing *Alu*-*Alu* PCR.

METHODS

Construction of Primary and Marker-Enriched Small-Insert Genomic Libraries

Construction of primary and marker-enriched small insert libraries was carried out according to the methods described earlier (Pulido and Duyk 1994).

Screening with ³²P-Labeled Probes

Single colonies from marker-selected libraries were picked into single wells in a 96-well microtiter plate containing 200 μ l of 2 $\times$ YT medium plus 100 μ g/ml of carbenicillin (cb) and 10% glycerol. Plates were incubated overnight at 37°C. Colonies were replicated onto LB agar plates and nylon filters (MSI). Nylon membranes were transferred to LB agar plus cb plates with colony side up. Colonies were allowed to grow overnight at 37°C. Colony lysis, DNA fixing, and denaturing were carried out according to Buluwela et al. (1989). Pre-hybridization, hybridization, and washing were carried out with modifications of standard protocols (Ausubel et al. 1988; Sambrook et al. 1989). For microsatellite screening a [GGAA]₁₀ oligonucleotide probe was used. Rescreening was carried out to confirm the positive clones before growing the colonies for sequencing. During rescreening, filters were probed with *Alu* probes and [GGAA]₁₀ was probed separately. Hybridization with *Alu* probes was done at 45°C.

Nonradioactive Screening

Screening was performed for primary, secondary, and *Alu* screening of the ATA clones using the Quick-lite hybridization system (FMC Corp.). Screening with the [ATA]_n probe was carried out at 32°C. *Alu* screening was performed at 55°C.

Primer Design

Primers were designed after aligning sequences with the PILEUP program of the GCG package, version 7.0. Consensus sequences in the area of *Alu* repeats were obtained by visual inspection of the aligned sequences. These primers were

Table 3. Sequences of *Alu* probes

Probe name	Sequence
<i>AluCon</i> ^{a,b}	AAGTGCTGGGATTACAGGTGTGAGCCACCA
<i>AluATA1</i> ^b	GGAGGCTGAGAGGCAGGAGAATCGCCTGA
<i>AluATA2</i> ^b	CCCAGGCTGGAGTGCAGTGGCA
<i>AluGGAA4</i>	GCACTCCAGCCTGGGTGACAGAG
<i>AluGAAA1</i>	ATCTCGGCTCACTGCAACCTCCGT
<i>AluGAAA2</i>	GAGACCAGCCTGGCCAACATGGT
<i>AluGAAA3</i>	GTGCCACTGCACTCCAGCCTGGG
<i>AluGAAA4</i>	CCCAGCTACTTGGAAGGCTGAGGCAGG
<i>AluConatcc1</i> ^{a,c}	CGACCTCGAGATCTCGGCTCACTGCAA
<i>AluConactcc2</i> ^{a,d}	AAGTCGCCGGCCGCTTGCAGTGAGGCCGAGAT

^aProbes designed **not** in the present study.^bAvailable from Life Codes, Inc.^cDiffers from American Type Culture Collection (ATCC, Rockville, MD) oligonucleotide probe *Aluconsensus* (designation 517) at position 15, which is C/T, and at position 16, which is G/A.^dAvailable from ATCC as *Aluconsensus* (designation 559) oligonucleotide probe.

checked for homology with sequences of GGAA, GAAA, and ATA repeat classes. Probes for screening GGAA repeat microsatellites were designed by using 103 GGAA and 45 ATA repeat sequences associated with *Alu* repeats. GGAA and ATA microsatellite sequences were generated in our laboratory. Probes for GAAA repeats were designed by using GAAA repeats containing sequences from GenBank. *Alu* probes used in this study, along with some commercially available probes, are shown in Table 3.

Classification of *Alu* Repeats

Microsatellite sequences with >10 repeats (repeats equal to 12 or more for CA; 6 or more for GGA, GAA, and GTT) were sent by e-mail to the Pythia server (pythia@anl.gov) and were classified into subfamilies according to the criteria described by Jurka and Milosavljevic (1991; Jurka 1993). Nomenclature of *Alu* subfamilies was changed according to the standard (Batzer et al. 1996). Sequences, except for GGAA and ATA, were obtained from the GenBank database.

