THYROID
Volume 12, Number 9, 2002
© Mary Ann Liebert, Inc.

Genotyping and Characterization of Two Polymorphic Microsatellite Markers Located Within Introns 29 and 30 of the Human Thyroglobulin Gene

Carina M. Rivolta, Christian M. Moya, Fernando M. Mendive, and Héctor M. Targovnik

The purpose of the present work was to characterize two new polymorphic microsatellite markers in the thyroglobulin gene. TGrI29 and TGrI30 repeats are located within introns 29 and 30, respectively. Genetic studies were carried out by using polymerase chain reaction (PCR) followed by denaturing polyacrilamide gel electrophoresis. TGrI29 exhibited clearly 4 distinguishable alleles ranging from 197 to 203 base pair (bp) in length and TGrI30 showed 8 alleles ranging from 502 to 542 bp. We characterized the two markers by determinating allele frequencies and measures of variation. The heterozygosities (HET) observed of TGrI29 and TGrI30 were 0.859 and 0.522, respectively. The polymorphism information contents (PIC) were 0.471 and 0.434, respectively. No significant differences from Hardy-Weinberg values were found for these two systems. The PCR products of each allele were cloned using the pGEM-T Easy vector and directly sequenced by *Taq* polymerase-based chain terminator method. Sequencing analysis indicated that both loci are complex repeats, TGrI29 containing two types of variable motifs $(\mathbf{tc})_n$ and $(\mathbf{tg})_n$, and TGrI30 a tetra-nucleotide tandem units $(\mathbf{atcc})_n$. In two TGrI29 alleles and one TGrI30 allele were found two different subtypes in each one, with the same molecular weights but different distribution of the tandem repeats. In conclusion, both microsatellites analyzed are highly informative polymorphic markers and can be used in linkage studies in families with congenital hypothyroidism or autoimmunity thyroid diseases.

Introduction

Thyroglobulin (Tg) is a large homodimer glycoprotein synthesized and secreted by the thyroid cells into the lumen of the thyroid follicle (1–3). Tg serves as the matrix for thyroid hormones synthesis, triiodothyronine, and thyroxine (2). The human Tg, mapped on chromosome 8q24, is coded by a single-copy gene 270 kb long and contains 8.5 kb of coding sequence divided into 48 exons (3–7). Several mutations in the Tg gene have been reported (8–14) and are associated with congenital goiter and variable degrees of hypothyroidism. The inheritance mode of this monogenic disease is autosomal recessive. The availability of highly informative polymorphic markers will allow indirect disease diagnosis by genetic linkage studies, such as in cases with no identified mutations and for rapid identification of affected newborns or gene carriers in families with Tg mutations. The term DNA polymorphism refers to a wide range of variations in nucleotide base composition (15), single nucleotide polymorphism (SNP), insertion and deletion sequences (Indel), or length of nucleotide repeats. This latter group includes two categories of multiallelic tandemly repeated DNA sequences (16,17). Loci with repeated motifs of a few base pairs are often referred to as short tandem repeats (STR) or microsatellites (16), while those with longer repeated motifs are referred to as variable number of tandem repeats (VNTR) or minisatellites (17). STRs proved to be the most suitable markers in linkage analysis between a disease locus and a genetic marker because of their diversity levels, high degree of resolution, relatively low mutation rates, being highly informative, and rapid typing. Recently the Tg gene has been identified as a major susceptibility gene for familial autoimmune thyroid diseases (AITD), by linkage analysis using a Tg microsatellite inside intron 27 (18).

We previously identified two compound microsatellite repeats located within introns 29 and 30 of the Tg gene (7). We present here the genotyping and characterization by cloning and sequencing of these two new polymorphic Tg mi-

