Am. J. Hum. Genet. 53:443–450, 1993

Characterization of the Patterns of Polymorphism in a "Cryptic Repeat" Reveals a Novel Type of Hypervariable Sequence

David P. Jacobson, Paula Schmeling, and Steve S. Sommer

Department of Biochemistry and Molecular Biology, Mayo Clinic/Foundation, Rochester, MN

Summary

Alternating purine and pyrimidine repeats (RY(i)) are an abundant source of polymorphism. The subset with long tandem repeats of GT or AC (GT(i)) have been studied extensively, but cryptic RY(i) (i.e., no single tandem repeat predominates) have received little attention. The factor IX gene has a polymorphic cryptic RY(i) of 142–216 bp. Previously, there were four known polymorphic alleles, of the form AB, A_2B, A_2B_2, and A_3B_2, where A = $(GT)(AC)_3(AT)_3(GT)(AT)_4$ and B = A with an additional 3′ AT dinucleotide. To further characterize this locus, we examined more than 1,700 additional human chromosomes and determined the sequences of the homologous sites in orangutans and chimpanzees. The novel alleles found in humans expand the repertoire of A/B alleles to $A_{0-4}B_1$ and $A_{1-3}B_2$. The A_nB_2 series are abundant in Caucasians but are absent in blacks and Asians. Conversely, the A_0B_1 allele is common in blacks but is not found in more than 1,700 Caucasian chromosomes. The data are compatible with a model in which recombination is more frequent than polymerase slippage at this locus. In orangutans, the RY(i) is present, but the sequence is markedly different. An A/B-type of pattern was discerned in which B differs from A by an additional six (AT) dinucleotides at the 3′ end. In chimpanzees, the size of the RY(i) locus was greatly expanded, and the sequence showed a novel pattern of hypervariability in which there are many tandem repeats of the form $(GT)_n(AC)_o(AT)_p(GT)_q(AT)_s$, where *n, o, p, q,* and *s* are different integers. The sequences of the factor IX intron 1 cryptic RY(i) in three primates provide perspective on the range of possible patterns of polymorphism. Analysis of the patterns suggests how the RY(i) can be conserved during evolution, while the precise sequence varies.

Introduction

GT or AC dinucleotide tandem repeats of 26 bp or more, (GT(26+)), are likely to be polymorphic (Weber 1990). GT(i) are capable of forming zDNA, which is recombinogenic in eukaryotes (Rich et al. 1984; Treco and Arnheim 1986; Wahls et al. 1990). However, the predominant mechanism generating polymorphism is thought to be polymerase slippage, because flanking markers do not exchange when alleles mutate and because common alleles tend to differ by only one or two repeat units (Weber and May, 1989; Luty et al. 1990; Hauge et al. 1991; Morral et al. 1991).

GT(26+) are a subset of a class of repetitive elements with alternating purine and pyrimidine segments (RY(i)) (see Nomenclature subsection in Material and Methods). A search of primate GenBank sequences revealed that RY(26+) is enhanced 770-fold relative to chance expectation. Thirty-three percent of RY(26+) do not contain GT(26+), and, when G(26+) are present, the majority are flanked by at least 5 additional alternating purine and pyrimidine bases (6% would be expected by random chance), and it is not infrequent to find 12 or more flanking RY(i) (.05% expected). The relative frequency of RY(i) increases dramatically with length (Sarkar et al. 1991). For example, RY(40+) are enhanced 4 million–fold in primate GenBank sequences and at least 1 million–fold in other eukaryotes, including yeast. By contrast, RY(i) are decreased in bacteria.

Received December 17, 1992; final revision received March 30, 1993.

Address for correspondence and reprints: Steve S. Sommer, Department of Biochemistry and Molecular Biology, Mayo Clinic/Foundation, Rochester, MN 55905.

