Am. J. Hum. Genet. 64:508–517, 1999

DNA Variation in a 5-Mb Region of the X Chromosome and Estimates of Sex-Specific/Type-Specific Mutation Rates

Theodore Anagnostopoulos,* Peter M. Green,* Gabriella Rowley, Cathryn M. Lewis, and Francesco Giannelli

Division of Medical and Molecular Genetics, Guy's, King's College, and St. Thomas' Hospitals Medical and Dental School, London, United Kingdom

Summary

We describe a new approach for the study of human genome variation, based on our solid-phase fluorescence chemical mismatch–cleavage method. Multiplex screening rates ⩾80 kb/36-lane gels are achieved, and accuracy of mismatch location is within ±2 bp. The density of differences between DNA from any two humans is sufficiently low, and the estimate of their position is accurate enough, to avoid sequencing of most polymorphic sites when defining their allelic state. Furthermore, highly variable sequences, such as microsatellites, are distinguished easily, so that separate consideration can be given to loci that do and do not fit the definition of infinite mutation sites. We examined a 5-Mb region of Xq22 to define the haplotypes of 23 men (9 Europeans, 9 Ashkenazim, and 5 Pygmies) by reference to DNA from one Italian man. Fifty-eight 1.5-kb segments revealed 102 segregating sites. Seven of these are shared by all three groups, two by Pygmies and Europeans, two by Pygmies and Ashkenazim, and 19 by Ashkenazim and Europeans. Europeans are the least polymorphic, and Pygmies are the most polymorphic. Conserved allelic associations were recognizable within 40-kb DNA segments, and so was recombination in the longer intervals separating such segments. The men showed only three segregating sites in a 16.5-kb unique region of the Y chromosome. Divergence between X- and Y-chromosome sequences of humans and chimpanzees indicated higher male mutation rates for different types of mutations. These rates for the X chromosomes were very similar to those estimated for the X-linked factor IX gene in the U.K. population.

Received September 1, 1998; accepted for publication December 2, 1998; electronically published January 25, 1999.

Address for correspondence and reprints: Prof. F. Giannelli, Division of Medical and Molecular Genetics, Guy's, King's College, and St. Thomas' Hospitals Medical and Dental School, 8th floor, Guy's Hospital Tower, London, SE1 9RT, United Kingdom. E-mail: f.giannelli@umds.ac.uk

* These authors contributed equally to this work.

0002-929799/6402-0020$02.00

Introduction

Interest in the study of human genome variation is increasing as the Human Genome Project progresses toward the definition of a complete reference sequence (Olson 1995; Bentley 1996; Collins et al. 1997; Rowen et al. 1997). This sequence information provides new ways to tackle old problems, such as the role of selection and random variation in evolution (Kimura 1968), the origin of humans and their colonization of the Earth (Cavalli-Sforza et al. 1994), and the identification of genes involved in specific single-gene disorders. However, as the focus of genetic research turns to multifactorial disease (Lander and Schork 1994; Cordell and Todd 1995; Collins et al. 1997), a new problem arises: how to differentiate the multitude of trivial polymorphisms from variations that cause susceptibility or resistance to common diseases. To address this problem, a greater understanding of the way in which mutation, selection, recombination, and population history combine to determine the patterns of variation along human chromosomes is essential. This, in turn, requires development of procedures for the analysis of DNA variation over extensive chromosome regions, and of new methods of statistical inference.

Investigation of human DNA variation, spearheaded by students of human genetic history, has concentrated mainly on nonrecombining DNA elements, such as mitochondria and nonrecombining regions of the Y chromosome, or on autosomal regions so small that they have little chance of recombination (Cann et al. 1987; Dorit et al. 1995; Hammer et al. 1995; Jobling and Tyler-Smith 1995; Whitfield et al. 1995; Stoneking and Soodyall 1996; Harding et al. 1997; Zerjal et al. 1997). Absence of recombination greatly simplifies the study of genealogies, and fully developed methods of statistical inference are limited to this situation (Hudson 1990; Felsenstein 1988). Nevertheless, an understanding of chromosome organization and evolution cannot be achieved without consideration of recombination, and development of mathematical statistical methods capable of analyzing genealogies of recombining DNA

regions has begun (Bandelt et al. 1995; Griffiths and Marjoram 1996; Hey and Wakeley 1997). Data on the distribution of DNA variations in haplotype format from recombining regions of the human genome should form the empiric basis of the development of statistical methods capable of analyzing genealogies of recombining DNA regions. With these methods, it should become possible not only to investigate human genetic history, by use of regions of the genome that are not so constrained, with regard to segregation and selection, as the mitochondria and the nonrecombining regions of the Y chromosome, but also to bring the reconstruction of historical genetic events to bear on the interpretation of the functional consequences of specific DNA variations.

