Am. J. Hum. Genet. 52:922–927, 1993

Incidence and Origin of "Null" Alleles in the (AC)n Microsatellite Markers

David F. Callen, Andrew D. Thompson, Yang Shen, Hilary A. Phillips, Robert I. Richards, John C. Mulley, and Grant R. Sutherland

Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, North Adelaide, South Australia

Summary

Twenty-three (AC)n repeat markers from chromosome 16 were typed in the parents of the 40 CEPH (Centre d'Étude du Polymorphisme Humain) families. Where parents were informative, the entire families were then typed. There were seven markers in which null alleles were demonstrated, as recognized by the apparent noninheritance, by a sib, of a parental allele. Four of these markers showed a null allele in a single sibship, while in the other three at least 30% of the CEPH sibships were shown to have a null allele segregating. One null allele was sequenced and shown to be the result of an 8-bp deletion occurring within the priming sequence for PCR amplification of the (AC)n repeats. In gene mapping or in application to diagnosis, the presence of a segregating null allele will not corrupt the linkage data but could result in loss of information. In isolated instances a segregating null allele may be interpreted as nonpaternity. The presence of a null allele may generate misleading data when individuals are haplotyped to determine the presence of linkage disequilibrium with a disease gene.

Introduction

Microsatellite markers which depend on the variability in the length of (AC)n repeats are powerful tools for the genetic analysis of human populations (Weber 1990). Since they can exhibit high heterozygosity and are PCR formatted, they are ideal "index markers" on human chromosomes, for the construction of genetic maps and for application to mapping and diagnosis by linkage in disease families.

Analysis of these markers is dependent on PCR using oligoprimers which flank the (AC)n repeat. Any mutation which is within the DNA sequence complementary to the oligoprimers may inhibit or completely prevent their binding, resulting in either reduced or complete loss of product. Where PCR amplification is entirely prevented, this will be evident as an absence of PCR product. These "null" alleles will not necessarily be recognized when there is a product from the other homologue, and this may lead to an underestimate of marker heterozygosity and to apparent incompatability of genotypes within a family. Null alleles have long been known for protein polymorphisms and, more recently, for VNTR markers (Chakraborty et al. 1992) and have been recognized, together with population subdivision, as a major factor in depression of observed heterozygosity, compared with that expected on the basis of Hardy-Weinberg equilibrium. This report presents a summary of the incidence of null alleles for a number of (AC)n markers on chromosome 16. The molecular basis of one of these null alleles has been determined.

Received August 25, 1992; revision received November 16, 1992.

Address for correspondence and reprints: Dr. David F. Callen, Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, North Adelaide, S.A. 5006, Australia.

0002-9297/93/5204-

0002-9297/93/5205-0009$02.00

Material and Methods

The (AC)n repeat markers from chromosome 16 which were examined for the presence of a null allele are listed in table 1. All (AC)n repeats listed were first typed in the parents of the 40 CEPH (Centre d'Étude

Table 1

Chromosome 16 Microsatellite Markers Genotyped on the CEPH Families

Locus (probe)	Size of Forward Primer, Size of Reverse Primer (bp)	Size of Product (bp)	No. of Alleles	Heterozygosity (%)		No. of CEPH Families	
				Observed	Expected	Typed	With Null Allele
Short arm:							
D16S67 (CRI-0391)[a]	37, 25	149–165	9	48	77	31	17
D16S79A (66F3)[b]	26, 24	153–169	9	52	79	33	13
D16S94 (VK5)[c]	24, 26	82–90	5	58	51	33	0
D16S287 (16XE81)[d]	25, 25	203–227	10	78	78	40	1
D16S291 (AC2.5)[e]	25, 25	154–170	9	83	78	38	0
D16S292 (AC2.3)[e]	24, 24	180–198	10	78	73	39	0
D16S294 (AC1)[e]	25, 24	136–138	2	38	49	23	1
D16S295 (62F3)[f]	24, 25	110–124	8	67	66	35	0
D16S296 (62B4)[f]	28, 25	150–162	8	76	75	38	0
D16S297 (15H1H)[f]	27, 25(R1)[g]	176–184	5	45	61	28	17
	27, 25(R2)	232–244	7	81	74	28	0
D16S298 (3.12)[e]	28, 25	172–192	11	70	79	37	0
D16S299 (6.17)[e]	25, 25	126–140	8	66	72	36	0
D16S300 (AC1.1)[e]	24, 25	163–175	7	76	77	36	0
D16S319 (AC7.14)[h]	20, 20	149–159	7	49	52	29	0
Long arm:							
D16S164 (AC16.15)[i]	29, 21	169–179	5	38	41	24	0
D16S186 (AC16.101)[i]	25, 25	130–178	10	57	63	35	0
D16S261 (MFD24)[j]	20, 20	88–100	6	67	71	36	0
D16S265 (MFD23)[j]	24, 24	160–184	13	77	79	38	0
D16S289 (AC7.46)[e]	28, 20	156–176	8	76	77	36	0
D16S301 (AC6.21)[e]	27, 25	142–152	6	64	64	35	0
D16S303 (AC6.26)[e]	27, 25	101–115	6	36	41	25	0
D16S304 (AC1.14)[e]	26, 28	132–154	12	61	60	33	1
D16S305 (AC1.15)[e]	25, 25	172–200	14	80	82	39	1