0002-9297/93/5302-0016$02.00

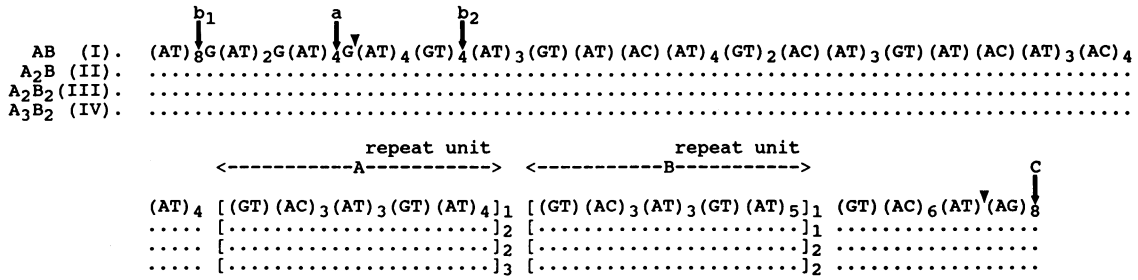


Figure 1 cRY(i) (between the arrowheads) and the repetitive flanking sequence, shown in a manner that indicates the organization of the repetitive sequence. The numbers after the brackets reflect the number of A and B segments in each polymorphic allele. The previously used roman-numeral designation of the alleles is shown. The lengths of alleles I–IV vary from 142 bp to 218 bp. Sites a, b_1, b_2, and c are sites of polymorphism (see the text and table 1). The polymorphisms found are $(AT)_2$ at site a, $(AT)_7$ and $(GT)_3$ at sites b_1 and b_2, respectively, and $(AG)_6$ at site c.

Some RY(i) lack long tandem identical repeats. We refer to these as "cryptic repeats" (cRY(i)) because the nature of the long tandem repeat is only appreciated when the sequence is categorized into purines and pyrimidines. Each of three cRY(i) examined by Sarkar et al. (1991) displayed multiple polymorphic alleles. As a source of polymorphism, cRY(i) have the practical advantage that PCR produces alleles of unambiguous size rather than the stuttered products associated with GT(i).

One cryptic repeat is in intron 1 of the factor IX gene. It was initially described as a *Hinf*I/*Dde*I polymorphism in which some individuals had an additional 50 bp of sequence (Winship et al. 1984; Yoshitake et al. 1985). By subsequent sequence analysis, two additional uncommon alleles were found, each differing in the number of A and B repeat units (fig. 1) (Sarkar et al. 1991). A is 24 bp, and B is identical except that an extra AT dinucleotide is added to the 3′ end. The alleles form a progression of A and B units: AB, A_2B, A_2B_2, and A_3B_2. Herein, we characterize this cryptic repeat in more detail and compare the human sequence with that of orangutans and chimpanzees. Most noteworthy are the discovery of a novel type of hypervariable locus and the evolutionary conservation of cRY(i) despite marked sequence differences between the species.

Material and Methods

Nomenclature

The frequency expected from random sequence of different patterns of purines and pyrimidines depends on the total length of the sequence. $(R)_{26}$, $(RY)_{13}$, and $(RRY)_{8.67}$ each have a length of 26 bp and an identical expected frequency (Sarkar et al. 1991). The adopted nomenclature refers to these sequences as "R(26)," "RY(26)," and "RRY(26)," respectively (Gostout et al. 1993). When these singlet, doublet, and triplet repeats are referred to in general terms, the terms "R(i)," "RY(i)," and "RRY(i)" are used.

The polymorphic repeat units found in the factor IX RY(i) are 24- and 26-bp sequence units referred to as "A" and "B." The number of repeat units of A and B are denoted by subscripts. Thus, A_2B_2 denotes two A units followed by two B units. A_nB_2 denotes an unspecified number of A units followed by two B units.

The RY(i) is extraordinarily polymorphic in chimpanzees (see Results). $(GT)_n(AC)_o(AT)_p(GT)_q(AT)_r$ denote n GT dinucleotide repeats, o AC dinucleotides, p AT dinucleotide repeats, etc., where n, o, p, q, and r are intergers that can vary.

PCR, GAWTS, and Factor IX Activity (FIX:C)

Genomic DNA was isolated from fresh or frozen blood according to a method described by Gustafson et al. (1987). PCR and direct genomic sequencing with GAWTS were performed as described elsewhere (Koeberl et al. 1990; Sommer et al. 1990). The sequenced non-Caucasian samples were from males (one X chromosome), except for samples from 1 black and 10 Asian females. FIX:C was measured by a one-stage clotting assay as described elsewhere (Bottema et al. 1990). Blood from male orangutans and chimpanzees was obtained from the Yerkes Regional Primate Research Center (Atlanta).