Sequencing

Plasmid DNA was purified by using Magicprep columns (Promega). Double-stranded DNA was sequenced on an Applied Biosystems 373A automatic sequencer using their *Taq* dye primer cycle sequencing kit.

Alu–*Alu* PCR

Alu–*Alu* PCR was carried out according to the method described in Zoghbi et al. (1991). The total volume of the PCR reactions was 100 μ l with a 2 μ M final primer concentration. A single YAC colony was resuspended in sterile water to serve as a template in a reaction of 10 mM Tris-HCl (pH 8.3), 1.5 mM MgCl₂, 50 mM KCl, 0.1 mg/ml of gelatin, 250 μ M each dNTP,

and 5 units of *Taq* polymerase (Boehringer Mannheim). PCR conditions included an initial denaturation at 94°C for 5 min, followed by 30 cycles of 1 min at 94°C, 1 min at 55°C, and 4 min at 72°C. A final elongation step consisted of 7 min at 72°C. Aliquots (20 μ l) were separated in a 1% SeaKem LE agarose gel (FMC Bioproducts). All primer sequences are listed in Table 3, except PDJ34 [TGAGC(C/T)(G/A)(A/T)GAT(C/T)(G/A)(C/T)(G/A)CC(A/T)CTGCACTCCAGCCTGGG] (Breukel et al. 1990).

Statistical Analysis

The significance of association of *Alu* repeats with microsatellites was analyzed using χ^2 analysis with StatView software (StatView 4.5 statistical software, from Abacus Concepts, Berkeley, CA). Deviations of expected values from observed values were tested by cell χ^2 .

ACKNOWLEDGMENTS

We would like to thank Dr. Aleksander Milosavljevic, Biology and Medical Research Division, Argonne National Laboratory, Argonne and Dr. Jerzy Jurka, Linus Pauling Institute, Palo Alto for the classification *Alu* repeats on the Pythia server. We also want to thank Dr. Daniel Eberl, Department of Genetics, Harvard Medical School, Boston for his suggestions. We thank Dr. George De Sanctis, Department of Medicine, Harvard Medical School for helping in the statistical analysis. This work was supported by grant NIH 1 P50 H00835-01 from the National Institutes of Health.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Arcot, S.S., Z. Wang, J.L. Weber, P.L. Deininger, and M.A. Batzer. 1995. *Alu* repeats: A source for the genesis of primate microsatellites. *Genomics* 29:136–144.
- Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. 1988. *Current protocols in molecular biology*. Vol. 1. Greene/John Wiley, New York, NY.
- Batzer, M.A., G.E. Kilroy, P.E. Richard, T.H. Shaikh, T.D. Desselte, C.L. Hoppens, and P.L. Deininger. 1990. Structure and variability of recently inserted *Alu* family members. *Nucleic Acids Res.* 18: 6793–6798.
- Batzer, M.A., P.L. Deininger, U. Hellmann-Blumberg, J. Jurka, D. Labuda, C.M. Rubin, C.W. Schmid, E. Zietkiewicz,

