Am. J. Hum. Genet. 65:1580–1587, 1999

Mutation Rates in Humans. II. Sporadic Mutation-Specific Rates and Rate of Detrimental Human Mutations Inferred from Hemophilia B

F. Giannelli, T. Anagnostopoulos, and P. M. Green

Division of Medical and Molecular Genetics, Guy's, King's, and St. Thomas' School of Medicine, Guy's Campus, London

Summary

We estimated the rates per base per generation of specific types of mutations, using our direct estimate of the overall mutation rate for hemophilia B and information on the mutations present in the United Kingdom's population as well as those reported year by year in the hemophilia B world database. These rates are as follows: transitions at CpG sites 9.7×10^{-8}, other transitions 7.3×10^{-9}, transversions at CpG sites 5.4×10^{-9}, other transversions 6.9×10^{-9}, and small deletions/insertions causing frameshifts 3.2×10^{-10}. By taking into account the ratio of male to female mutation rates, the above figures were converted into rates appropriate for autosomal DNA—namely, 1.3×10^{-7}, 9.9×10^{-9}, 7.3×10^{-9}, 9.4×10^{-9}, 6.5×10^{-10}, where the latter is the rate for all small deletion/insertion events. Mutation rates were also independently estimated from the sequence divergence observed in randomly chosen sequences from the human and chimpanzee X and Y chromosomes. These estimates were highly compatible with those obtained from hemophilia B and showed higher mutation rates in the male, but they showed no evidence for a significant excess of transitions at CpG sites in the spectrum of Y-sequence divergence relative to that of X-chromosome divergence. Our data suggest an overall mutation rate of 2.14×10^{-8} per base per generation, or 128 mutations per human zygote. Since the effective target for hemophilia B mutations is only 1.05% of the factor IX gene, the rate of detrimental mutations, per human zygote, suggested by the hemophilia data is ~1.3.

Received March 15, 1999; accepted for publication September 15, 1999; electronically published November 2, 1999.

Address for correspondence and reprints: Prof. F. Giannelli, Division of Medical and Molecular Genetics, 8th Floor, Guy's Hospital Tower, London Bridge, London SE1 9RT, United Kingdom. E-mail: francesco.giannelli@kcl.ac.uk

0002-9297/1999/6506-0013$02.00

Introduction

Data on spontaneous human mutation rates are still scarce. Locus mutation rates based on dominantly inherited diseases have been obtained (Vogel and Motulsky 1997), but dominant diseases present several difficulties. They often show incomplete, unpredictable penetrance. Some dominant diseases—for example, Huntington disease (MIM 14310)—arise by a distinct mutational mechanism: triplet expansion (The Huntington's Disease Collaborative Group 1993). Furthermore, in general, dominant mutations causing gain of function or dominant negative effects are likely to show narrow spectra and rather restricted intragenic targets. Mutation rates more representative of the changes responsible for the evolution of the bulk of the genome can be obtained from X-linked recessive diseases, which are particularly convenient because the natural X monosomy of the male allows detection of any recessive mutation. However, extrapolation from mutation rates in X-linked genes to those in the whole genome requires determination of the ratio of sex-specific mutation rates. Among the X-linked recessive diseases the hemophilias are ideal, because they are caused by mutations that do not result in balanced polymorphisms, and the mutation-selection equilibrium that has existed until recently ensures a substantial proportion of recent mutations in the hemophilia populations (Tuddenham and Giannelli 1994). Furthermore, as mentioned in the accompanying paper (Green et al. 1999 [in this issue]), full ascertainment of hemophilia patients is closely approached in countries where national patient registers have been established. Of the two hemophilias (MIM 30670 and MIM 30690), hemophilia B (MIM 30690) is the most useful at present, because the standard size of the factor IX gene allowed the development of satisfactory mutation-detection procedures 10 years ago, and this has led to the collection of abundant information on hemophilia B mutations. A world database of such mutations was assembled in 1990 and is updated annually (Giannelli et al. 1990, 1998).