PCR and Sequencing Primers

The following primers were used:

A—I1(5719)-17U, 5′ TAC ATC TAC CTA GTC TG 3′;
B—I1(5337)-17D, 5′ GTC CAT TGA CCA AA 3′;

C—(T7-29)I1(5697)-45U, 5′ GGT ACC TAA TAC GAC TCA CTA TAG GGA GAG ACA CTC CTG AAC TCT 3′;
D—(Sp6-29)I1(5235)-46D, 5′ GGT ACC ATT TAG GTG ACA CTA TAG AAT ACA CTG TAG GCA ATG GTA A 3′;
E—(T7-29)I1(5781)-47U, 5′ GGT ACC TAA TAC GAC TCA CTA TAG GGA GAT GTT ATA GGA ATT GCA GT 3′;
F—I1(5353)-17D, 5′ AAT GTC ATT GTG CAG CA 3′;
G—I1(5256)-18D, 5′ GGT ATT TGT GTA TCA AAA 3′;
H—(SP6-23)I1(5337)-40D, 5′ ATT TAG GTG ACA CTA TAG AAT ACG TCC ATC ATT GAC CAA A 3′;
I—I2(6535)-17U, 5′ ACA TAA ATA TAT ATG GA 3′.

Primers A and B were used to amplify the cRY(i) in the initial screen. Primers B and C were used to amplify the cRY(i) and adjacent sequences from humans and chimpanzees. Primers D and E were necessary to amplify this region from orangutans. Primer F was used for sequencing with amplimers B and C. Primer G was used for sequencing with amplimers D and E. Primers A and H were used to amplify the cRY(i) region for sequencing in the opposite direction, and primer I was used as the sequencing primer. In chimpanzees, only partial sequence was obtained from the cRY(i), because it was too long to sequence in its entirety from the 5′ direction and because a flanking polypurine stretch caused polymerase stuttering, thereby precluding sequencing from the 3′ direction.

Results

Six New Polymorphic Alleles

Previously, four alleles were defined in the RY(i) in intron 1 of the factor IX gene by sequencing 76 Caucasian X chromosomes from individuals of western European origin. By contrast, only one allele was found in 30 X chromosomes from Asians of predominantly South Korean origin. To assess whether other alleles exist in western Europeans, 1,645 chromosomes from Caucasians were screened for sequence changes, by agarose gel electrophoresis. Any sample that migrated differently from the common AB and A_2B_2 alleles was sequenced.

About two-thirds of the patients in this sample (80% of the X chromosomes) were females. Females are particularly informative, because changes of only a few base pairs can be detected by altered migration of heteroduplexes. The bulge created by two or four extra bases on one strand places kinks in DNA and substantially affects migration (Bhattacharyya and Lilley 1989; Hsieh and Griffith 1989; Bhattacharyya et al. 1990). We confirmed this in our system by finding altered migration of heteroduplexes in samples heterozygous for a novel allele due to six rather than eight AG dinucleotides at site c, a polypurine stretch that directly adjoins the RY(i) (fig. 1). In total, 12 of about 670 Caucasian females (1% of chromosomes) were heterozygotes for $(AG)_6$.

Despite the large number of Caucasian X chromosomes screened, only one new allele, AB_2, was found in the A/B region of the cRY(i). When X chromosomes from African-Americans, Hispanics, and Asians were sequenced, the AB allele was found to be the most common allele in all groups. Three novel variants of the A/B region were found in blacks: B, A_3B, and A_4B (table 1). A_4B also was found in Hispanics, who are a mixture of Caucasians, blacks, and Native Americans. Thus the observed pattern of repeats is $A_{0-4}B$ and $A_{1-3}B_2$ (fig. 2). In addition, a variant AB allele, AB^a, was found in blacks. AB^a has two rather than four AT dinucleotides at site a (shown in fig. 1).

Some alleles are characteristic of particular races. Three (9%) of 32 black chromosomes contain the B allele, which is not found in 1,721 Caucasian chromosomes (1,645 screened chromosomes and 76 previously sequenced chromosomes). In addition, 27% of Caucasian chromosomes have the $A_{1-3}B_2$ series, compared with 0 of 92 black and Asian chromosomes. The presence of A_2B_2 in Hispanics presumably reflects the contribution of Caucasian genes.

RY(i) and FIX:C

Alternating purine and pyrimidine residues can form zDNA, which can relax supercoiled helices and mediate recombination (reviewed in Rich et al. 1984; Blaho and Wells 1989). In addition, RY(i) have been implicated in the regulation of gene expression (Naylor and Clark 1990; Wittig et al. 1991). To determine whether the RY(i) in intron 1 affects FIX:C, the cRY(i) genotype was determined for DNA from 45 Caucasian males without bleeding diatheses, in whom FIX:C had previously been determined (table 2). The averages of FIX:C were identical in Caucasians with the AB and A_2B_2 alleles.