[a] Reverse primer sequence = 5'CCACAGCTGGAAACTGGAGAATCAG3'; forward primer sequence = 5'GACTCTGTCTCAAAAAAAAA-AAAAATACAATATATGA3'.

[b] Reverse primer sequence = 5'CAAAGTCTTGGCAGGTATGGACAACT3'; forward primer sequence = 5'ATGAGCCACCAAGCCCG-GGCAGAA3'

[c] Details of primer sequences and PCR conditions are given by Aksentijevich et al. (submitted).

[d] Details of primer sequences and PCR conditions are given by Phillips et al. (1991*b*).

[e] Details of primer sequences and PCR conditions are given by Thompson et al. (1992).

[f] Details of primer sequences and PCR conditions are available from Genome Data Base, Baltimore.

[g] Two different reverse primers were used (see text).

[h] Details of primer sequences and PCR conditions are given by Shen et al. (1992).

[i] Details of primer sequences and PCR conditions are given by Phillips et al. (1991*a*).

[j] Details of primer sequences and PCR conditions are given by Weber et al. (1990).

du Polymorphisme Humain) families. Where parents were informative, the entire families were then typed. As described elsewhere (Thompson et al. 1992), the products of PCR reactions were electrophoresed on denaturing polyacrylamide gels to allow resolution of the lengths of (AC)n repeats.

Observed heterozygosity was determined from the proportion of the 80 CEPH parents who were heterozygous. The expected heterozygosity was calculated from the allele frequencies estimated from these CEPH parents (160 chromosomes).

Sequencing of PCR-amplified DNA was a modified version of Murray's (1989) method. PCR product was purified for sequencing by using the Magic™ PCR Preps DNA Purification System (Promega). PCR and PCR sequencing used the same reaction conditions and