We have conducted a national study of the United Kingdom's hemophilia B population that has allowed direct estimates of the overall and sex-specific mutation rates (Green et al. 1999). Here, we use the results of this

Table 1

Estimation of Hemophilia B Point Mutation Target Size in the Factor IX Coding Sequence

Point Mutation Type	No. of Possible Changes[a]	No. of Disease-Causing Changes Found				
		Up to 1997 (x)	In 1998 and 1999		Estimated Changes Not Seen up to 1997[b] (q)	Target Size[c]
			Previously Observed (y)	Not Previously Observed (z)		
Transitions:						
CpG	31	20	18	0	0	20
Non-CpG	892	185	50	26	96.2	281.2
Transversions [CpG transversions]	2,352 [65]	208 [19]	29 [1]	44 [1]	315.6 [19]	523.6 [38]
Total	3,275	413	97	70	411.8	824.8

[a] Nonsynonymous base substitutions only. The total number of base substitutions is 4,086, of which 120 are at CpG sites.

[b] Derived by use of the formula $q = xz/y$, where each letter corresponds to the appropriate value for each type of mutation in columns x, y, z, and q. Data in column x are from Giannelli et al. (1997), and those in columns y and z are from Giannelli et al. (1998) and Burks et al. (1999).

[c] Derived from $x + q$: it indicates the number of base substitutions expected to cause hemophilia B. Note that, since each base can undergo three different substitutions, a total target of 825 is equivalent to 275 bases.

study, plus information from the world database of hemophilia B mutations, to estimate the rates of different base substitutions and small deletions/insertions. In addition, we estimate mutation rates from interspecies sequence divergences, between humans and chimpanzees, that were observed in both X and Y chromosome sequences. These rates are compared with those obtained from hemophilia B. Finally, using the data on hemophilia B, we attempt to estimate the rate of detrimental mutations per human zygote.

Material and Methods

Identification of Hemophilia B Mutations and Determination of their Independent Origin

The methods used to characterize hemophilia B mutations and the criteria to determine their independent origin have been described in the accompanying paper (Green et al. 1999 [in this issue]).

Sequencing of Chimpanzee DNA and of a Reference Human DNA

Using the human genomic sequence available in the public databases, we have designed 42 PCR reactions of ~1.5 kb each in an Xq22 region spanning ~5 Mb (for details, see Anagnostopoulos et al. [1999]). These were arranged in seven clusters representing cosmid inserts. Each cosmid sequence contained six PCR reactions placed an average of 5 kb apart. Eleven similar but overlapping reactions were designed from unique Y-chromosome sequence (Whitfield et al. 1995). Twenty two X-chromosome and six Y-chromosome reactions were successful on both human and chimpanzee DNAs. The ends of the PCR products obtained by the above reaction in the chimpanzee's and a human male's DNA were sequenced with the PCR primers used in dRhodamine™ dye terminator cycle sequencing, according to the manufacturer's protocol (PE Biosystems).

Results

Type-Specific Mutation Rates

In the accompanying paper (Green et al. 1999 [in this issue]), we have obtained the overall and sex-specific rates for hemophilia B mutations, and have observed that the sex ratio for different types of mutations does not differ significantly. To determine the rates of specific types of mutations from the hemophilia B data, it is necessary to know the effective target size for the different hemophilia B mutations. This can be fairly accurately determined for hemophilia B mutations that are caused by base substitutions resulting in nonsense or missense mutations.

Since the coding sequence of the FIX gene specifies 39 amino acids of prepropeptide and 415 amino acids of circulating FIX protein, it contains 1,362 bp and can experience 4,086 different base substitutions. Of these, 811 are synonymous and 24 occur at CpG sites. The remaining 3,275 result in nonsense or missense mutations (table 1). The latter subdivide as follows: 31 transitions at CpG sites; 892 other transitions; and 2,352 transversions, of which 65 are at CpG sites. Not all these mutations can be expected to result in hemophilia B, and, to determine the fraction of disease-causing mutations, we use the following rationale. The 1997 world database of hemophilia B mutations, with 1,552 entries (Giannelli et al. 1997), gives the fullest list up to 1997 of different base substitutions causing hemophilia B mutations. Within the factor IX coding sequence this amounts to 20 transitions at CpGs, 185 other transitions, and 208 transversions, of which 19 are at CpG sites. The number of mutations of these types that re-