RY(i) in Primates

In order to assess the evolutionary conservation of the cRY(i), the region from five male orangutans and six male chimpanzees was amplified by utilizing primers

Table 1

Polymorphic Alleles at the Factor IX Intron 1 RY(i)

Allele Type[a]	No. of Caucasians[b]	Hispanics	African-Americans	Asians[c]
B			3	
AB	1,154[d]	7	20	16 (46)[e]
AB[a]			1	
A_2B	23		6	1
A_3B			1	
A_4B^b		1	1	
AB_2	1			
A_2B_2	462[d]	5		
A_3B_2	5			
Total no. sequenced without screening	0[f]	13	32	17 (47)

[a] Superscripts denote alterations outside the A/B region. AB[a] is identical to AB, except that 4 bp $((AT)_2)$ are missing near the 5′ end of the RY(i) at site "a" (fig. 1). The A_4B^b allele lacks 2 bp (AT) and 2 bp (GT) near the 5′ end of RY(i) at sites "b1" and "b2," respectively (fig. 1).

[b] A total of 1,645 chromosomes were screened by PCR. If there was any question about altered mobility (76 chromosomes), the sample was sequenced. Forty-six of the chromosomes had the common AB or A_2B_2 alleles. Many of the sequenced AB alleles had a 2-bp deletion 3′ of the RY(i) (see fig. 1).

[c] From China, Japan, Philippines, and India.

[d] Estimated by determining the fraction of AB and A_2B_2 alleles in a random subsample of the Caucasian chromosomes.

[e] The 30 Asians previously sequenced (Sarkar et al. 1991) had AB. Thus, a total of 47 Asian alleles were examined.

[f] In this study, 1,645 Caucasian chromosomes were screened for size differences. In the previous study (Sarkar et al. 1991), 76 consecutive Caucasian chromosomes were sequenced without screening.

designed from flanking human sequences. The orangutan sequence contained a long cRY(i), but neither the precise sequence nor the order of the dinucleotides resembled the human sequence. It is interesting that an A/B type of pattern occurs in which B is identical to A, except for an additional six (AT) dinucleotides at the 3′ end (fig. 3, *underline*).

In chimpanzees, the amplified region was large and variable in size. Sequence from the 5′ end revealed a long, highly polymorphic cRY(i), which contains repeat units of the form $(GT)_n(AC)_o(AT)_p(GT)_q(AT)_r$, where *n*, *o*, *p*, *q*, and *r* are integers (fig. 4 and table 3). There are some imperfect repeat units that differ by a transition and, in one case, by a transversion, which disrupts the RY(i) (table 3). In one chimpanzee, unambiguous sequence could not be obtained, because of somatic mosaicism.

Discussion

Screening of the factor IX intron 1 cRY(i) in Caucasians revealed a total of five alleles. Additional alleles that differed by a few base pairs might have been missed in hemizygous males, but this is much less likely in females, in whom heteroduplexes between a rare and a common allele generally will migrate aberrantly on agarose gels, as discussed in Results.

An analysis of the factor IX cRY(i) in different racial and ethnic groups revealed multiple novel alleles. Alleles of the form $A_{0-4}B$ and $A_{1-3}B_2$ were found. AB is the most frequent in all ethnic groups analyzed. B is relatively common in blacks (9%), yet it was not found in 1,721 Caucasian chromosomes. Conversely, the $A_{1-3}B_2$ set of alleles is found in Caucasians but not in Asians or blacks. Precise estimates of the rarity of A_nB_2 alleles in blacks will be difficult to attain from our sample of African-Americans, because of some admixture of this population with Caucasians. FIX:C in Caucasians was independent of whether an individual had AB or A_2B_2, suggesting that factor IX gene expression is generally unaffected by the cRY(i) genotype. Although other interpretations are possible, the pattern of polymorphism in humans invites speculation that homologous recom-