MICROSATELLITE ASSOCIATION WITH ALU REPEATS

- and E. Zuckerkandl. 1996. Standardized nomenclature for Alu repeats. *J. Mol. Evol.* 42: 3–6.
- Beckmann, J.S. and J.L. Weber. 1992. Survey of human and rat microsatellites. *Genomics* 12: 1–5.
- Breukel, C., J. Wijnen, C. Tops, H. Klift, H. Dauwerse, and P.M. Khan. 1990. *Nucleic Acids Res.* 18: 3097.
- Britten, R.J., W.F. Baron, D.B. Stout, and E.H. Davidson. 1988. Sources and evolution of human Alu repeated sequences. *Proc. Natl. Acad. Sci.* 85: 4770–4774.
- Buluwela, L., A. Forster, T. Boehm, and T.H. Rabbitts. 1989. A rapid procedure for colony screening using nylon filters. *Nucleic Acids Res.* 17: 452.
- Dietrich, W.F., J.C. Miller, R.G. Steen, M. Merchant, D. Damron, R. Nahf, A. Gross, D.C. Joyce, M. Wessel, R.D. Dredge, A. Marquis, L.D. Stein, N. Goodman, D.C. Page, and E.S. Lander. 1994. A genetic map of the mouse with 4006 simple sequence length polymorphisms. *Nature Genet.* 7: 220–245.
- Economou, E.P., A.W. Bergen, A.C. Warren, and S.E. Antonarakis. 1990. The poly(A) tract of Alu repetitive elements is polymorphic in the human genome. *Proc. Natl. Acad. Sci.* 87: 2951–2954.
- Edwards, A., A. Civitello, H.A. Hammond, and C.T. Caskey. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* 49: 746–756.
- Edwards, A., H.A. Hammond, L. Jin, C.T. Caskey, and R. Chakraborty. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12: 241–253.
- Gastier, J.M., J.C. Pulido, S.L. Sunden, T. Brody, K.H. Beutow, J.C. Murray, J.L. Weber, T.J. Hudson, V.C. Sheffield, and G.M. Duyk. 1995. Survey of trinucleotide repeats in the human genome: Assessment of their utility as genetic markers. *Hum. Mol. Genet.* 4: 1829–1836.
- Gyapay, G., J. Morissette, A. Vignal, C. Dib, C. Fizames, P. Millasseau, S. Marc, G. Bernardi, M. Lathrop, and J. Weissenbach. 1994. The 1993–94 Genethon human genetic linkage map. *Nature Genet.* 7: 246–339.
- Hearne, C.M., S. Ghosh, and J.A. Todd. 1992. Microsatellites for linkage analysis of genetic traits. *Trends Genet.* 8: 287–294.
- Jin, L., Y. Zhong, and R. Chakraborty. 1994. The exact numbers of possible microsatellite motifs. *Am. J. Hum. Genet.* 55: 582–583.
- Jurka, J. 1993. A new subfamily of recently retroposed human Alu repeats. *Nucleic Acids Res.* 21: 2252.
- Jurka, J. and A. Milosavljevic. 1991. Reconstruction and analysis of human Alu genes. *J. Mol. Evol.* 32: 105–121.
- Jurka, J. and C. Pethiyagoda. 1995. Sequences from primates: Compilation and analysis. *J. Mol. Evol.* 40: 120–126.
- Jurka, J. and T. Smith. 1988. A fundamental division in the Alu family of repeated sequences. *Proc. Natl. Acad. Sci.* 85: 4775–4778.
- Kapitonov, V. and J. Jurka. 1996. The age of Alu subfamilies. *J. Mol. Evol.* 42: 59–65.
- Lagercrantz, U., H. Ellegren, and L. Andersson. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* 21: 1111–1115.
- LeBlank-Straceski, J.M., K.T. Montgomery, H. Kissel, L. Murtaugh, P. Tsai, D.C. Ward, K.S. Krauter, and R. Kucherlapati. 1994. Twenty-one polymorphic markers from human chromosome 12 for integration of genetic and physical maps. *Genomics* 19: 341–349.
- Litt, M. and J.A. Luty. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44: 397–401.
- Matera, A.G., U. Hellmann, M.F. Hintz, and C.W. Schmid. 1990. Recently transposed Alu repeats result from multiple source genes. *Nucleic Acids Res.* 18: 6019–6023.
- Murray, J.C., K.H. Beutow, J.L. Weber, S. Ludwigsen, T. Scherpbier-Heddema, F. Manion, J. Quillen, V.C. Sheffield, S. Sunden, G.M. Duyk, J. Weissenbach, G. Gyapay, C. Dib, J. Morissette, G.M. Lathrop, A. Vignal, R. White, N. Matsunami, S. Gerken, R. Melis, H. Albertsen, R. Plaetke, S. Odelberg, D. Ward, J. Dausset, D. Cohn, and H. Cann. 1994. A comprehensive human linkage map with centimorgan density. *Science* 265: 2049–2054.
- Polymeropoulos, M.H., D.S. Rath, H. Xio, and C.R. Merrill. 1991. Tetranucleotide repeat polymorphism at the beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2). *Nucleic Acids Res.* 19: 1432–1432.
- Puers, C., H.A. Hammond, L. Jin, C.T. Caskey, and J.W. Schumm. 1993. Identification of repeat heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]_n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. *Am. J. Hum. Genet.* 53: 953–958.
- Pulido, J.C. and G.M. Duyk. 1994. Construction of small insert libraries enriched for short-tandem-repeat sequences by marker selection. In *Current protocols in human genetics* (ed. A. L. Boyle), pp. 2.2.1–2.2.31. John Wiley and Sons, New York, NY.
- Quentin, Y. 1992. Origin of the Alu family: A family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. *Nucleic Acids Res.* 20: 3397–3401.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