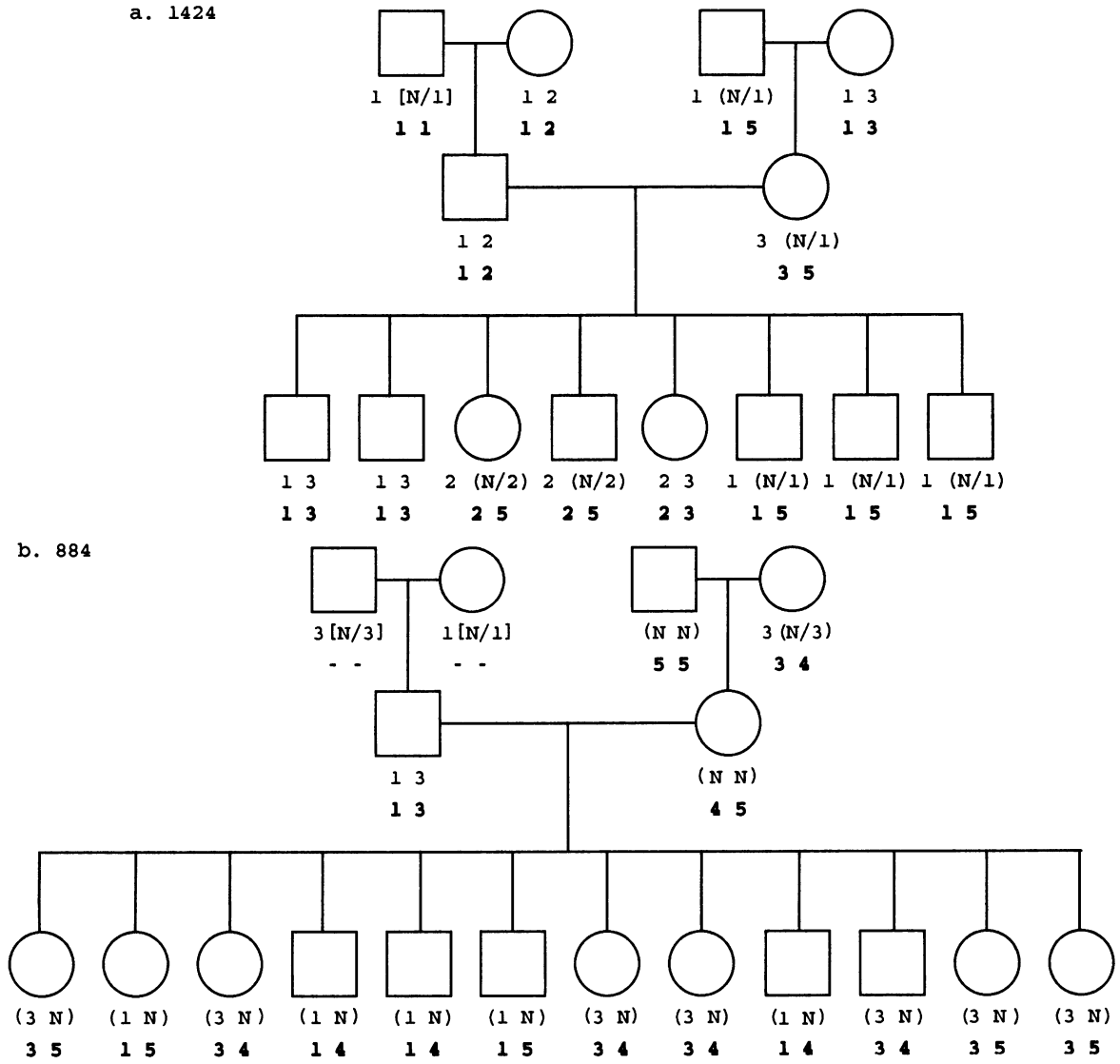

Figure 1 CEPH pedigrees 1424 (*a*) and 884 (*b*) genotyped for *D16S297*. The upper row of genotypes corresponds to those obtained using the reverse primer R1. Inferred genotypes involving null alleles are indicated in parentheses, beside the marker phenotype. Genotypes which could potentially involve null alleles are indicated by square brackets beside the marker phenotype. The lower row of genotypes was scored using the R2 reverse primer which enabled detection of all null alleles. The alleles correspond to the following size classifications: 1, $(AC)_{15}$; 2, $(AC)_{13}$; 3, $(AC)_{12}$; 4, $(AC)_{15}\Delta 8$ bp; 5, $(AC)_{14}\Delta 8$ bp; and N, null.

step-cycle file as described elsewhere (Thompson et al. 1992), except that, for sequencing, dNTPs were at 7.5 μM, and $MgCl_2$ was at 4.5 mM. Purified product, primer (120 ng of 24mer), and 1.5 U *Taq* polymerase (Perkin Elmer Cetus) were divided into four 20-μl sequencing reactions, to each of which was added either 2.5 mM ddATP, 1.0 mM ddCTP, 0.5 mM ddGTP, or 5.0 mM ddTTP. Five microliters of each reaction was mixed with 4 μl formamide loading buffer and electrophoresed on a 6% wedge denaturing polyacrylamide gel, and the result was visualized by autoradiography.

Results

Table 1 shows the 14 (AC)n repeats analyzed on the short arm of chromosome 16 and the 9 analyzed on the

long arm. The segregation of a null allele in a sibship was recognized by the apparent noninheritance, by a sib, of a parental allele (fig. 1). Of the 23 (AC)n repeats studied, 7 (30%) could be demonstrated to have null alleles. In four cases there was only a single CEPH sibship, of the 40 CEPH studied, where a null allele could be detected, while in the other three cases at least 30% of the CEPH sibships were shown to have a null allele segregating.

Where only a single CEPH sibship demonstrated the presence of a null allele, the number of sibs for whom a null allele was inferred to preserve Mendelian inheritance was four (*D16S287*), one (*D16S294*), four (*D16S304*), and five (*D16S305*). For each sibship the typing of the DNA samples was repeated, and the same results were obtained. The possibility of error due to incorrect DNA samples was eliminated, since the same CEPH sibship DNA samples have been used to determine the genotypes of all the (AC)n repeats listed in the table, and these data were all consistant with Mendelian inheritance.