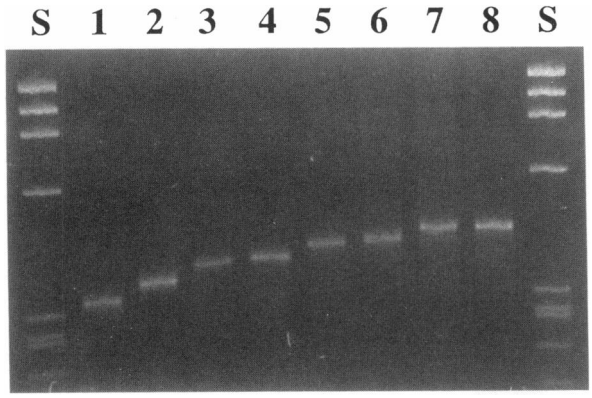
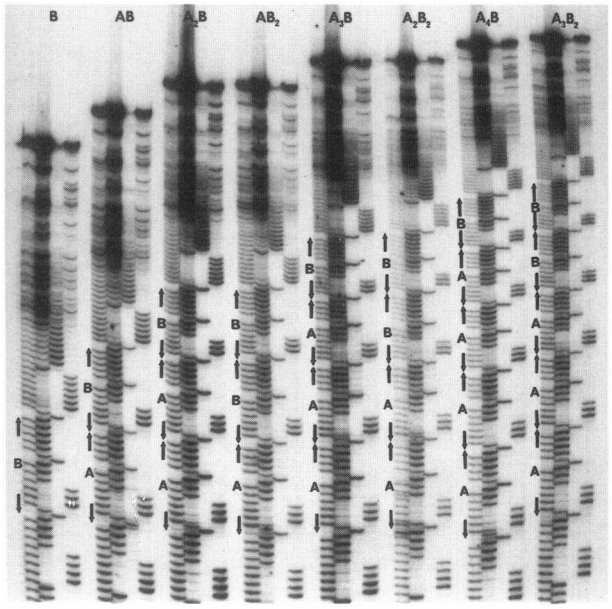


Figure 2 Amplification of the polymorphic $A_{0-4}B$ and $A_{1-3}B_2$, which yields segments of discrete size—in contrast to the common microsatellite repeats, GT(i). *Above*, Agarose gel electrophoresis of PCR. Expected PCR sizes plus promoters are in parentheses. Lanes S, Molecular-weight standards produced by *Hae*III digestion of ΦX174 DNA (sizes of markers, from top to bottom, are 1,353, 1,078, 872, 603, 310, 281, 271, and 234 bp, respectively). Lane 1, Allele B (338 bp). Lane 2, Allele AB (362 bp). Lane 3, Allele A_2B (386 bp). Lane 4, Allele AB_2 (388 bp). Lane 5, Allele A_3B (410 bp). Lane 6, Allele A_2B_2 (412 bp). Lane 7, Allele A_4B (430 bp). Lane 8, Allele A_3B_2 (436 bp). *Right*, Sequence of the alleles.



bination (which includes sister-chromatid exchange, which does not change flanking markers) and polymerase slippage are the dominant mechanisms that generate the observed polymorphisms. Recombination of the primordial AB allele can produce alleles of the form A_nB_1, since B is A with a 3′ AT dinucleotide (fig. 5*A*). However, polymerase slippage is necessary to convert a B_1-type allele to a B_2-type allele. Polymerase slippage seems the less common event, since (1) the A_nB_3 and A_nB_4 sets of alleles were *not* observed, yet they are readily produced by polymerase slippage (fig. 5*B*); (2) A_nB_2 alleles arose from a relatively recent and uncommon event, since these alleles are restricted to Caucasians; and (3) the minor AB_2 and A_3B_2 alleles found in Caucasians are compatible with those produced by homologous recombination with the common A_2B_2 allele.

The cRY(i) is present in orangutans and chimpanzees. The primate sequences show marked interspecies differences and a high degree of polymorphism. Orangutans have an A/B type of pattern in which B differs from A only in the number of (AT) dinucleotides at the 3′ end. In addition, polymorphisms arise by presumptive polymerase slippage and deletion events. The chimpanzee cRY(i) is much larger than the others. The alleles

```
Gelar          (AT)4 (AC)  (AT)  [G(AT)2]2 (AT)2  (GT)4 (AT)4  (GT)  (AT)2
Chatek         ..........................................*(AT)5..........
Solok,
Molek, &
Kichil         ........................(AC)(AT)..............  --    --

Gelar          (AC)  (AT)3 (GT)2 (AT)4 (GT)5 (AT)3 (GT)  (AT)2 (AC)  (AT)3
Chatek         ..........................................................
Solok,
Molek, &
Kichil          --    --    --    --  ..............................*(AT)2

Gelar          (GT)2 (AT)4 (GT)5 (AT)9 (AC)3 (AG)10 (164)
Chatek         ................*(AT)10.............(168)
Solok,
Molek, &
Kichil         ................*(AT)8..............(134)
```

Figure 3 Alignment of the five orangutan sequences, listed by individual names. Areas of polymorphism due to presumptive polymerase slippage at AT dinucleotide are indicated by asterisks (*). A deleted segment is indicated by double dashes (– –).