To contribute to this progress, we have developed a rapid, relatively inexpensive procedure to screen large DNA regions for sequence variations, and we have obtained a body of data in haplotype format. This was done by studying the X chromosome, because haploidy in men allows direct experimental definition of variation haplotypes. Our procedure for the study of DNA variation yields data akin to those obtained by sequencing, but much more rapidly and at lower cost, and has enabled us to report the results of a systematic analysis of DNA variation along the X chromosome in a sample of 24 men, comprising Europeans, Ashkenazim, and Pygmies. These results are contrasted with those on a small nonrecombining region of the Y chromosome, whereas the discrepancy of human X- and Y-chromosome sequences from the homologous sequences of the chimpanzee are used to estimate mutation rates.

Subjects and Methods

Subjects

The DNA samples examined were from 24 men and a male chimpanzee (National Institute of General Medical Sciences [NIGMS] cell strain GM03452). An Italian man was chosen as the human reference. The test-human samples were from three Britons, three Greeks, and three Swedes (collectively termed "Europeans"); nine Ashkenazim; and five Pygmies (two Biaka and three Mbuti). The use of these samples for the current work was approved by the ethical committee at Guy's, King's College, and St. Thomas' Hospitals Medical and Dental School.

DNA Extraction

DNA from the European and Ashkenazi men was extracted from blood lymphocytes, by standard procedures (Miller et al. 1988), whereas the Pygmy and chimpanzee DNA similarly was extracted from cell cultures obtained from NIGMS (code numbers GM10469A, GM10470A, GM10492A, GM10494A, GM10495A, and GM03452). The Pygmy cell cultures were lymphoblasts, and the chimpanzee cell cultures were fibroblasts.

Solid-Phase Fluorescent Chemical–Mismatch Cleavage

Italian reference ("probe") DNA was amplified, by use of biotinylated primers, and was hybridized to test DNA, and amplified with the same primers in the presence of one of three fluorescent dUTP, as described elsewhere (Rowley et al. 1995). For the present work, we elected to screen DNA in stretches of 1.5 kb, as this is within the useful range of the method, and the segments are not so large as to lead to several differences commonly occurring between any two randomly chosen sequences. Triplexes were then prepared, for modification of T and C residues at mismatches, with osmium tetroxide and hydroxylamine, respectively, and cleavage was then performed with piperidine. Thus, each reaction tube and lane of the gel examined ~4.5-kb DNA. These chemical treatments were performed as described elsewhere (Montandon et al. 1989), but were done on a solid phase created by binding the DNA to streptavidin-coated magnetic beads (Rowley et al. 1995). This allows rapid manipulation and eliminates time-consuming ethanol precipitation steps. Reacted DNA was mixed with size markers (Rox 2500, PE-Applied Biosystems), and the cleavage products were resolved and sized by electrophoresis in denaturing gels in the ABI 377 prism, by GENESCAN 2.0.2 (PE Applied Biosystems).

The position of each mismatch was determined accurately with the help of internal markers and by reference to the smaller of the two cleavage bands. DNA sequencing of PCR products was performed by use of dRhodamine dye-terminator–cycle sequencing, according to the manufacturer's protocol (PE Applied Biosystems). Accession numbers for the sequence information used in this study are: European Molecular Biology Laboratory (EMBL) Z68331, Z68871, Z69367, Z69721, Z69722, Z70224, Z70226, Z70227, Z70274, Z73900, Z73965, and Z75745; and Genbank X96421 and L35265.

DNA Variation in the X Chromosome

Using sequence information obtained at the Sanger Centre, we have designed 72 PCR reactions, of ~1.5 kb each, in a region of Xq22 spanning 5 Mb. These were arranged in 12 clusters representing cosmid inserts, selected at the intervals detailed in figure 1. Each cosmid contained six PCR reactions, placed ~4.5 kb from each other. Fourteen of the 72 PCR reactions did not amplify satisfactorily and were excluded, without replacement.