Laboratorio de Biología Molecular, Cátedra de Genética y Biología Molecular, Facultad de Farmacia y Bioquímica. Universidad de Buenos Aires, Buenos Aires, Argentina.

```
................................gggactgcaaacagtgtggtgatgagctagagttagcagggct
gttgtgttgctgaagtgcaaatcaagacaatccgagttgaggctcaaaatgcaagtagagcccagattaagaaaagaa
tctggacttattttctgtaggcagtggagactccaggaagagctgggatcatggcatgtaaatacctgaagacagagt
cagggaatggttgagctaaaactttcaagtttgggttaagtggagctgaaagctgactttgtagatgaagactaccat
tttggagagaggctggacttccatgtggcttcagaagacaaaaagaggatcagtggttcaaagttacaaaaagaggac
ttgactatacaacttcctaatgatggagtgggctgctgtggttagcacaatgatgtcatgtggagctatagaggcatt
agacagaagtctccaggggaaaacttctgcattaggtgaaaaggtggaccagataacttccaaggctttatccaaacc
cagagtctaagatttgccatcccataattttaggacttcctggggccttctgaactattcctgtctgacccaactc  TGrI29F
ccagctctactgacttcaatcctgctctggtttcacacagcataaaaatgtctctgtctctatctctgacactttctc
tctctgtctctctctgtgtgtgtgtgtgtgtgtgtttctttcttgtgtttttctcatattcactagagttaaaggtgc  TGrI29R
aggaagttgactccctgcttctttttcag GTTTGACAACAGAACTTTTCTCCCCTGTGGACCTCAACCAGGTCATTG
Exon 30                       L   T   T   E   L   F   S   P   V   D   L   N   Q   V   I   V
TCAATGGAAATCAATCACTATCCAGCCAGAAGCACTGGCTTTTCAAGCACCTGTTTTCAGCCCAGCAGGCAAACCTAT
  N   G   N   Q   S   L   S   S   Q   K   H   W   L   F   K   H   L   F   S   A   Q   Q   A   N   L   W
GGTGCCTTTCTC gtaagtatccttagaactcattcttcttcttccagacactgtagtcaggcatcacaggccaagtc
  C   L   S   R
tggcaactcctgagacacagataaggccaaattttgtgtcataactcatggatcactgtggatgatataaaagaaaga  TGrI30F
tagcacataacccatatcttcaagaaggttacagtgtagttaatctctactctgcttccataacctatcttcatccac
ctgtctctccatctatccgtccattcacccatccatccatccatccatctacccatccatcccattcatccatccatc
tacccatttatccatcccatccatccttccatcccatccacccatccatccatccatccatccatccatcccattcat
ccgtccatctttgcatcccatccatccatccattcgatccatccatccatccatccatccatccatccatccatccat
ccacccatccaaccatccatccatccatgtatccatgtatccatcccatccaaccatccatccatccatccatccatc
catccatccatccatccatctatccatccatccatttggcagttattaagtatgcattatgttcaaagctttgtgttt
gattttggagaagaaagagaacacccaaaatgaatcaaatcaaccatgttccttgcccttggggagctcatatatacc  TGrI30R
aggggcacaggctgacttgggcacaggtgcagggagaccattcacaaagttcacagttttacagtcctgtgacctct
ggtctgcagatcctgacttcctctgagatggccatgggtagatccaatgcaatgttactcttttgcctcttgtattgc
tgagtttgaggtttgtcccttatgccttcaactctgcctttattcaggagcacacggccccctccccagctagcagaa
ccccagtgtttttgcatgggccttccagactggatgatcatgacagtctcaggcctctgccctaactaggagggatct
ctaccaggcatttccccacccagagaatcctgtagagattattctcccctgtagaccccacaaaaactaaaatcacac
tactctcttgcctgtag GTTGTGTGCAGGAGCACTCTTTCTGTCAGCTCGCAGAGATAACAGAGAGTGCATCCTTGT
Exon 31                 C   V   Q   E   H   S   F   C   Q   L   A   E   I   T   E   S   A   S   L   Y
ACTTCACCTGCACCCTCTACCCAGAGGCACAGGTGTGTGATGACATCATGGAGTCCAATGCCCAGGGCTGCAGACTGA
  F   T   C   T   L   Y   P   E   A   Q   V   C   D   D   I   M   E   S   N   A   Q   G   C   R   L   I
TCCTGCCTCAGATGCCAAAGGCCCTGTTCCGGAAGAAAG gtgagcacttggagagatctgcataaactgtatttcca
  L   P   Q   M   P   K   A   L   F   R   K   K   V
atgttctgctggctctgtggtttagaaagatccacattagtttcttatttaaaaaacattttctttatacttttatt
ggtttcaagaagatgggaagatggatccaaactttcacggttagaatcaataaactactacgtaattgtctcagttta
aacaaaaaatatgaaaagggttatttgggggcagtatt......................................
```

FIG. 1. Sequence data of the exons 30 and 31, 3′ end of intron 29, complete intron 30 and 5′ end of intron 31 of the thyroglobulin gen, from λ dash 171. The exonic sequences are indicated by capital letters and the intronic sequences by lower-case letters. The amino acid sequence is represented by the one-letter code. Shaded areas indicate the microsatellites of introns 29 (TGrI29) and 30 (TGrI30). The positions of the primers for TGrI29 (forward: TGrI29F, reverse: TGrI29R) and TGrI30 (forward: TGrI30F, reverse: TGrI30R) microsatellites are underlined.

crosatellite markers. In order to improve the accuracy of linkage analysis, we have established allele and genotype frequencies, and measures of variation of these loci.

Materials and Methods

Genomic DNA isolation

Peripheral blood samples collected in sodium ethylenediaminetetraacetic acid (EDTA) from unrelated individuals without thyroid pathology were used. Genomic DNA was isolated from white blood cells by the sodium dodecyl sulfate (SDS)-proteinase K method.

Microsatellite analysis

The microsatellites present in introns 29 and 30 of the Tg gene are referred as TGrI29 and TGrI30, respectively. The tandem repeat sequences and the primers used for amplification of these regions are shown in Figure 1.

Polymerase chain reaction (PCR) amplification reactions

S_1 S_2 S_3 S_4 S_5 S_6 S_7

29.4: 203 bp
29.3: 201 bp
29.2: 199 bp
29.1: 197 bp

FIG. 2. Autoradiograph showing alleles at TGrI29 locus for unrelated individuals. Molecular weights are indicated in base pairs (bp).