Table 2

FIX:C Associated with RY(i)

Allele	No.	Average Age (years)	Average FIX:C (%)
AB	30	63.9	124.4
A_2B_2	15	67.1	125.8

NOTE.—The average FIX:C for young adults is 100%, but it increases with age (Meade et al. 1977; Brandt et al. 1990). The age distributions of individuals with AB and A_2B_2 alleles are similar (data not shown).

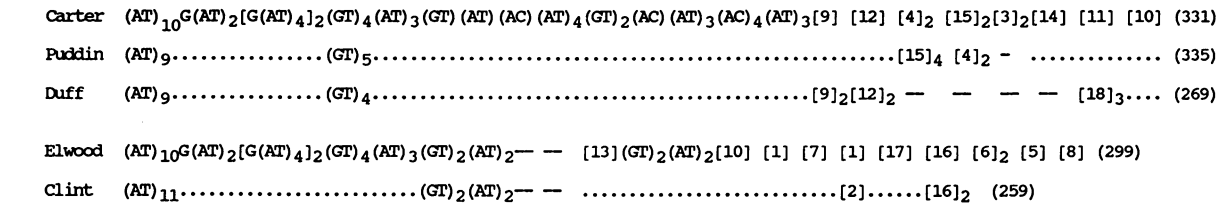


Figure 4 Alignment of the five chimpanzee sequences. The $(GT)_n(AC)_o(AT)_p(GT)_q(AT)_r$ repeat units found at the 3′ half of the sequence are numbered as indicated in table 3. Two subgroups are set apart by a skipped line. Within the upper subgroup, three repeats do not align perfectly. Repeat units 15 and 4 differ by a T→C transition. The same is true of repeat units 11 and 18.

are so variable that the locus is likely to undergo a very high rate of germ-line mutation. In addition, some somatic instability is suggested by the somatic mosaicism found in one of six male chimpanzees.

The hypervariable chimpanzee cRY(i) shows properties in common with microsatellites and VNTRs. Like microsatellites, the sequence can be expressed in the form of a short tandem repeat. Like VNTRs, the polymorphic sequence has a repeat unit of moderate size (approximately 20 bp; see table 3). However, the cRY(i) repeat is *not* composed of one predominant sequence; rather, the repeat is a constellation of sequences of the form $(GT)_n(AC)_o(AT)_p(GT)_q(AT)_r$. The hypervariability observed in chimpanzees is associated with a substan-

Table 3

Repeat Units in the Chimpanzee cRY(i)

Repeat Number	No. of Tandem Repeats (GT)	(AC)	(AT)	(GT)	(AT)	Length (bp)	Occurrences in Chimpanzees[a]
Perfect units:							
1	1	1	3	1	3	18	3
2	1	1	3	1	4	20	1
3	1	1	3	2	3	20	2
4	1	1	4	2	3	22	4
5	1	2	2	1	3	18	1
6	1	2	2	2	3	20	2
7	1	2	2	3	3	20	2
8	1	2	3	1	2	18	1
9	1	2	3	2	2	20	4
10	1	2	3	2	3	20	5
11	1	2	4	1	3	22	2
12	1	2	4	2	3	24	3
13	1	3	3	2	2	22	2
Imperfect units:							
14	1	1	4	1	3[b]	20	2
15	1[c]	1	4	2	3	22	7
16	1	2	2[d]	2	1[e]	32	3
17	1[c]	2	3	1	3	20	2
18	1	2	4[f]	1	3	22	3

[a] No. of times the repeat unit appears in the sequence of the five chimpanzees (fig. 4).

[b] $T^2 \rightarrow C$, $T^4 \rightarrow C$; i.e., transitions at base numbers 2 and 4 of the AT repeat yield the sequence $(AC)_2(AT)$.

[c] $T^2 \rightarrow C$.

[d] $A^1 \rightarrow C$; this transversion disrupts the RY(i).