mained undetected up to 1997 is obviously proportional to the fraction of different mutations enlisted in the following 2 years (Giannelli et al. 1998; Burks 1999) that had not been observed previously. Thus, the estimate of the number of different mutations that remained undetected up to 1997 is given, for each class, by $q = xz/y$, where x is the number of different mutations in the 1997 database, y the number of different mutations reported to the database by 1998 and 1999 that had been seen previously, and z the number of mutations novel in these last 2 years of referrals (table 1). This approach could overestimate the target for hemophilia B mutations and thus could result in an underestimation of mutation rates per nucleotide if the submissions to the hemophilia B database in 1998 and 1999 were biased in favor of mutations that had not been previously observed. The likelihood of this event is small, because contributions to the database are always made year after year by the same set of national coordinators, who provide largely unpublished information on the precise understanding that all well-characterized mutations are equally valuable irrespective of whether they have been observed previously. The sum of q and x for the above groups of mutations indicates that the number of different missense or nonsense mutations that can cause hemophilia B should consist of 20 transitions at CpG sites, 281.2 transitions at other sites, and 523.6 transversions, of which 38 are at CpG sites, thus defining, in the coding sequence of the factor IX gene, a total target of 824.8 different point mutations severe enough to cause hemophilia B. This is a target equivalent to 275 nucleotides. The target size for small deletion/insertion mutations can also be determined. A convenient approximation is to consider the creation of a frameshift as the means by which these mutations cause hemophilia B. In a coding region of 1,362 nucleotides, there are as many positions at which frameshifts may occur, and these may be caused by deletions or insertions of one or two bases or multiples thereof. Therefore, frameshifts can occur at 1,362 positions because of a variety of mutations involving deletions, insertions, or both deletions and insertions.

If we now consider both the proportion of independent mutations of the above different types that were observed in the United Kingdom and the estimate of the overall mutation rate obtained from the same population, we can calculate the type-specific mutation rates per nucleotide per generation (table 2), using the formula $n_i\mu y_i/tx_i$ where n_i is 1 for transitions and small deletions/insertions, 2 for transversions, and 3 for all base substitutions, μ is the overall mutation rate, y_i the number of independent mutations of a specific type found in the United Kingdom in the factor IX coding sequence, t the total number of independent mutations observed in the United Kingdom, and x_i the number of events defining the target for each specific type of mutation. This formula partitions μ into its components by attributing to each mutation type a rate value proportional to its occurrence in the U.K. population (y_i/t). In addition, these rates are related to the hemophilia B effective target in nucleotide equivalents for each type of mutation (n_i/x_i). Thus, the mutation type–specific rates are expressed per nucleotide per generation, rather than per gene per generation as done for μ. The rates estimated are as follows: transitions at CpG sites, 9.73×10^{-8}; other transitions, 7.28×10^{-9}; transversions, 6.84×10^{-9}; all base substitutions, 2.10×10^{-8}; frameshifts, 3.19×10^{-10} and all mutations combined, 2.13×10^{-8}.

Table 2

Estimate of Type-Specific Mutation Rates based on Factor IX Coding Sequences

Mutation Type	Target Size (x)[a]	No. of Independent Mutations (y)[b]	Estimated Mutation Rate[c] ($\times 10^{10}$)
Transitions:			
CpG	20	76	$\mu_1 = 973$
Non-CpG	281.2	80	$\mu_2 = 72.8$
Transversions:			
All	523.6	70	$\mu_3 = 68.4$
Non-CpG	485.6	66	$\mu_u = 69.6$
CpG	38	4	$\mu_v = 53.9$
All base substitutions	824.8	226	$\mu_4 = 210$
Small deletions/insertions causing frameshifts	1,362	17	$\mu_5 = 3.19$
All above types			$\mu_6 = 213$

[a] No. of different base substitutions expected to cause hemophilia B or of positions where frameshift mutations may occur through small deletions/insertions.

[b] Observed in the factor IX coding sequence of hemophilia B patients from the United Kingdom.

[c] The formula used for estimation is $n_i\mu y_i/tx_i = \mu_i$ (see text). $\mu = 7.733 \times 10^{-6}$; $t = 302$; $i = 1, 2, 3, 4, 5, 6, u$, or v. Note that μ_i is expressed per base per generation and that μ and t have been determined in the accompanying article (Green et al. 1999 [in this issue]).

From Hemophilia B to the Bulk of the Genome

To apply the above rates to the genome as a whole, one has to consider three additional facts. Since the X chromosome—and, hence, the factor IX gene—spends one-third of its life in the male, whereas the autosomes spend equal times in both sexes, it follows that X-linked genes will show a lower rate of mutation than the autosomes whenever the mutation rates are higher in the male.