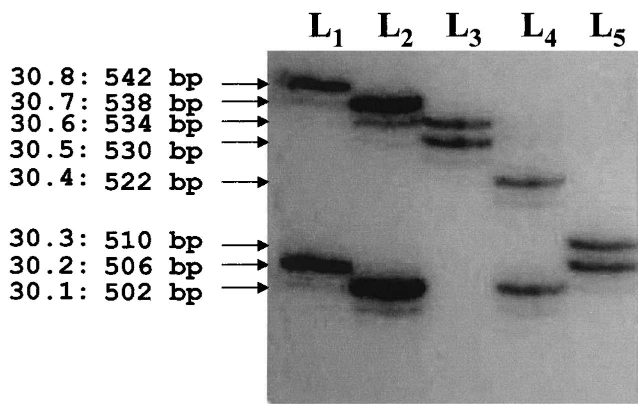

FIG. 3. Autoradiograph showing alleles at TGrI30 locus for unrelated individuals. Molecular weights are indicated in base pairs (bp).

were performed in a total volume of 15 μL using a standard PCR buffer (Invitrogen-Life Technologies, Carlsbad, CA) containing 125 ng of DNA, 14 pmol of each primer (the forward primer was ^{32}P end-labeled using polynucleotide kinase), 0.2 mM each deoxyribonucleotide triphosphate, 2.5 mM $MgCl_2$, 4% dimethyl sulfoxide (DMSO) and 0.3 units of *Taq* polymerase (Invitrogen-Life Technologies). Samples were heated to 95°C for 3 minutes, followed by 25 cycles of DNA denaturation (95°C for 30 seconds), annealing (59°C for 30 seconds), and polymerization (72°C for 1 minute). After the last cycle, the samples were incubated for an additional 10 minutes at 72°C. Amplification was carried out in a MJ Research PTC 100 thermoblock (MJ Research, Watertown, MA).

Aliquots of PCR products were mixed with one half volume of formamide dye solution (95% formamide, 10 mM NaOH, 0.05% bromophenol blue, and 0.05% xylene cyanol), heated to 70°C for 2 minutes, and electrophoresed on a 6% (TGrI29) or 5% (TGrI30) denaturing polyacrylamide sequencing gel. Autoradiography of dried gels was performed for 24 hours. Alleles were sized by comparison with a M13 mp18 sequencing standard electrophoresed on the same gel.

Cloning of microsatellites

Genomic DNA of individuals carrying the different alleles to the TgRi29 and TgRi30 microsatellites was amplified by PCR. The total volume of 100 μL contained the standard PCR buffer (Invitrogen-Life Technologies), 250–300 ng of DNA, 50 pmol of each forward and reverse primers (Fig. 1), 0.2 mM each deoxyribonucleotide triphosphate, 2.5 mM $MgCl_2$, 4% DMSO, and 2 units of *Taq* polymerase (Invitrogen-Life Technologies). The same cycles as those described above were used except that the annealing temperature was 58°C for TGrI30. Amplification was carried out for 40 cycles. Control reactions without added DNA were included in every set of amplification.

The PCR products were purified from a 2% agarose gel on Concert Rapid Gel Extraction System (Invitrogen-Life Technologies) and cloned using the pGEM-T Easy Vector (Promega, Madison, WI). JM109 High Efficiency Competent Cells (Promega) were used for transformations. After that recombinant clones were identified by color screening on indicator plates containing IPTG (isopropylthio-β-galactoside) and X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) plasmid DNA was isolated with the Concert Rapid Plasmid Miniprep System (Invitrogen-Life Technologies).

DNA sequencing

Sequencing of the alleles was performed with the *Taq* polymerase-based chain terminator method (fmol; Promega) from TGrI29 and TGrI30 clones, using the amplification reverse primers (Fig. 1). The results were analyzed using the PC gene computer program (Intelligenetics, Geneva, Switzerland).

Statistical analysis

A standard χ^2 analysis of the observed and expected genotypes for TGrI29 and TGrI30 microsatellites was carried out in order to test for Hardy-Weinberg equilibrium (19).

TABLE 1. SUMMARY OF MEASURES OF VARIATION

STR	*Location*	*Allele*	*PCR product size (bp)*	*Allele frequencies*	*HET*	*PIC*
TGrI29	Intron 29	29.1	197	0.336	0.859	0.471
		29.2	199	0.200		
		29.3	201	0.450		
		29.4	203	0.014		
TGrI30	Intron 30	30.1	502	0.099	0.522	0.434
		30.2	506	0.115		
		30.3	510	0.013		
		30.4	522	0.007		
		30.5	530	0.003		
		30.6	534	0.036		
		30.7	538	0.720		
		30.8	542	0.007		

HET, Heterozigosity; PIC, Polymorphism information content.

TABLE 2. GENOTYPE FREQUENCIES OF THE TGrI29 SYSTEM

Genotype	*Observed (n)*	*Expected (n)*
29.1/29.1	7	7.902
29.1/29.2	9	9.408
29.1/29.3	22	21.168
29.1/29.4	2	0.659
29.2/29.2	4	2.800
29.2/29.3	11	12.600
29.2/29.4	0	0.392
29.3/29.3	15	14.175
29.3/29.4	0	0.882
29.4/29.4	0	0.014

$\chi^2 = 4.9357$; $p > 0.50$ (not significant); $df = 6$; $n = 70$.