The 36 successful reactions in cosmids 2–8 and the European DNA samples were selected at the start of the study, to screen DNA variation in two identical but independent experiments, carried out on separate occa-

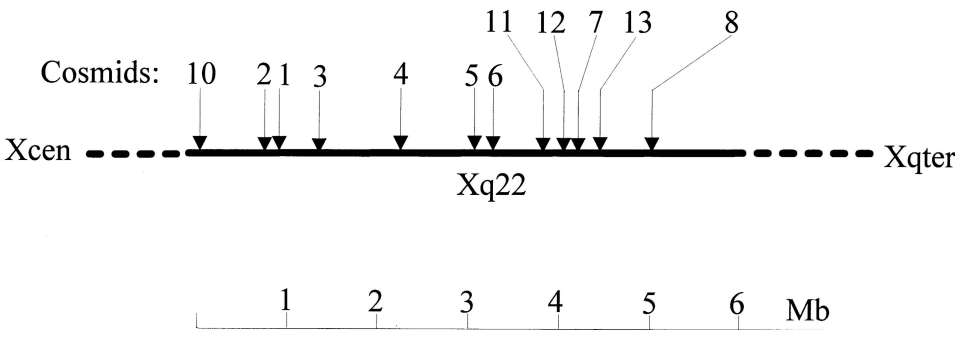

Figure 1 Map of the X-chromosome region analyzed. The position of cosmids chosen for the analysis of DNA variation is indicated with arrows.

sions, to confirm the expected reproducibility and specificity of the mismatch-detection procedure. Each experiment detected the same identical set of 72 mismatches as pairs of fast-running bands on ABI denaturing gels (fig. 2).

We obtained further confirmation of the reliability of the method by sequencing the 6 PCR products from cosmid 2 in our sample of 24 men, because this showed only 21 sequence changes that perfectly explained all of the observed mismatch bands (see table 1 and fig. 3). The position of the 21 mismatches estimated, prior to sequencing, from the size of the smaller of the two cleavage bands, differed on average from the exact position, later determined by sequencing, by $< \pm 2$ nucleotides (nt; table 1). This indicates that our procedure can resolve >150 different variants of a 1.5-kb segment, which is a number that is unlikely to be reached.

Since any difference between target and probe DNA resulting from transition, transversion, small deletion/insertion, or blockage of two or more base changes causes a mismatched and hence a cleavable heteroduplex, unequivocal distinction between these events can be achieved only by sequencing. However, different events at the same position should be very rare. Figure 2*B* shows that our procedure detects and hence can distinguish short tandem repeats and long monopurine or monopyrimidine tracks, even when sequence information is incomplete. These sequences frequently are associated with mutation rates and probabilities of back mutations in vivo much higher than those of most DNA (Brinkmann et al. 1998). In vitro, they cause PCR stuttering, which is detected by mismatch cleavage within segments, amplified from the hemizygous reference DNA as broad, diffuse bands. This bears witness to the sensitivity of the technique that detects mutant DNA, even when mixed with a 10-fold excess of reference sequence (A. Montandon, personal communication; Verpy et al. 1994). There is therefore a negligible risk of false-negative results. The risk of false positives is similarly small and is further reduced because genuine cleavages must produce two fragments that add to the size of the uncut segment.

Segregating sites in the segments corresponding to the 58 successful PCR reactions were identified by comparison of the reference Italian man with the 23 test men, and the most divergent pair of sequences analyzed showed, on average, <1 difference per 1.5 kb. Consequently, two human X chromosomes can be expected to differ only rarely >2 independent positions within any 1.5-kb segment. It follows that, in general, alleles at segregating sites can be identified without recourse to sequencing, thus restricting this slower and more cumbersome technique to exceptional circumstances, such as the presence of $\geqslant 3$ differences in a pair of sequences. Thus, we have screened, almost exclusively by using mismatch cleavage, 2 Mb (24 × 87 kb) of Xq22 sequence and have found a total of 102 segregating sites (fig 3). Some segregating sites were specific to one group, whereas others were shared. Thus, 7 sites were present in all groups, 2 were present in Pygmies and Europeans, 2 were present in Pygmies and Ashkenazim, and 19 were present in Ashkenazim and Europeans. The European group was the least polymorphic, with 44 segregating sites, as compared with 48 sites in Ashkenazim and 47 sites in Pygmies. In addition, the Italian reference had two changes absent from all the test samples. Average pairwise differences within populations indicated that Pygmies showed more diversity (23.6 compared with 16.8 for Ashkenazim and 15.0 for Europeans). Pairwise differences between groups were assessed by use of a permutation test (Hudson et al. 1992) and showed significant differences between Pygmies and the other two groups ($P < .001$ in each case) but less difference between Europeans and Ashkenazim ($P = .043$).