For a ratio of male to female mutation rate $v/u = 8.64$ the autosome–to–X-chromosome mutation-rate ratio is $[3 \times (8.64 + 1)]/[2 \times (8.64 + 2)] = 1.359$. In addition, the rate of frameshifts in the coding region of the factor IX gene must be multiplied by 1.5, to account for the small deletions/insertions of 3 bp or multiples

thereof that do not cause frameshifts and that, because of their unpredictable phenotypic effect, were not considered in the context of hemophilia B. Of course, this assumes that deletions/insertions of 3 bp or multiples thereof are as likely to occur as those of 1 or 2 bp and their respective multiples (that cause frameshifts). In this way, an estimate for the rate of small deletions/insertions is obtained. Finally, one has to correct for the excessive representation of CpG sites, since we have previously noted evidence suggesting a fourfold enrichment of CpG sites in critical positions of the factor IX gene (Green et al. 1992). The present data offer the following evidence of CpG enrichment in the critical target for hemophilia B. The proportion of CpG-site substitutions capable of causing hemophilia B (58:96—see first columns of tables 1 and 2) is 2.5-fold greater than that at other sites in the factor IX coding sequence (766.8:3,179). Furthermore, the CpG-site substitutions that are expected to be able to cause hemophilia B represent 7% of the hemophilia B effective target (58:824.8), whereas the frequency of nucleotides forming CpG sites in our randomly chosen X-chromosome sequences (table 3) was 1.4% (194:13,690)—that is, five times lower. The randomly chosen X-chromosome sequences should be more representative of the human genome, because the coding sequence of the factor IX gene is exposed to strong selection against detrimental mutations. After the fivefold overrepresentation of CpG sites in the target site for hemophilia B mutations is corrected for and the other adjustments mentioned above are made, the following values are obtained for the autosomal rates of mutation per base per generation: 1.32×10^{-7} transitions at CpG sites, 9.88×10^{-9} transitions at non-CpG sites, 9.28×10^{-9} transversions, 2.08×10^{-8} base substitutions, 6.49×10^{-10} small deletions/insertions, and 2.14×10^{-8} small mutations combined, where the latter is an average weighted by target size.

Comparison of Mutation Rates Derived from Hemophilia B and Interspecies Sequence Divergence Data

Recently we have compared X- and Y-chromosome sequences from a man and a male chimpanzee (Anagnostopoulos et al. [1999]). The sequences comprised 13,960 bp from the X and 4,670 bp from the Y chromosome (table 3), representing short stretches of DNA at many randomly selected sites. The X-chromosome sequences showed 64 transitions (including 10 at CpG sites), 33 transversions and 4 small deletion/insertions, and the Y chromosome showed 41 transitions (including 9 at CpG sites), 16 transversions, and 6 small deletion/insertion mutations. It is worth noting that the ratio of transitions at CpG to those not at CpG sites is similar for the X- and Y-chromosome sequences. Logistic regression clearly showed that, whereas there are significant differences between transitions at and transitions not at CpG sites ($P < .0001$) and between transitions on the X and Y chromosomes ($P = .0014$), there is no significant effect of chromosome on the ratio of transitions occurring at CpG to those at other sites, as tested by the interaction between chromosomes and CpG/non-CpG sites ($P = .56$). This negates a particular and marked excess of CpG transitions in the male, relative to the female, germline.

From the sequence-divergence data, the following rates of mutations per base per year can be obtained for the X chromosome (table 3), if it is assumed that there has been 5 million years of separation between the two species (White et al. 1994) and, hence, a combined 10 million years of mutation accumulation during divergence: 5.15×10^{-9} transitions at CpG sites, 3.92×10^{-10} transitions at non-CpG sites, 2.36×10^{-10} transversions, 6.94×10^{-10} base substitutions, 2.86×10^{-11} small deletions/insertions, and 7.23×10^{-10} for

Table 3

Estimates of Type-Specific Mutation Rates for X and Y Chromosome Sequences, Based on Chimpanzee-to-Human Sequence Divergence

	X Chromosome			Y Chromosome		
Mutation Type	Target Size (bp)	No. of Mutations Observed	Estimated Rate[a] ($\times 10^{10}$)	Target Size (bp)	No. of Mutations Observed	Estimated Rate[a] ($\times 10^{10}$)
Transitions:						
CpG	194	10	$\mu_1 = 51.5$	78	9	$\nu_1 = 115$
Non-CpG	13,766	54	$\mu_2 = 3.92$	4,592	32	$\nu_2 = 7.00$
Transversions	13,960	33	$\mu_3 = 2.36$	4,670	16	$\nu_3 = 3.44$
All base substitutions	13,960	97	$\mu_4 = 6.94$	4,670	57	$\nu_4 = 12.2$
Small deletions/insertions	13,960	4	$\mu_5 = 0.286$	4,670	6	$\nu_5 = 1.29$
All mutations	13,960	101	$\mu_6 = 7.23$	4,670	63	$\nu_6 = 13.5$