Results

Two microsatellites TGrI29 and TGrI30 located within introns 29 and 30, respectively, were found during the sequencing of the TG gene in the clone named λ dash 171 (Fig. 1) (7). TGrI29 is a complex STR, with a repeat region containing a mixture of two types of dinucleotide tandem repeats $(\mathbf{tc})_n$ and $(\mathbf{tg})_n$ and TGRi30 consists of one type of tetranucleotide tandem unit $(\mathbf{atcc})_n$. In order to study the possible polymorphism of these repetitive sequences we amplified them by PCR using flanking primers. Then, the labeled products were separated in a polyacrylamide denaturing sequencing gel. The genotyping were carried out for TGrI29 and TGrI30 in a population sample of 70 and 152 unrelated individuals, respectively. The fragments amplified with TGrI29 primers were in a size range between 197 and 203 base pairs (bp) (Fig. 2). The alleles observed varied in length by 2 bp. So, we have assigned numbers for them begining with 29.1 to indicate the 197-bp allele and incrementing at 2-bp intervals (from 29.1 to 29.4, see Fig. 2). The fragments amplified with TGrI30 primers were in a size ranging between 502 and 542 (Fig. 3). We have coded these alleles with numbers begining with 30.1 to indicate the 502-bp allele (from 30.1 to 30.8, see Fig. 3). The allele sizes were easily determined by comparison with the M13 mp18 sequences. In addition to the main amplification product, each allele presents a typical shadow product that is 2 (TGrI29) or 4 bp (TGrI30) smaller and less intense. The reason may be slippage during PCR amplification or incomplete extension by the polymerase.

The allele and genotype frequencies of these microsatellites are summarized in Tables 1, 2, and 3. TGrI29 locus has four alleles, three of them being relatively common in the population of 140 chromosomes studied (Table 1). Of the 10 possible genotypes from the 4 observed alleles, 7 were seen in the population survey (Table 2).

A total of 8 different TGrI30 alleles were observed in the 304 chromosomes included in this study, 4 of them being the most common (Table 1). Of the 36 possible genotypes from

TABLE 3. GENOTYPE FREQUENCIES OF THE TGrI30 SYSTEM

Genotype	*Observed (n)*	*Expected (n)*
30.1/30.1	0	1.490
30.1/30.2	3	3.461
30.1/30.3	0	0.391
30.1/30.4	1	0.211
30.1/30.5	0	0.090
30.1/30.6	2	1.083
30.1/30.7	24	21.669
30.1/30.8	0	0.211
30.2/30.2	1	2.010
30.2/30.3	1	0.454
30.2/30.4	0	0.245
30.2/30.5	0	0.105
30.2/30.6	1	1.259
30.2/30.7	27	25.171
30.2/30.8	1	0.245
30.3/30.3	0	0.026
30.3/30.4	0	0.028
30.3/30.5	0	0.012
30.3/30.6	0	0.142
30.3/30.7	3	2.845
30.3/30.8	0	0.028
30.4/30.4	0	0.007
30.4/30.5	0	0.006
30.4/30.6	0	0.077
30.4/30.7	1	1.532
30.4/30.8	0	0.015
30.5/30.5	0	0.001
30.5/30.6	1	0.033
30.5/30.7	0	0.657
30.5/30.8	0	0.006
30.6/30.6	0	0.197
30.6/30.7	7	7.880
30.6/30.8	0	0.077
30.7/30.7	78	78.797
30.7/30.8	1	1.532
30.8/30.8	0	0.007

$\chi^2 = 40.3540$; $p > 0.05$ (not significant); $df = 28$; $n = 152$.

TABLE 4. HAPLOTYPE FREQUENCIES BETWEEN TGrI29 AND TGrI30 GENOTYPES

Haplotype	*Frequency*
29.1/29.1–30.1/30.2	0.06
29.1/29.1–30.1/30.4	0.02
29.1/29.1–30.2/30.3	0.02
29.1/29.1–30.7/30.7	0.02
29.1/29.2–30.1/30.6	0.02
29.1/29.2–30.1/30.7	0.02
29.1/29.2–30.2/30.2	0.02
29.1/29.2–30.2/30.7	0.04
29.1/29.3–30.1/30.7	0.14
29.1/29.3–30.2/30.7	0.10
29.1/29.3–30.2/30.8	0.02
29.1/29.3–30.6/30.7	0.02
29.1/29.3–30.7/30.7	0.06
29.1/29.4–30.1/30.7	0.02
29.1/29.4–30.2/30.7	0.02
29.2/29.2–30.5/30.6	0.02
29.2/29.3–30.6/30.7	0.04
29.2/29.3–30.7/30.7	0.16
29.3/29.3–30.6/30.7	0.02
29.3/29.3–30.7/30.7	0.16
Other	0.00

$n = 50$.

29.

4 tg (tc)$_2$ tg (tc)$_2$ ta (tc)$_2$ tg acactt (tc)$_4$ tg (tc)$_4$ (tg)$_{13}$
3b tg (tc)$_2$ tg (tc)$_2$ ta (tc)$_2$ tg acactt (tc)$_4$ tg (tc)$_4$ (tg)$_{12}$
3a tg (tc)$_2$ tg (tc)$_2$ ta (tc)$_2$ tg acactt (tc)$_4$ tg (tc)$_6$ (tg)$_{10}$
2 tg (tc)$_2$ tg (tc)$_2$ ta (tc)$_2$ tg acactt (tc)$_4$ tg (tc)$_5$ (tg)$_{10}$
1b tg (tc)$_2$ tg (tc)$_2$ ta (tc)$_2$ tg acactt (tc)$_4$ tg (tc)$_3$ (tg)$_{11}$
1a tg (tc)$_2$ tg (tc)$_2$ ta (tc)$_2$ tg acactt (tc)$_4$ tg (tc)$_4$ (tg)$_{10}$

FIG. 4. Nucleotide sequence comparison of the cloning alleles of the TGrI29 microsatellite. The alleles and subtypes are aligned for optimal homology. The dinucleotide tandem repeats are shown in parentheses.

the 8 observed alleles, 15 were seen in the population survey (Table 3).