The haplotypes of the 23 test individuals depicted in figure 3 show clear linkage disequilibrium, at least within cosmids, leading to variant configurations conserved among each population group. Particularly evident are

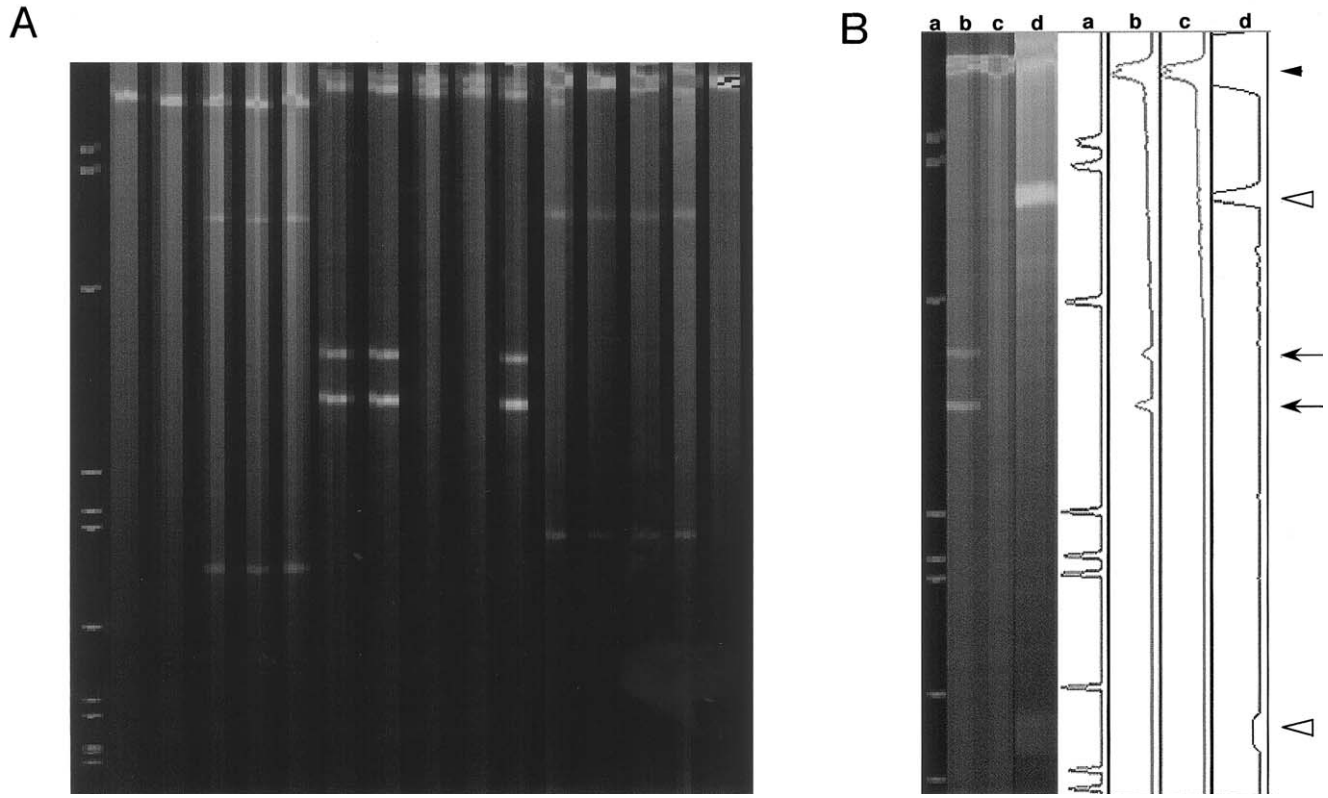

Figure 2 Detection of mismatches and of PCR stuttering at a CA microsatellite. *A,* Black-and-white reproduction of four-color gel. Five lanes of the multiplex are resolved into its components: red or size markers shown only once (track 1). Tracks 2–6, blue. Tracks 7–11, yellow. Tracks 12–16, green. Cleavage bands are present in tracks 4, 5, and 6; 7, 8, and 11; and 12, 13, 14, and 15. The cleaved mismatches arise from allelic differences, between reference and test DNA, at three different segregating sites. *B,* Black-and-white reproduction of four-color gel image tracks (*left*) and corresponding electropherograms (*right*). *Lane a,* red size markers used as internal standards but here are shown separately, for clarity. *Lane b,* single mismatch cleavage, dividing full-size green-labeled DNA (blackened arrowheads) into two smaller products (long arrows). *Lane c,* negative control, showing reference hemizygous green-labeled DNA, with no cleavage. *Lane d,* reference hemizygous blue-labeled DNA, cleaved at multiple sites within a CA_{18}, because of mismatches caused by PCR stuttering. Note broad diffuse cleavage bands (unblackened arrowheads).

these arrangements in the region covering cosmids 10 and 2. In this region, a single crossover in the interval between the cosmids, for example, may explain the multiple differences seen among several individuals (i.e., see, in fig. 3, rows 2 and 9 vs. rows 4, 6, 8, and 10, and row 3 vs. row 5 in Europeans; or rows 11, 13, 14, and 18 vs. rows 12, 15, and 19 in Ashkenazim). Furthermore, single changes most likely caused by more recent mutations are easily recognized (i.e, cosmid 2 in row 7).

Difference between DNA Variation Density Found in the X Chromosome and a Y-Chromosome–Specific Region