[a] Mutation rates μ_i and ν_i are expressed per base per year; μ_4 and ν_4 are obtained by addition of μ_3 and ν_3, respectively, to the weighted average of μ_1 and μ_2 or ν_1 and ν_2; the time when species diverged is taken to be 5 million years ago. The above data are from Anagnostopoulos et al. (1999).

small deletions/insertions and base substitutions combined. Similarly, data on the Y chromosome (table 3) gave the following rates per nucleotide per year: 11.5×10^{-9} transitions at CpG sites, 7.00×10^{-10} transitions at non-CpG sites, 3.44×10^{-10} transversions, 1.22×10^{-9} base substitutions, 1.29×10^{-10} deletions/insertions, and 1.35×10^{-9} for small deletions/insertions and base substitutions combined. Furthermore, comparison of the sequence-divergence rates in the X and Y sequences suggests an average ratio of male to female mutation rates of 3.5 over the 10 million years of separate evolution. When this factor is taken into account, the rate of mutations for the autosomes or the bulk of the genome over the period of species divergence becomes 6.32×10^{-9} CpG transitions, 4.81×10^{-10} non-CpG transitions, 2.89×10^{-10} transversions, 8.51×10^{-10} base substitutions, 3.51×10^{-11} small deletion/insertions, and 8.87×10^{-10} all small mutation types combined per nucleotide per year. As mentioned earlier, the frequency of CpG sites in the above X-chromosome sequences is fivefold lower than that observed in the effective hemophilia B target. The above rates calculated from divergence data and those obtained from hemophilia B are in good agreement, particularly if one considers the uncertainties about the precise period of divergence and the most appropriate generation times for the two species over this time. When the type-specific mutation rates per base per generation that are calculated from the hemophilia B data are divided by the mutation rate per base per year calculated from the divergence of human and chimpanzee sequences, values are obtained that are compatible with the length of human generations (table 4).

Estimate of the Spontaneous Rate of Detrimental Mutations in Humans

Data on the rates of spontaneous mutations recently have been reviewed, and the question of whether the high spontaneous rate is a health risk has been raised (Crow 1997; Drake et al. 1998). These reviews underscore the difficulties presented by our species for estimation of the genomic rate of detrimental mutations, as well as the need for additional human data. We have used our study of hemophilia B in the United Kingdom to provide information on spontaneous-mutation rates. The use of a disease phenotype is particularly apt for consideration of the distinction between the general spontaneous rate of mutation and that of detrimental mutations, where "detrimental" is defined as a clinically detectable effect in the hemizygous male such that genetic fitness is likely to be decreased. Our data indicate that a human zygote with a genome of 6×10^9 bp should contain 128.4 mutations. However, only a small fraction of these mutations may be detrimental. This fraction can be roughly estimated by consideration of the nucleotide equivalent of the effective target for hemophilia B mutations and the nucleotide content of the factor IX gene. Since 824.8 different base substitutions in the factor IX coding sequence are expected to cause hemophilia B, and 3 substitutions may occur at any nucleotide, this effective target is equivalent to 275 nucleotides. Small deletions and insertions causing hemophilia B by frameshift mutation may occur at 1,362 nucleotides, but such mutations occur at a rate ~66-fold lower than that of base substitutions, and, therefore, the effective target for hemophilia B mutations affecting the coding region is [275 + (1,362/66)] = 295.6 nucleotides. Some hemophilia B mutations occur in the noncoding region of factor IX or are in-frame deletions/insertions or gross deletions. In the United Kingdom, these mutations represent 19.5%, of the total. Thus, the effective target for hemophilia B can be further increased by 19.5% to 353.2 nucleotides. Since the factor IX is 33,500 nucleotides long, only 1.05% of mutations occurring in the gene can be expected to cause hemophilia B. If one applies this figure to the whole genome, only 1.35 of the 128.4 mutations of a human zygote should be detrimental, according to our specific definition of this term.