To determinate the informativeness of the systems studied, we calculated measures of variation (Table 1). Thus, the observed heterozygosities (HET) of two loci TGrI29 and TGrI30 were 0.859 and 0.522, respectively, and the polymorphism information contents (PIC) were 0.471 and 0.434, respectively. Both STR loci identified are polymorphic and can be used as genetic markers. The χ^2 analyses of observed and expected genotypes for each microsatellites showed a nonsignificant p value, indicating that the sampled population do not deviate from Hardy-Weinberg equilibrium for the TGrI29 and TGrI30 loci (Tables 2 and 3).

The haplotype frequencies between TGrI29 and TGrI30 genotypes analyzed in a sample of 50 unrelated individuals are summarized in Table 4. The most common associations are 29.1/29.3–30.1/30.7, 29.2/29.3–30.7/30.7, and 29.3/29.3–30.7/30.7 with frequencies of 0.16, 0.14, and 0.16, respectively.

Mendelian transmission of the both STRs alleles was verified in two families. However, we also studied the allelic segregation for the two microsatellites in three different fam-

30.

8 atccgtccattcaccc (atcc)$_4$ atctaccc (atcc)$_2$ cattc (atcc)$_2$ atctacccattt (atcc)$_2$ c
7 .... (atcc)$_4$ .... (atcc)$_2$ ..... (atcc)$_2$ .... (atcc)$_2$.
6 .... (atcc)$_4$ .... (atcc)$_2$ ..... (atcc)$_2$ .... (atcc)$_2$.
5 .... (atcc)$_4$ .... (atcc)$_2$ ..... (atcc)$_2$ .... (atcc)$_2$.
4 .... (atcc)$_4$ .... (atcc)$_2$ ..... (atcc)$_2$ .... (atcc)$_2$.
3 .... (atcc)$_4$ .... (atcc)$_2$ ..... (atcc)$_2$ .... (atcc)$_2$.
2B .... (atcc)$_4$ .... (atcc)$_2$ ..... (atcc)$_2$ .... (atcc)$_2$.
2A .... (atcc)$_4$ .... (atcc)$_2$ ..... (atcc)$_2$ .... (atcc)$_2$.
1 .... (atcc)$_4$ .... (atcc)$_2$ ..... (atcc)$_2$ .... (atcc)$_2$.

8 (atcc)$_2$ ttcc (atcc)$_1$ c (atcc)$_1$ accc (atcc)$_7$ cattc (atcc)$_1$ gtccatctttgc (atcc)$_1$ c
7 (atcc)$_2$ .... (atcc)$_1$. (atcc)$_1$ accc (atcc)$_7$ ..... (atcc)$_1$ .... (atcc)$_1$.
6 (atcc)$_2$ .... (atcc)$_1$. (atcc)$_1$ accc (atcc)$_7$ ..... (atcc)$_1$ .... (atcc)$_1$.
5 (atcc)$_2$ .... (atcc)$_1$. (atcc)$_1$ (atcc)$_1$ (atcc)$_5$ ..... (atcc)$_1$ .... (atcc)$_1$.
4 (atcc)$_2$ .... (atcc)$_1$. (atcc)$_1$ accc (atcc)$_8$ ..... (atcc)$_1$ .... (atcc)$_1$.
3 (atcc)$_2$ .... (atcc)$_1$. (atcc)$_1$ accc (atcc)$_6$ ..... (atcc)$_1$ .... (atcc)$_1$.
2B (atcc)$_2$ .... (atcc)$_1$. (atcc)$_1$ accc (atcc)$_6$ ..... (atcc)$_1$ .... (atcc)$_1$.
2A (atcc)$_2$ .... (atcc)$_1$. (atcc)$_1$ accc (atcc)$_6$ ..... (atcc)$_1$ .... (atcc)$_1$.
1 (atcc)$_2$ .... (atcc)$_1$. (atcc)$_1$ accc (atcc)$_6$ ..... (atcc)$_1$ .... (atcc)$_1$.