Eleven successful PCR reactions of ~1.5 kb were designed, in an 18.3-kb region of the Y chromosome known to be unique and that had been sequenced by Whitfield et al. (1995), and DNA variations among the Y chromosomes of the 24 men considered in the present study were sought. Only three segregating sites were found in the 0.4 Mb of DNA screened (fig. 4). The rarer allele was present at the first site in one European, at the second site in two Europeans, and at the third site in three Ashkenazim and in two Mbuti Pygmies.

Whereas the 24 men analyzed show 102 segregating sites in the 87 kb of Xq22 DNA screened, or one segregating site per 853 nt, the same men show a much lower density of segregating sites in the nonrecombining region of the Y chromosome that was examined. Only three segregating sites were found in a Y-chromosome region of 16.5 kb, or one segregating site per 5,500 bp. This represents a 6.4-fold excess of segregating sites on the X chromosome, relative to the Y chromosome. Since there are three X chromosomes for every Y chromosome in the population, the deficit of variation in the human

Table 1

Correspondence between Estimated Mismatch Position and True Position of Sequence Change

Name[a]	Mismatch Position[b]	Nucleotide Position[c]	Nucleotide Change[d]
2.1a	573	575	G→A
2.1b	292	294	ATTA insertion
2.1c	332	334	G→C
2.1d	1450[e] ± 50	1454	C→G
2.1e	180	180	A→C
2.2a	405	406	G→T
2.2b	561	563	G→T
2.2c	495	494	A→G
2.2d	253	254	C→T
2.3a	382	382	T→C
2.3b	661	665	G→C
2.3c	375	372	C→T
2.5a	361	361	C→T
2.5b	448	451	A→G
2.5c	588	587	G→C
2.6a	243	244	T→C
2.6b	528	528	G→A
2.6c	213	213	C→A
2.6d	407	408	G→A
2.6e	531	531	C→T
2.6f	205	204	G→A

[a] The first digit refers to the cosmid, the second digit to the individual 1.5-kb segment, and the letter to the individual segregating site.
[b] Measurement of base pairs from either end of segment.
[c] Measurement of base pairs from start of forward strand or reverse strand (underlined), relative to orientation of sequence deposited in EMBL database.
[d] Changes in reverse strand are underlined.
[e] The smaller cleavage product was not visible in this case.