YANDAVA ET AL.

Sheffield, V.C., J.L. Weber, K.H. Buetow, J.C. Murray, D.A. Even, K. Wiles, J.M. Gastier, J.C. Pulido, C. Yandava, S.L. Sundén, G. Mattes, T. Businga, A. McClain, J. Beck, T. Scherpiër, J. Gilliam, J. Zhong, and G.M. Duyk. 1995. A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide maps. *Hum. Mol. Genet.* 4: 1837–1844.

Sinnett, D., J. Deragon, L.R. Simard, and D. Labuda. 1990. Alu polymorphisms-Human DNA polymorphisms detected by polymerase chain reaction using Alu-specific primers. *Genomics* 7: 331–334.

Stallings, R.L., A.F. Ford, D. Nelson, D.C. Torney, C.E. Hildebrand, and R.K. Moyzis. 1991. Evolution and distribution of (GT)_n repetitive sequences in mammalian Genomes. *Genomics* 10: 807–815.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17: 807–815.

Warne, D., C. Watkins, P. Bodfish, K. Nyberg, and N.K. Spurr. 1991. Tetranucleotide repeat polymorphism at the human-actin related pseudogene 2 (ACTBP2) detected using the polymerase chain reaction. *Nucleic Acids Res.* 19: 6980.

Weber, J.L. 1990. Human DNA polymorphisms based on length variations in simple sequence tandem repeats. In *Genome analysis, volume 1: Genetic and physical mapping* (ed. S. Tishman and K. Davies), pp. 159–181. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Weber, J.L. and P.E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44: 388–396.

Weissenbach, J., G. Gyapay, C. Dib, A. Vignal, J. Morissette, P. Millasseau, G. Vaysseix, and M. Lathrop. 1992. A second-generation linkage map of the human genome. *Nature* 359: 794–801.

Zoghbi, H.Y., A.E. McCall, and F. LeBorgne-Demarquoy. 1991. Sixty-five radiation hybrids for the short arm of human chromosome 6: Their value as a mapping panel and as a source for rapid isolation of new probes using repeat element-mediated PCR. *Genomics* 9: 713–720.

Zuliani, G. and H. Hobbs. 1990. A high frequency of length polymorphisms in repeated sequences adjacent to alu sequences. *Am. J. Hum. Genet.* 46: 963–969.

Received January 7, 1997; accepted in revised form May 29, 1997.

Characterization of *Alu* Repeats That Are Associated with Trinucleotide and Tetranucleotide Repeat Microsatellites

Chandri N. Yandava, Julie M. Gastier, Jacqueline C. Pulido, et al.

Genome Res. 1997 7: 716-724

Access the most recent version at doi:[10.1101/gr.7.7.716](https://doi.org/10.1101/gr.7.7.716)

References

This article cites 36 articles, 4 of which can be accessed free at:
<http://genome.cshlp.org/content/7/7/716.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Affordable, Accurate
Sequencing.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>