8 (atcc)$_3$ attcg (atcc)$_{12}$ accc (atcc)$_1$ aacc (atcc)$_1$ (atcc)$_1$ (atcc)$_1$ atgt (atcc)$_1$ atgt
7 (atcc)$_3$ ..... (atcc)$_{11}$ .... (atcc)$_1$ .... (atcc)$_1$ (atcc)$_1$ (atcc)$_1$ .... (atcc)$_1$ ....
6 (atcc)$_3$ ..... (atcc)$_{10}$ .... (atcc)$_1$ .... (atcc)$_1$ (atcc)$_1$ (atcc)$_1$ .... (atcc)$_1$ ....
5 (atcc)$_3$ ..... (atcc)$_{11}$ .... (atcc)$_1$ .... (atcc)$_1$ (atcc)$_1$ (atcc)$_1$ .... (atcc)$_1$ ....
4 (atcc)$_3$ ..... (atcc)$_{11}$ .... (atcc)$_1$ .... (atcc)$_1$ aacc (atcc)$_3$ .... (atcc)$_1$ ....
3 (atcc)$_3$ ..... (atcc)$_{10}$ .... (atcc)$_1$ .... (atcc)$_1$ aacc (atcc)$_3$ .... (atcc)$_1$ ....
2B (atcc)$_3$ ..... (atcc)$_{10}$ .... (atcc)$_1$ .... (atcc)$_1$ aacc (atcc)$_3$ .... (atcc)$_1$ ....
2A (atcc)$_3$ ..... (atcc)$_9$ .... (atcc)$_1$ .... (atcc)$_1$ aacc (atcc)$_3$ .... (atcc)$_1$ ....
1 (atcc)$_3$ ..... (atcc)$_9$ .... (atcc)$_1$ .... (atcc)$_1$ aacc (atcc)$_3$ .... (atcc)$_1$ ....

8 (atcc)$_2$ c (atcc)$_1$ aacc (atcc)$_{10}$ (atct)$_1$ (atcc)$_3$ atttggcagttattaagtatgcattatgtt
7 (atcc)$_2$. (atcc)$_1$ .... (atcc)$_{10}$ (atct)$_1$ (atcc)$_3$ ....
6 (atcc)$_2$. (atcc)$_1$ .... (atcc)$_{10}$ (atct)$_2$ (atcc)$_2$ ....
5 (atcc)$_2$. (atcc)$_1$ .... (atcc)$_{10}$ (atct)$_1$ (atcc)$_3$ ....
4 (atcc)$_2$. (atcc)$_1$ .... (atcc)$_6$ accc-------- ....
3 (atcc)$_2$. (atcc)$_1$ .... (atcc)$_7$ ------------ ....
2B (atcc)$_2$. (atcc)$_1$ .... (atcc)$_6$ ------------ ....
2A (atcc)$_2$. (atcc)$_1$ .... (atcc)$_7$ ------------ ....
1 (atcc)$_2$. (atcc)$_1$ .... (atcc)$_6$ ------------ ....

FIG. 5. Nucleotide sequence comparison of the cloning alleles of the TGrI30 microsatellite. The alleles and subtypes are aligned for optimal homology. The dotted line denotes the conserved nucleotides for each allele and the discontinuous line indicates the absent nucleotides in the lightest alleles. The tetra-nucleotide tandem repeats are shown in parentheses.

ilies with congenital goiter caused by mutations in the Tg gene. The analysis of pedigrees showed that the two markers are informative, indicating a linkage between the mutation and TGrI29 and TgGrI30 alleles and provided a convenient mean to explore the segregation of Tg alleles (data not shown).

In order to explore the origin of the variability, we cloned and sequenced the different alleles of TGrI29 and TGrI30 loci. For TGrI29, identical sequences were observed in all the sequenced alleles except for the variable number of $(\mathbf{tc})_n$ and $(\mathbf{tg})_n$ motifs (Fig. 4). Additionality, both alleles 29.1 and 29.3 contain *each one* two subtypes with the same molecular weights and different distribution of the tandem repeats (29.1a and 29.1b; 29.3a and 29.3b). In TGrI30 alleles we found tetranucleotide tandem units $(\mathbf{atcc})_n$ sometimes appearing in a variable number and in other cases in a conserved repeat number (Fig. 5). The repetitive sequences are disrupted by conserved regions within the same locus. The allele 30.2 presents two subtypes with the same molecular weight of 506 bp and different distribution of the tandem repeats (30.2a and 30.2b). At the end of the repetitive core, some alleles presented additional tetranucleotide units: alleles 30.5, 30.7 and 30.8 include one **atct**, allele 30.6 two **atct** while 30.4 showed an **accc** unit instead of **atct.**

Discussion

The usefulness of TGrI29 and TGrI30 microsatellites as a simple STR system for family linkage studies is clear; it is an informative polymorphism, without genotyping problems, which is readily amplified by PCR.