Y chromosome, relative to the X chromosome, found in our experiments is 2.15-fold. This, however, does not take into account any difference in mutation rate between the X and Y chromosomes. Since the Y chromosome is confined exclusively to males, whereas the X chromosome spends only one-third of its time in this sex, a difference in sex-specific mutation rates would result inevitably in different mutation rates between the X and Y chromosomes. If the higher male mutation rate (see below) is taken into account, the deficit of variation we found in the Y chromosome, relative to the X chromosome, is fourfold ($2.15 \times 1.33/0.695 = 4.11$).

The Divergence of Human and Chimpanzee DNA Sequences in the X and Y Chromosomes

To measure the divergence between human and chimpanzee DNA, we used the human PCR primers to amplify the DNA from a male chimpanzee. Six of the 11 Y-chromosome reactions, and 22 of the 36 for the X-chromosome cosmids 2–8, were successful. The primers of these reactions were used to sequence segments from each end of the amplified regions. We have thus obtained short stretches of sequence, haphazardly distributed within the examined Xq22 and Y-chromosome regions. The X sequences totalled 13,960 bp, or one-sixth of the region screened by our mismatch-detection procedure. The Italian reference sequence differed from that of the male chimpanzee at 97 positions, because of the following mutations: 61 transitions (including 10 at the chimpanzee's CpG sites), 28 transversions, 4 double substitutions, 1 3-bp deletion, and, finally, 1 2-bp and 2 4-bp insertions. This gives a total average of 0.695% mutational changes between the X-chromosome sequence of the human and chimpanzee lineages. The Y-chromosome sequences of our reference human and chimpanzee DNA totaled 4,650 bp and showed 62 differences. These were due to 40 transitions, including 9 at the chimpanzee CpG sites; 15 transversions; 1 2-bp substitution; 2 1-bp, 1 2-bp, 1 3-bp, and 1 4-bp insertion; and 1 18-bp deletion. The level of evolutionary divergence observed by us in the Y-chromosome sequences is 1.33%.

Sex- and Type-Specific Mutation Rates

The ratio of male-to-female mutation rates in humans and chimpanzees during the 5 million years of their separate evolution (White et al. 1994) can be estimated from the sequence divergence between the two species, observed for the DNA of the X and Y chromosomes. Thus, the 1.33% divergence between the chimpanzee and human Y chromosome can be expressed as the male mutation rate $v = 1.33K$ (in which K indicates a rate of mutation per base per year), and the 0.695 divergence of X-chromosome sequences can be expressed as

$$\frac{1v}{3} + \frac{2\mu}{3} = 0.695K,$$

in which μ is the female mutation rate. It follows that

$$\frac{v}{\mu} = \frac{1.33}{3 \times 0.695 - 1.33} \times 2 = 3.52$$

Consequently, the autosomes, which spend equal time in the two sexes, should have mutation rates 1.227-fold greater than those of the X chromosome. This factor is obtained by considering the 3.5 value of the ratio of male-to-female mutation rates; the average mutation rate of autosomes to X chromosomes should then be: $[3\,(3.5 + 1)]/[2\,(3.5 + 2)]$

Our analysis of divergence between X- and Y-chromosome sequences of humans and chimpanzees also suggests the following mutation rates per nucleotide site per year, in the Y and X chromosome, respectively: 11.5 and 5.15×10^{-9} transitions at CpG sites; 7 and 3.92×10^{-10} transitions at non-CpG sites; 3.44 and 2.36 ×