There was no apparent relationship between the sequence or position of the oligoprimers flanking the (AC)n repeats and the presence of a null allele. For those (AC)n repeats with no demonstrated null allele, oligoprimers were constructed ranging from immediately adjacent to either side of the repeat (*D16S303*) to more than 20 bp either side of the repeat (*D16S296*). For those (AC)n repeats with a demonstrated null allele, the distance of the primers from the repeat was (presented as forward and reverse) 1 and 55 (*D16S67*), 2 and 69 (*D16S79A*), 14 and 104 (*D16S287*), 20 and 5 (*D16S294*), 32 and 53 (*D16S297*), 34 and 13 (*D16S304*), and 55 and 1 (*D16S305*).

The frequency of null alleles will be underestimated, since there may be no child within the pedigree who has the appropriate genotype to enable detection of a null allele. Not all 40 of the CEPH sibships were typed for each marker, with the families not typed being discarded as noninformative on the basis of parental genotypes. Since these apparently homozygous parents may in fact be heterozygous for a null allele, this will also underestimate the allele frequency.

Except for those markers with an appreciable frequency of null alleles, there is good agreement between expected and observed heterozygosity; hence population subdivision can be excluded as a contributor to the depression of heterozygosity observed for *D16S67*, *D16S79A*, and *D16S297*. Several CEPH sibships are known to be related, but this would have little effect on the calculation of expected heterozygosities. Where a null allele was present at an appreciable frequency, this was reflected in a lower observed heterozygosity, compared with that calculated from allele frequencies (table 1). For the (AC)n repeat at *D16S297*, 16 of the CEPH sibships were shown to be segregating a null allele. One of these sibships (sibship 884) had individuals homozygous for the null allele and therefore did not generate a PCR product (fig. 1*b*). This was investigated further by constructing an additional reverse oligoprimer—flanking the original reverse oligoprimer—to generate a larger product. Use of this reverse primer enabled (1) a PCR product to be detected for these null alleles and (2) two new alleles to be detected (table 1 and fig. 1). Typing of those CEPH sibships with suspected null alleles increased the observed heterozygosity from 45% to 81% (table 1), which is comparable to the expected frequency of 74%.

For the *D16S297* locus there were individuals who initially were homozygous for a null allele but who, with amplification using the additional reverse primer, were shown to be homozygous for one of the two new alleles. A homozygote for each of these new alleles was selected, and DNA was amplified by PCR using the additional reverse primer and was sequenced. For each individual there was an 8-bp deletion, and 4 bp within this deletion were part of a small direct repeat present in the undeleted sequence (fig. 2). This 8-bp deletion was located at the 3′ end of the sequence from which the first reverse primer was designed. This could either prevent primer extension or cause the reverse primer to fail to bind, resulting in absence of a PCR product, which would therefore be scored as a null allele. This deletion was associated with two alleles with different numbers of the (AC) repeat—i.e., alleles $(AC)_{14}\Delta 8$ bp and $(AC)_{15}\Delta 8$ bp—with frequencies of .03 and .22, respectively. The alleles which had dinucleotide-repeat number equivalent to these but which did not have the 8-bp deletion—i.e., alleles $(AC)_{14}$ and $(AC)_{15}$—had frequencies of .39 and .04, respectively.

Discussion

This study of (AC)n microsatellite repeats on chromosome 16 demonstrates that null alleles can commonly occur. In the one case investigated in detail, the null allele was shown to be the result of an 8-bp deletion occurring within the sequence chosen as a priming sequence for PCR amplification of the (AC) repeat. This problem was overcome by synthesis of a new oli-

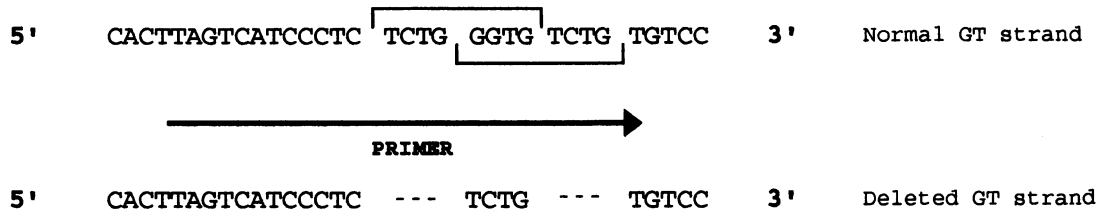

Figure 2 Sequence of null alleles at *D16S297*, showing 8-bp deletion at site of reverse primer, i.e., R1. The two possible sites of the deletion on the intact strand are indicated in the upper sequence. The sequence used for the R1 primer is indicated on the lower sequence.

goprimer which did not include the site of this deletion. In the case of *D16S67* the forward primer includes the poly(A) sequence adjacent to an Alu repeat, and mutations in this sequence are likely to be responsible for the null allele. Examples of oligoprimers which fail to amplify alleles in some individuals have been previously noted for *D16S287* (Phillips et al. 1991*b*) and for *IL9* on chromosome 5 (Weber et al. 1991) and are likely to be due to a mechanism similar to that detailed in the present report. Alleles have been reported where there is inhibition of the PCR reaction, resulting in faint bands—e.g., *D16S186* (Phillips et al. 1991*a*) and *D14S34* (Weber et al. 1991). These cases are likely to be the result of polymorphisms within the binding site, which merely inhibits, rather than completely prevents, the amplification by PCR.