The incidence of congenital hypothyroidism is estimated to be 1 in 4000. Patients with this heterogeneous disorder can be divided into two groups: goitrous and nongoitrous hypothyroidism. The presence of congenital goiter is usually associated with mutations in genes responsibles for thyroid hormonogenesis: sodium iodide (Na^+/I^-) symporter (20), Tg (8–14), thyroid peroxidase (21,22), and pendrin (23) genes. In families presenting hypothyroidism supposedly caused by a Tg synthesis defect, a linkage analysis could be performed using TGrI29 and TGrI30. A positive family history of goiter was obtained in more than 50% of these patients and frequently more than one sibling was affected per generation (1). In this context, the screening programs for neonatal hypothyroidism ensure early treatment and prevent profound peripheral tissue hypothyroidism. The diagnosis is based on the presence of neonatal goiter associated with hypothyroidism, with low serum thyroid hormone levels and high serum thyrotropin (TSH) concentrations. The use of TGrI29 and TGrI30 markers in the affected newborns would be of value. It may be important for carrier detection and prenatal diagnosis in families with Tg defect. A remarkable advantage of genetic linkage analysis is that there is no need of knowing the disease-causing mutations in the target gene. Nongoitrous neonatal hypothyroidism results from thyroid gland agenesis, dysgenesis and hypoplastic thyroid gland. In most cases, the ethiology is unknown. In some patients it is associated with mutations in genes responsible for the development of thyroid follicular cells: thyrotropin-β (TSH-β) (24), TTF-1 (thyroid transcription factor) (25), *PAX*-8 (paired box transcription factor 8) (26), and TSH receptor genes (27,28). Interestingly, the Tg region has been shown to be consistently linked to familial nongoitrous congenital hypothyroidism (29). Strong evidence also suggests that the Tg gene is a major susceptibility gene for familial autoimmune thyroid disease (AITD) (18). Polymorphisms in the Tg gene may be involved in the ethiology of AITD. The use of TGrI29 and TGrI30 microsatellites is also convenient in linkage analysis in nongoitrous congenital hypothyroidism or familial AITD. These new markers could be used in combination with other DNA polymorphisms with comparable characteristics.

However, tandem sequence repeats were found to modulate the splicing mechanism of different genes (30,31). In previous studies we have identified a 138-bp deletion in two siblings with congenital goiter in the Tg mRNA corresponding to exon 30 (11). Genomic DNA analysis confirmed a **g** to **t** transversion at position +1 in the donor splice site of intron 30 (14). However, it is possible that the TGrI29 and TGrI30 microsatellites within flanking introns of exon 30 can modulate the splicing and consequently lead to an aberrant splicing signal therefore contributing to the molecular mechanism responsible for the disease.

In conclusion, microsatellite analysis using the markers within the Tg gene described and characterizated here is well suited for routine use in laboratories engaged in linkage studies in families with congenital hypothyroidism or autoimmunity thyroid diseases.

Acknowledgments

C.M. Rivolta is a research fellow of the Argentine National Research Council (CONICET). C.M. Moya and F.M. Mendive are research fellows of the Universidad de Buenos Aires H.M. Targovnik is a established investigator of the CONICET.

This work was supported by grants from Universidad de Buenos Aires (TB 80/98, B 087/2001), CONICET (0853/98), FONCYT (05-00000-01591/ PICT 97, 05-08838/ PICT 2000).

The nucleotide sequence data reported in this paper have been submitted to the GenBank, EMBL and DDJ databases under the accession number AF169659.

References

1. Medeiros-Neto G, Targovnik HM, Vassart G 1993 Defective thyroglobulin synthesis and secretion causing goiter and hypothyroidism. Endocr Rev **14:**165–183.
2. Dunn JT 1996 Thyroglobulin: Chemistry and biosynthesis. In: Braverman LE, Utiger RD (eds) Werner's & Ingbar's The Thyroid. J.B. Lippincott Company, Philadelphia, pp. 85–95.
3. van de Graaf SAR, Ris-Stalpers C, Pauws E, Mendive FM, Targovnik HM, de Vijlder JJM 2001 Up-to-date with the human thyroglobulin sequence. J Endocrinol **170:**307–321.
4. van de Graaf SAR, Pauws E, de Vjilder JJM, Ris-Stalpers C 1997 The revised 8307 base pair coding sequence of human thyroglobulin transiently expressed in eukaryotic cells. Eur J Endocrinol **136:**508–515.
5. Mendive FM, Rivolta CM, Vassart G, Targovnik HM 1999 Genomic organization of the 3′ region of the human thyroglobulin gene. Thyroid **9:**903–912.
6. Moya CM, Mendive FM, Rivolta CM, Vassart G, Targovnik HM 2000 Genomic organization of the 5′ region of the human thyroglobulin gene. Eur J Endocrinol **143:**789–798.
7. Mendive FM, Rivolta CM, Moya CM, Vassart G, Targovnik HM 2001 Genomic organization of the human thyroglobu-