Discussion

The spectrum of mutations observed in hemophilia B is not unlike that of sequence divergence between closely related species. Point mutations dominate, followed by small deletion/insertions and only a few gross deletions or other gross rearrangements. The latter are probably influenced by features of chromosome organization that are as yet poorly delineated, such as the distribution of replication start sites, the arrangement of chromosomes on the nuclear matrix, and the presence, extent, orientation, and identity of repeated sequences; it is therefore difficult to obtain estimates of their occurrence that may be truly representative of gross deletions, insertions, or inversions in the whole genome. More-reliable estimates are possible for point and small mutations. We have determined the rates of transitions at CpG and non-CpG sites, of transversions, and of small deletions/insertions found in hemophilia B, and we have used both these values and the ratio of male to female mutation rates to obtain rates relevant to the whole genome.

To arrive at the type-specific mutation rates for hemophilia B, we had to define the relevant biological targets and the numbers of independent mutations in our U.K. sample. Both tasks were performed by use of economical and conservative assumptions. We defined the target for hemophilia B mutations by using observational data, rather than theoretical inferences based on conservation and presumed relevance of changes in the factor IX sequence, and we considered as a single event identical mutations that could not be positively shown

Table 4

Comparison of Autosomal Type-Specific Mutation Rates Derived from U.K. Hemophilia B Population and Interspecies Divergence of X and Y Chromosome Sequences

	Type of Mutation					
	Transitions					
Data	CpG	Non-CpG	Transversions	All Base Substitutions	Small Deletions/ Insertions	All
A	1.32×10^{-7}	9.88×10^{-9}	9.28×10^{-9}	2.08×10^{-8}	6.49×10^{-10}	2.14×10^{-8}
B	6.32×10^{-9}	4.81×10^{-10}	2.89×10^{-10}	8.51×10^{-10}	3.51×10^{-11}	8.87×10^{-10}
A/B	20.9	20.5	32.1	24.4	18.4	24.0

Note.—Row A: estimates of rates per base per generation, from hemophilia B, adjusted for autosomes, as explained in text. Row B: estimates of rates per base per year, from divergence of X-chromosome sequences, adjusted for autosomes, as explained in text. Row A/B: quotient of rates shown in rows A and B; this should reflect length of current human generation, if the estimates in rows A and B are sufficiently coherent.

to be of independent origin. The latter may have resulted in slight underestimation. Our values are higher than the rates calculated by Ketterling et al. (1993) and by Sommer and Ketterling (1996). Those authors, however, relied on data in the literature, rather than attempting to estimate the overall mutation rate for hemophilia B; they do not take into account the ratio of male to female mutation rates when extrapolating from hemophilia B to the whole genome, and their definition of the mutational target for hemophilia B is less purely based on observation in humans than is ours.

The base-substitution rate that we calculate from the hemophilia B data is in reasonable agreement with evolutionary rates for synonymous base substitution and point mutations in pseudogenes (Li and Graur 1991; Crow 1993). Similarly, the type-specific mutation rates that we have obtained from our hemophilia B work are in keeping with the rates derived from our study of the divergence between X- and Y-chromosome sequences in humans and chimpanzees (Anagnostopoulos et al. 1999; present study). This would be expected if we have correctly defined the target of mutations able to cause hemophilia B and if this disease represents a good model for the estimation of mutation rates in man. Interestingly, we find clear evidence of overrepresentation of CpG sites in the effective target for hemophilia B. This contributes to the reported excess of CpG transition mutations in hemophilia B (Koeberl et al. 1989; Green et al. 1992; Giannelli et al. 1998). We have previously argued that the tendency toward "CpG suppression" characteristic of vertebrate DNA (Bird 1980) will be opposed by selection with a strength directly proportional to the phenotypic effect of mutations affecting this dinucleotide. This should result in preferential conservation of CpG sites at essential positions within genes (Green et al. 1992). X-linked genes are maximally exposed to this effect of natural selection, because of the monosomy for the X chromosome in the heterogametic sex. Thus, a relative enrichment for CpG sites in the critical target of genes, especially X-linked genes, should have been expected.

Our data on interspecies sequence divergence were not from genes that are present on both the X and Y chromosomes. Nevertheless, we have sought to ensure that the sequences were representative both of the respective chromosomes and of regions without unusual evolutionary constraints. This was done by the sampling of small DNA stretches at many haphazardly chosen locations in unique regions of the X and Y chromosomes.