A null allele segregating in a single CEPH sibship was demonstrated for four of the (AC)n repeats on chromosome 16. For each of these (AC)n repeats, all other CEPH sibships typed were consistent with Mendelian inheritance. The same DNA samples were used for the determination of genotypes for all the loci in table 1. Therefore, it is unlikely that the genotypes interpreted as null alleles were due to technical factors. Further evidence for a null allele at *D16S287* has been described elsewhere (Phillips et al. 1991*b*). An additional pair of oligoprimers flanking this (AC) repeat were constructed and used to amplify the alleles in the family where segregation of a null allele was suspected. Mendelian inheritance was then observed, confirming the presence of a null allele.

Either in gene mapping or in application to diagnosis, the presence of a segregating null allele will not corrupt the linkage data. If undetected, a null allele will merely result in that individual being scored as a homozygote, resulting in loss of informativeness. The presence of a null allele at an appreciable frequency can be suspected (within a homogeneous population) when the observed heterozygosity is markedly less than the expected heterozygosity. The presence of an undetected null allele may generate misleading data when individuals are haplotyped to determine the presence of linkage disequilibrium of an (AC)n repeat with a disease gene.

Within a pedigree, vertical transmission of a null allele through apparent homozygotes can result in an individual's genotype being apparently inconsistent with classical Mendelian inheritance. In such cases an explanation of nonpaternity is often invoked, but, if the genotypes have been determined on the basis of PCR amplification, then the presence of a null allele should be considered as an alternative explanation. Synthesis of alternative oligoprimers should alleviate such problems.

Recently, Chakraborty et al. (1992) have suggested that, for forensic applications in DNA typing, PCR-based polymorphisms provide the ideal systems, since use of VNTRs can result in underestimation of heterozygotes, because of nondetection of small alleles on Southern blots. In view of the occurrence of null alleles in PCR-based systems, this recommendation should be approached with caution.

Acknowledgments

This work was supported by U.S. Department of Energy contract DEFG0289ER60863.

References

Aksentijevich I, Pras E, Gruberg L, Shen Y, Holman K, Helling S, Prosen L, et al. Refined physical mapping of the gene causing familial Mediterranean fever: a comparison of likelihood methods with homozygosity mapping (submitted)

Chakraborty R, de Andrade M, Daiger SP, Budowle B (1992) Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann Hum Genet 56:45–57

Murray V (1989) Improved double-stranded DNA sequenc-

ing using the linear polymerase chain reaction. Nucleic Acids Res 17:8889

Phillips HA, Harris P, Richards RI, Sutherland GR, Mulley JC (1991*a*) Dinucleotide repeat polymorphisms at the D16S164, D16S168 and D16S186 loci at 16q21-q22.1. Nucleic Acids Res 19:6964

Phillips HA, Hyland VJ, Holman K, Callen DF, Richards RI, Mulley JC (1991*b*) Dinucleotide repeat polymorphism at D16S287. Nucleic Acids Res 19:6664

Shen Y, Thompson AD, Holman K, Callen DF, Sutherland GR, Richards RI (1992) Four dinucleotide repeat polymorphisms on human chromosome 16 at D16S289, D16S318, D16S319 and D16S320. Hum Mol Genet 1:773

Thompson AD, Shen Y, Holman K, Sutherland GR, Callen DF, Richards RI (1992) Isolation and characterisation of (AC)n microsatellite genetic markers from human chromosome 16. Genomics 13:402–408

Weber JL (1990) Informativeness of human (dC-dA)n-(dG-dT)n polymorphisms. Genomics 7:524–530

Weber JL, Kwitek AE, May AE (1990) Dinucleotide repeat polymorphisms at the D16S260, D16S261, D16S265, D16S266 and D16S267 loci. Nucleic Acids Res 18:4034

Weber JL, Polymeropoulos MH, May PE, Kwitek AE, Xiao H, McPherson JD, Wasmuth JJ (1991) Mapping of human chromosome 5 microsatellite DNA polymorphisms. Genomics 11:695–700