lin gene. The complete intron-exon structure. Eur J Endocrinol **145:**485–496.
8. Ieiri T, Cochaux P, Targovnik HM, Suzuki M, Shimoda S-I, Perret J, Vassart G 1991 A 3′ splice site mutation in the thyroglobulin gene responsible for congenital goiter with hypothyroidism. J Clin Invest **88:**1901–1905.
9. Targovnik HM, Medeiros-Neto G, Varela V, Cochaux P, Wajchenberg BL, Vassart G 1993 A nonsense mutation causes human hereditary congenital goiter with preferential production of a 171-nucleotide-deleted thyroglobulin ribonucleic acid messenger. J Clin Endocrinol Metab **77:**210–215.
10. Corral J, Martín C, Pérez R, Sánchez I, Mories MT, San Millan JL, Miralles JM, González-Sarmiento R 1993 Thyroglobulin gene point mutation associated with non-endemic simple goitre. Lancet **341:**462–464.
11. Targovnik H, Vono J, Billerbeck AEC, Cerrone G, Varela V, Mendive F, Wajchenberg BL, Medeiros-Neto G 1995 A 138-nucleotide deletion in the thyroglobulin ribonucleic acid messenger in a congenital goiter with defective thyroglobulin synthesis. J Clin Endocrinol Metab **80:**3356–3360.
12. Hishinuma A, Takamatsu J, Ohyama Y, Yokozawa T, Kanno Y, Kuma K, Yoshida S, Matsuura N, Ieiri T 1999 Two novel cysteine substitutions (C1263R and C1995S) of thyroglobulin cause a defect in intracellular transport of thyroglobulin in patients with congenital goiter and the variant type of adenomatous goiter. J Clin Endocrinol Metab **84:**1438–1444.
13. van de Graaf SAR, Ris-Stalpers C, Veenboer GJM, Cammenga M, Santos C, Targovnik HM, de Vijlder JJM, Medeiros-Neto G 1999 A premature stopcodon in thyroglobulin mRNA results in familial goiter and moderate hypothyroidism. J Clin Endocrinol Metab **84:**2537–2542.
14. Targovnik HM, Rivolta CM, Mendive FM, Moya CM, Medeiros-Neto G 2001 Congenital goiter with hypothyroidism due to a 5′ splice site mutation in the thyroglobulin gene. Thyroid **11:**683-688.
15. Housman D 1995 Human DNA polymorphism. N Engl J Med **332:**318–320.
16. Bull LN, Pabón-Peña CR, Freimer NB 1999 Compound microsatellite repeats: practical and theoretical features. Genome Res **9:**830–838.
17. Jeffreys AJ, Wilson V, Thein SL 1988 Hypervariable "minisatellite" regions in human DNA. Nature **314:**67–73.
18. Tomer Y, Greenberg DA, Concepcion E, Ban Y & Davies TF 2002 Thyroglobulin is a thyroid specific gene for the familial autoimmune thyroid diseases. J Clin Endocrinol Metab **87:**404–407.
19. Sokal RR, Rohlf FJ 1995 Biometry: The principles and practice of statistics in biological research. W.H. Freeman, New York.
20. Miki K, Harada T, Miyai K, Takai S-I, Amino N. 1997 Congenital hypothyroidism caused by a mutation in the Na^+/I^- symporter. Nat Genet **16:**124–125.
21. Abramowicz MJ, Targovnik MH, Varela V, Cochaux P, Krawiec L, Pisarev MA, Propato FEV, Juvenal G, Chester HA, Vassart G 1992 Identification of a mutation in the coding sequence of the human thyroid peroxidase gene causing congenital goiter. J Clin Invest **90:**1200–1204.
22. Bikker H, Vulsma T, Baas F, de Vijlder JJM 1995 Identification of five novel inactivating mutations in the human thyroid peroxidase gene by denaturating gradient gel electrophoresis. Hum Mutat **6:**9–16.
23. Kopp P 2000 Pendred's syndrome and genetic defects in thyroid hormone synthesis. Rev Endocr Metab Disorders **1:**109–121.
24. Medeiros-Neto G, Herodotou DT, Rayan S, Kommerede S, Lacerca L, Sandrini R, Boguszewski MCS, Hollenberg AN, Rodovick S, Wondisford FE 1996 A circulating biologically inactive thyrotropin caused by a mutation in the beta subunit gene. J Clin Invest **97:**1250–1256.
25. Acebron A, Aza-Blanc P, Rossi DL, Lamas L, Santisteban P 1995 Congenital human thyroglobulin defect due to low expression of the thyroid-specific transcription factor TTF-1. J Clin Invest **96:**781–785.
26. Macchia PE, Lapi P, Krude H, Pirro MT, Missero C, Chiovato L, Souabni A, Baserga M, Tassi V, Pinchera A, Fenzi G, Gruters A, Busslinger M, Di Lauro R 1998 PAX8 mutations associated with congenital hypothyroidism caused by thyroid dysgenesis. Nat Genet **19:**83–86.
27. Sunthornthepvarakul T, Gottschalk ME, Hayashi Y, Refetoff S 1995 Resistance to thyrotropin caused by mutations in the thyrotropin-receptor gene. N Engl J Med **332:**155–160.
28. Abramowicz MJ, Duprez L, Parma J, Vassart G, Heinrichs C 1997 Familial congenital hypothyroidism due to inactivating mutation of the thyrotropin receptor causing profound hypoplasia of the thyroid gland. J Clin Invest **99:**3018–3024.
29. Ahlbom BE, Yaqoob M, Gustavsson P, Abbas HG, Annerén G, Larsson A, Wadelius C 2002 Linkage analysis identifies the thyroglobulin gene region as a major locus for familial congenital hypothyroidism. Hum Genet **110:**145–147.
30. Hedjran F, Yeakley JM, Huh GS, Hynes RO, Rosenfeld MG 1997 Control of alternative pre-mRNA splicing by distributed pentameric repeats. Proc Natl Acad Sci USA **94:** 12343–12347.
31. Pagani F, Buratti E, Stuani C, Romano M, Zuccato E, Niksic M, Giglio L, Faraguna D, Baralle FE 2000 Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. J Biol Chem **275:**21041–21047.

Address reprint requests to:
Héctor M. Targovnik
Laboratorio de Biologia Molecular
Cátedra de Genética y Biología Molecular
Facultad de Farmacia y Bioquímica
Universidad de Buenos Aires
Av. Córdoba 2351
4to piso-sala 5
1120-Buenos Aires
Argentina

E-mail: htargovn@huemul.ffyb.uba.ar