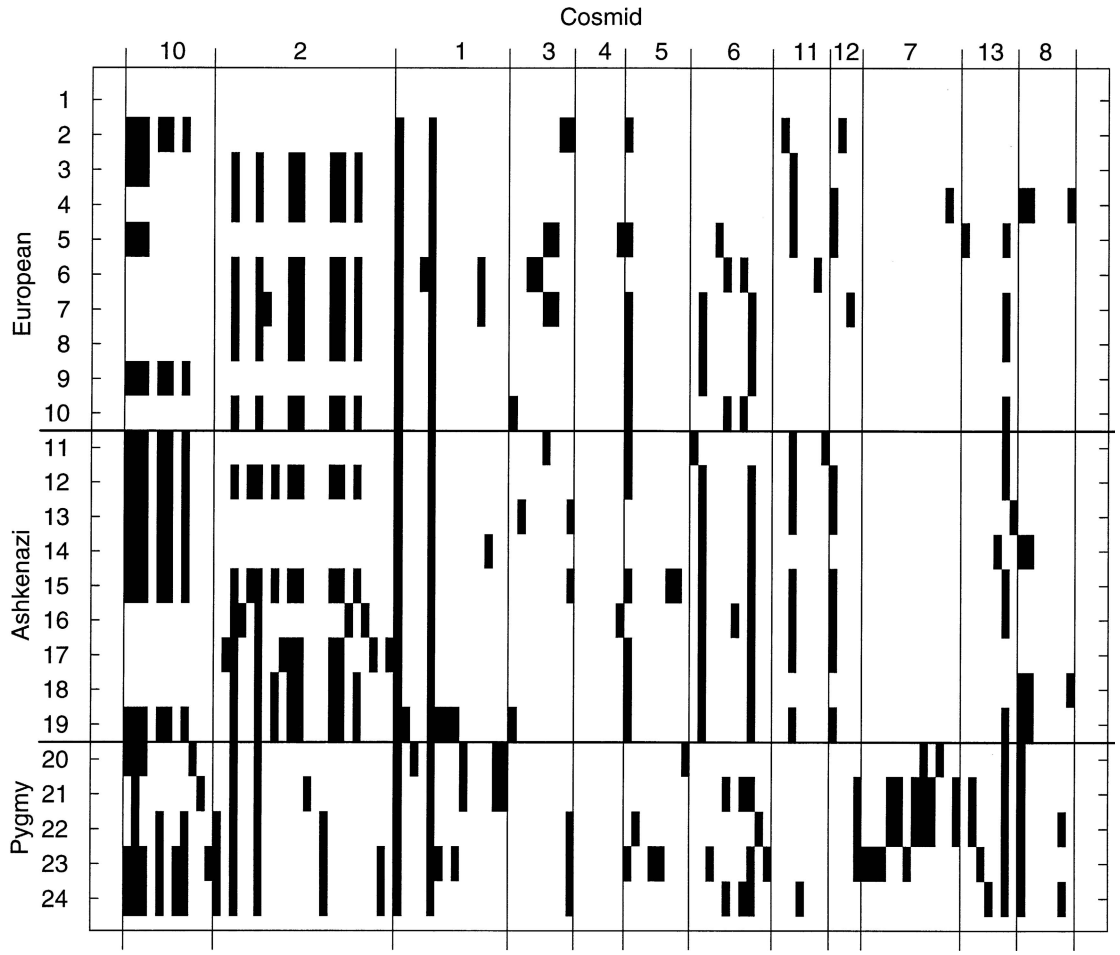

Figure 3 X-chromosome polymorphism matrix. Segregating sites are shown, indicating when individuals differ from (blackened areas) or are identical to (unblackened areas) the reference man in row 1. Columns represent 1.5-kb segments of cosmid inserts. Segments with no segregating site are indicated by a single column, whereas those with segregating sites are represented by a number of columns equal to the number of segregating sites. The numbers of 1.5-kb segments examined in cosmids 1–8 and 10–13 are 5, 6, 4, 6, 4, 5, 5, 6, 4, 5, 3, 4. Vertical lines are used to separate cosmids.

10^{-10} transversions; and 1.29 and 0.29 × 10^{-10} small deletions and/or insertions (fig. 5).

Discussion

The variation screening strategy we have developed has shown several desirable properties. The rate of screening is 81 kb DNA per ABI gel run, using an apparatus with 36 wells, and can be increased proportionally by using gels containing a higher number of lanes (e.g., 64 or 96). The chemistry at the root of the method has been tested extensively by use over the past 10 years in our laboratory and many others around the world (Cotton et al. 1988; Grompe et al. 1989; Montandon et al. 1989; Akli et al. 1991; Rowley et al. 1995; Verpy et al. 1996). The only mismatch that is not detected when labeling is limited to the probe DNA is the T.G mismatch, in which the T is in the probe strand and lies 3′ to a G (Forrest et al. 1991). Similarly, only mismatches in the 50 bp at each end of the 1.5-kb segments used in our experiments are expected to produce cleavage bands that are either too weakly labeled or are too similar in size to the broad full-size band to be detected reliably. The mismatch detection reactions are performed in ≤5 h and can be batched in 96-well microtitre plates. The chemistry is very reliable, since it is not affected by batch variation, and the stocks of chemicals are stable for ≤3 mo. Furthermore, the chemicals are inexpensive. They are toxic, but the amounts used in our procedure are very small, and appropriate handling in a fume cup-