Our data are in keeping with the concept of male-driven evolution (Miyata et al. 1987; Shimmin et al. 1993; Ellegren and Fridolfsson 1997). The ratio of sex-specific mutation rates calculated for hemophilia B, 8.64, is greater than the ratio of 3.5 obtained for the period of separate evolution of humans and chimpanzees, but, of course, the average paternal age of modern man is likely to be greater than that of his ancestors and the chimpanzee. The ratio of sex-specific mutation rates in hemophilia B is indeed fairly close to the ratio of germline cell division in human males and females of parental ages appropriate to our time (Drost and Lee 1995). This underscores the role of cell replication in germline mutagenesis but, of course, does not exclude the contribution of other differences in the biology of the male and female germlines to the definition of the sex-specific mutation rates.

Our estimate of the spontaneous rate of detrimental mutations in humans (1.35 per diploid) is in close agreement with the recent estimate of detrimental mutation rates in hominids, based on a comparison of synonymous and nonsynonymous rates in protein-coding regions of different species, made by Eyre-Walker and Keightley (1999). Those authors indirectly estimate the number of all mutations that could negatively affect fitness in evolving populations and suggest 4.2 amino acid–altering mutations per diploid per generation. On the basis of this estimate they suggest a rate of 1.6 deleterious mutations per diploid genome. They consider this as a minimal value, because mutations in noncoding regions are not accounted for, and they also take a conservative estimate of the number of genes in the human genome—namely, 60,000.

Our estimate is based on different premises, since we do not consider the number of genes in the genome and the level of conservation of several proteins in different species but instead argue from a direct estimate of the hemophilia B mutation rates, the effective target for such a phenotypically detectable trait relative to the full size of the factor IX gene, and the known size of the human genome. We take into account mutations in noncoding regions and all deletion/insertion changes and therefore do not feel that ours is an underestimate on this account. One could even argue that our estimate is likely to be a maximum estimate, because the factor IX gene has no functional homologues and therefore there is no redundancy of its function in the human genome. This probably does not apply to all human genes. A great variety of missense mutations cause hemophilia B, whereas mutations in many other human genes lead to abnormal phenotypes only when they are more disruptive; for example, mutations of the dystrophin (Roberts et al. 1994) or the adenomatous polyposis coli gene (Nagase and Nakamura 1993) mostly result in recognizable phenotypes when they cause premature termination of translation. The human genome is known to contain large regions of very low gene density (e.g., the centromeric region) where base substitutions may more frequently be neutral. However, sometimes structural features of some regions of the genome may result in unexpected incidence of gross rearrangements, as found, for example, in hemophilia A (Naylor et al. 1995), but such predisposing structures are expected to be rare. On the whole, our data suggest that a rate of detrimental mutations per human zygote of ~1 is highly plausible. Of course, the greater the rate of detrimental mutations, the greater is the problem of accounting for the survival of the species. Sexual reproduction and recombination provide much-needed flexibility in the face of selection, and Crow (1999), considering rates of detrimental mutations to be $\gg 1$, suggests that elimination of mutations in bunches has probably had a role in the maintenance of our species.

In conclusion, the study of hemophilia B mutations in the population of the United Kingdom has allowed the determination of "type-specific mutation rates" in the X chromosome and autosomes. These rates are in keeping with those which we obtain from consideration of sequence divergences in the X and Y chromosomes of humans and chimpanzees. The data from hemophilia B and those from sequence divergence are both in keeping with the concept of male-driven evolution. We also show that the target for hemophilia B mutations is rich in CpG nucleotides and suggest that this is because of selection acting against the tendency of CpG suppression that is characteristic of vertebrate DNA.

Finally, by considering the size of the target for hemophilia B mutations, relative to the whole factor IX gene, we provide a definition-based estimate of the rate of detrimental mutations per human zygote, operationally different from that generally used in investigations of genetic load, since it relies on extrapolation from a specific clinically detectable phenotype.

Acknowledgments

We are very grateful to the U.K. hemophilia-center directors and to the U.K. hemophilia B patients and their families for collaborating with our investigations of hemophilia B. We also wish to thank Drs. Jane Montandon and Gabriella Rowley and Ms. Samia Saad for their help during phases of our work and Adrienne Knight for secretarial help. We are also indebted to Dr. Cathryn Lewis for criticism of the manuscript and for statistical advice. This work was supported by Action Research and the Medical Research Council and by the Biotechnology and Biological Sciences Research Council.