[e] $(AT)_3(GT)(AC)_2(AT)_3$ substitutes for (AT).

[f] $T^8 \rightarrow C$.

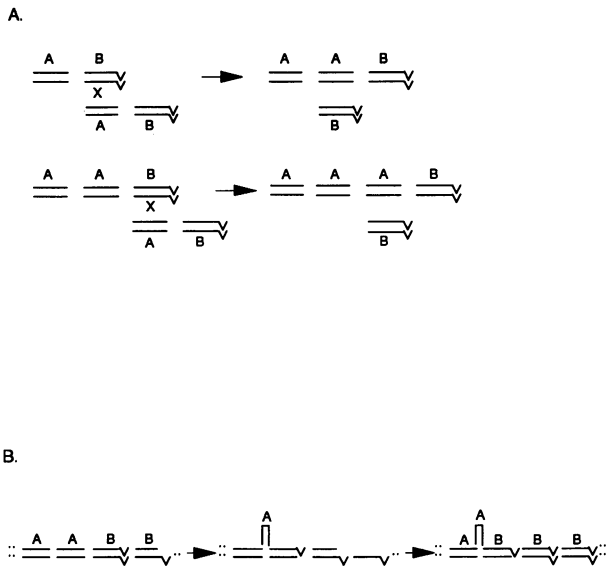


Figure 5 *A*, Examples of recombination, to illustrate how the AB primordial allele in humans can give rise to A_nB by recombination. The 3′ AT dinucleotide that distinguishes a B unit from an A unit is symbolized by vee (v). Out-of-register AB × AB recombination could occur by sister-chromotid exchange or by exchange between heterologous chromosomes. In the former case, flanking markers would not be exchanged. The B and A_3B alleles were found in the present study. Recombination cannot generate A_nB_2 alleles from AB_1. However, once one B_2-type allele is formed, recombination can generate the series of A_nB_2 alleles but not A_nB_3 or A_nB_4 alleles. *B*, Example of polymerase slippage, which illustrates how an A_2B_3 allele can be formed from A_2B_2. An A repeat "loops out" in the incompletely replicated strand to form an allele with five repeat units instead of four. Conversely, when the template strand loops out during replication, shorter products such as AB_2 are formed (not shown). Once A_2B_3 is formed, further polymerase slippage events can produce A_2B_4 and A_2B_5. Thus, polymerase slippage can generate many alleles, in addition to those generated by recombination.

tially longer cRY(i). This is reminiscent of the situation in myotonic dystrophy and fragile X, in which mutations occur at a relatively slow rate in individuals with "normal alleles" in which the number of CAG or CCG triplets is less than about 50 (reviewed in Caskey et al. 1992). However, once the repeat number is over 50, the mutation rate rapidly approaches 1 in the female germ line. In the present situation, the observed cRY(i) in humans and orangutans may be viewed as analogous to the normal polymorphic alleles, while chimpanzees have the larger, highly unstable form. If DNA could be obtained from families of chimpanzees, it should be possible to assess directly the stability of cRY(i) in the male and female germ lines.

In aggregate, these data suggest how a combination of polymerase slippage, recombination, point mutations, and deletions can alter the sequence but maintain a large stretch of RY(i) through evolution. If polymerase slippage and recombination are the most frequent types of mutation, then they will tend to maintain the cRY(i) over time. Occasional in-frame deletions will juxtapose two previously separated parts of the cRY(i). This will often produce a novel sequence, which can be further changed by subsequent polymerase slippage and recombination events. Likewise, a transition will generate a novel sequence of dinucleotides, which subsequently can be further modified by polymerase slippage and recombination events. Out-of-frame deletions and transversions will disrupt the cRY(i). However, if the remaining cRY(i) on either side is still susceptible to polymerase slippage and recombination, then the cRY(i) will persist.

Acknowledgments

We thank Harold McClure, D.V.M., and Eushin M. Lee at the Yerkes Primate Center for providing us with blood samples of orangutans and chimpanzees. We thank Mary Johnson for expert secretarial assistance. The work was supported by NIH grant HL39762.