References

Anagnostopoulos T, Green PM, Rowley G, Lewis CM, Giannelli F (1999) DNA variation in a 5-Mb region of the X chromosome and estimates of sex-specific/type-specific mutation rates. Am J Hum Genet 64:508–517

Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res 8:1499–1503

Burks C (1999) Molecular biology database list. Nucleic Acids Res 27:1–9

Crow JF (1993) How much do we know about spontaneous human mutation rates? Environ Mol Mutagen 21:122–129

——— (1997) The high spontaneous mutation rate: is it a health risk? Proc Natl Acad Sci USA 94:8380–8386

——— (1999) The odds of losing at genetic roulette. Nature 397:293–294

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutations. Genetics 148:1667–1686

Drost JB, Lee WR (1995) Biological basis of germline mutations: comparisons of spontaneous germline mutation rates among Drosophila, mouse and human. Environ Mol Mutagen 25:48–64

Ellegren H, Fridolfsson A-K (1997) Male-driven evolution of DNA sequences in birds. Nat Genet 17:182–184

Eyre-Walker A, Keightley PD (1999) High genomic deleterious mutation rates in hominids. Nature 397:344–347

Giannelli F, Green PM, High KA, Lozier JM, Lillicrap DP, Ludwig M, Olek K, et al (1990) Haemophilia B: database of point mutations and short additions and deletions. Nucleic Acids Res 18:4053–4059

Giannelli F, Green PM, Sommer SS, Poon M-C, Ludwig M, Schwaab R, Reitsma PM, et al (1997) Haemophilia B: database of point mutations and short additions and deletions, 7th edition. Nucleic Acids Res 25:133–135

——— (1998) Haemophilia B: database of point mutations and short additions and deletions, 8th edition. Nucleic Acids Res 26:265–268

Green PM, Montandon AJ, Bentley DR, Ljung R, Nilsson IM, Giannelli F (1990) The incidence and distribution of CpG→TpG transitions in the coagulation factor IX gene: a

fresh look at CpG mutational hotspots. Nucleic Acids Res 18:3227–3231

Green PM, Saad S, Lewis CM, Giannelli F (1999) Mutation rates in man. I. Overall and sex-specific rates obtained from a population study of haemophilia B. Am J Hum Genet 65: 1572–1579 (in this issue)

Huntington's Disease Collaborative Research Group, The (1993) A novel gene containing a trinucleotide that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Ketterling RP, Vielhaber E, Bottema CDK, Schaid DJ, Cohen MP, Sexaur CL, Sommer SS (1993) Germ-line origins of mutation in families with hemophilia B: the sex ratio varies with the type of mutation. Am J Hum Genet 52:152–166

Koeberl DD, Bottema CDK, Buerstedde J-M, Sommer SS (1989) Functionally important regions of the factor IX gene have a low rate of polymorphisms and a high rate of mutations in the oligonucleotide CpG. Am J Hum Genet 45: 448–457

Li WH, Graur D (1991) Fundamentals of molecular evolution. Science Associates, Sunderland, MA

Miyata T, Hayashida H, Kuma K, Mitsuyasa K, Yasunaga T (1987) Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harb Symp Quant Biol 52:863–867

Nagase H, Nakamura Y (1993) Mutations of the APC (adenomatous polyposis coli) gene. Hum Mutat 2:425–434

Naylor JA, Buck D, Green P, Williamson H, Bentley D, Giannelli F (1995) Investigation of the factor VIII intron 22 repeated region (*int22h*) and the associated inversion junctions. Hum Mol Genet 4:1217–1224

Roberts RG, Gardner RJ, Bobrow M (1994) Searching for the 1 in 2,400,000: a review of dystrophin gene point mutations. Hum Mutat 4:1–11

Shimmin LC, Chang BHJ, Li W-H (1993) Male-driven evolution of DNA sequences. Nature 362:745–747

Sommer SS, Ketterling RP (1996) The factor IX gene as a model for analysis of human germline mutations: an update. Hum Mol Genet 5:1505–1514

Tuddenham EGD, Giannelli F (1994) Molecular genetics of haemophilia A and B. In: Bloom A, Forbes CD, Thomas DP, Tuddenham EGD (eds) Haemostasis and thrombosis. Churchill-Livingstone, Edinburgh, pp 859–886

Vogel F, Motulsky AG (1997) Human genetics: problems and approaches. 3rd ed. Springer-Verlag, New York

White TD, Suwa G, Asfaw B (1994) *Australopithecus ramidus,* a new species of early hominid from Aramis, Ethiopia. Nature 371:306–312

Whitfield LG, Sulston JE, Goodfellow PN (1995) Sequence variation of the human Y chromosome. Nature 378: 379–380