References

Bhattacharyya A, Lilley DMJ (1989) The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). Nucleic Acids Res 17:6821–6840

Bhattacharyya A, Murchie AIH, Lilley DMJ (1990) RNA bulges and the helical periodicity of double-stranded RNA. Nature 343:484–487

Blaho JA, Wells RD (1989) Left-handed Z-DNA and genetic recombination. Prog Nucleic Acid Res Mol Biol 37:107–126

Bottema CDK, Koeberl DD, Ketterling RP, Bowie EJW, Taylor SAM, Lillicrap D, Shapiro A, et al (1990) A past mutation at isoleucine^{397} is now a common cause of moderate/mild haemophilia B. Br J Haematol 75:212–216

Brandt JT, Arkin CF, Bovill EG, Rock WA, Triplett DA (1990) Evaluation of APTT reagent sensitivity to factor IX and factor IX assay performance. Arch Pathol Lab Med 114:135–141

Caskey CT, Pizzuti A, Fu Y, Fenwick RG Jr, Nelson DL (1992) Triplet repeat mutations in human disease. Science 256:784–789

Gostout B, Liu Q, Sommer SS (1993) "Cryptic" repeating triplets of purines and pyrimidines (cRRY(i)) are frequent and polymorphic: analysis of coding cRRY(i) in the proopiome-

lanocortin (POMC) and TATA-binding protein (TBP) genes. Am J Hum Genet 52:1182–1190

Gustafson S, Proper JA, Bowie EJW, Sommer SS (1987) Parameters affecting the yield of DNA from human blood. Anal Biochem 165:294–299

Hauge XY, Grandy DK, Eubanks JH, Evans GA, Civelli O, Litt M (1991) Detection and characterization of additional DNA polymorphisms in the dopamine D2 receptor gene. Genomics 10:527–530

Hsieh C-H, Griffith JD (1989) Deletions of bases in one strand of duplex DNA, in contrast to single-base mismatches, produce highly kinked molecules: possible relevance to the folding of single-stranded nucleic acids. Proc Natl Acad Sci USA 86:4833–4837

Koeberl DD, Bottema CDK, Sarkar G, Ketterling RP, Chen S-H, Sommer SS (1990) Recurrent nonsense mutations at arginine residues cause severe hemophilia B in unrelated hemophiliacs. Hum Genet 84:387–390

Luty JA, Guo Z, Willard HF, Ledbetter DH, Ledbetter S, Litt M (1990) Five polymorphic microsatellite VNTRs on the human X chromosome. Am J Hum Genet 46:776–783

Meade TW, North WRS, Chakrabarti R, Haines AP, Stirling Y (1977) Population-based distributions of haemostatic variables. Br Med Bull 33:283–288

Morral N, Nunes V, Casals T, Estivill X (1991) CA/GT microsatellite alleles within the cystic fibrosis transmembrane conductance regulator (CFTR) gene are not generated by unequal crossingover. Genomics 10:692–698

Naylor LH, Clark EM (1990) $d(TG)_n \cdot d(CA)_n$ sequences upstream of the rat prolactin gene form Z-DNA and inhibit gene transcription. Nucleic Acids Res 18:1595–1601

Rich A, Nordheim A, Wang AH-J (1984) The chemistry and biology of left-handed Z-DNA. Annu Rev Biochem 53:791–846

Sarkar G, Paynton C, Sommer SS (1991) Segments containing alternating purine and pyrimidine dinucleotides: patterns of polymorphism in humans and prevalence throughout phylogeny. Nucleic Acids Res 19:631–636

Sommer SS, Sarkar G, Koeberl DD, Bottema CDK, Buerstedde J-M, Schowalter DB, Cassady JD (1990) Direct sequencing with the aid of phage promoters. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, New York, pp 197–205

Treco D, Arnheim N (1986) The evolutionary conserved repetitive sequence $d(TG \cdot AC)_n$ promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. Mol Cell Biol 6:3934–3947

Wahls WP, Wallace LJ, Moore PD (1990) The Z-DNA motif $d(TG)_{30}$ promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture. Mol Cell Biol 10:785–793

Weber JL (1990) Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. Genomics 7:524–530

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Winship PR, Anson DS, Rizza CR, Brownlee GG (1984) Carrier detection in haemophilia B using two further intragenic restriction fragment length polymorphisms. Nucleic Acids Res 12:8861–8872

Wittig B, Dorbic T, Rich A (1991) Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei. Proc Natl Acad Sci USA 88:2259–2263

Yoshitake S, Schach BG, Foster DC, Davie EW, Kurachi K (1985) Nucleotide sequence of the gene for human factor IX (anti-hemophiliac factor B). Biochemistry 24:3736–3750