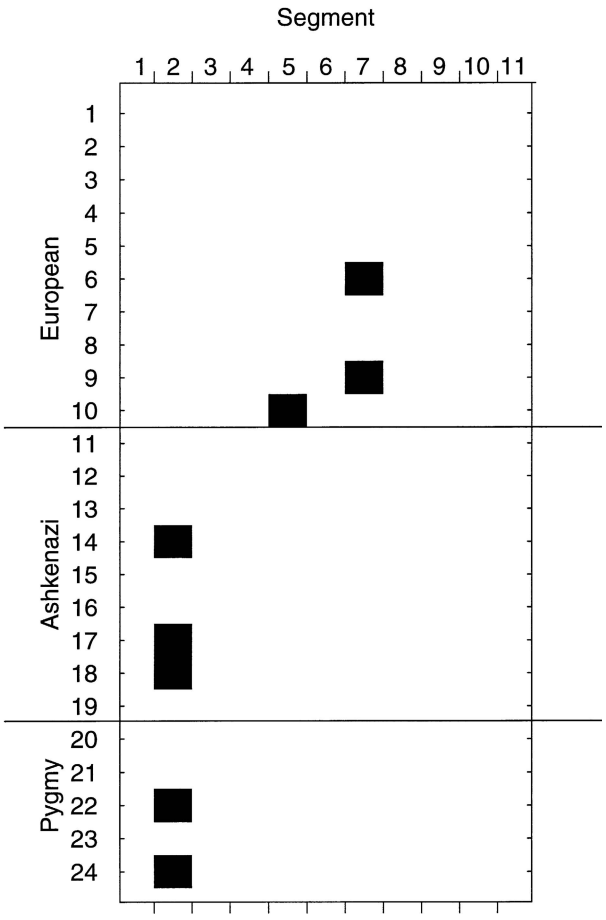

Figure 4 Y-chromosome polymorphism matrix. Segregating sites are shown, indicating when individuals differ from (blackened areas) or are identical to (unblackened areas) the reference man in row 1. Columns represent 1.5-kb segments, with or without segregating sites. No segment has more than one segregating site.

board ensures safety. The reproducibility of the procedure has been proved by past experience and in this work, because the 324 mismatch reactions, which were repeated twice, did not show even a single discrepancy. The location of the mismatch bands is so precise that, in most cases, variant alleles are identified without recourse to sequencing. Furthermore, mismatch detection also can be used to verify the identity of variant alleles, thus providing a rapid method for confirming allele sharing between individuals and populations.

Our approach revealed interesting patterns of variation over a 5-Mb region of the X chromosome in the three populations analyzed. The greater level of diversity observed among the Pygmies is consistent with the idea that modern humans originated in Africa (Brauer 1989), and the greater similarity between Europeans and Ashkenazim is not unexpected. In the latter two populations, polymorphisms are most abundant in cosmid 2 and are rare in cosmids 4 and 7, whereas in Pygmies, the last cosmid was the most polymorphic. This suggests that variation in the density of polymorphic loci in the different cosmids reflects the genetic history of different populations, rather than different patterns of mutations along the Xq22 region.

We have observed a fourfold deficit of DNA variation in the Y chromosome, relative to the X chromosome of the men analyzed. Dorit et al. (1995) and Whitfield et al. (1995) have commented on the rarity of polymorphisms on the human Y chromosome; more recently, Underhill et al. (1997) have reported greater success in the detection of Y-chromosome diallelic polymorphisms, using their own screening procedure, which is based on denaturing high-performance liquid chromatography.

The deficit of variation in the Y chromosome that we have observed could result from a number of different events—namely, selective sweeps, male migration, and polygyny. Analysis of further unique, well-mapped sequences of the human Y chromosome would be useful to confirm the above sequence-variation deficit.

To examine the sequence divergence between humans and chimpanzees, we have chosen to examine multiple short sequences at randomly selected locations. This should ensure that they are representative of the chromosome and are not biased unduly by the inclusion of sequences with special evolutionary constraints. The 1.33% divergence we found for the Y chromosome is very similar to that (1.3%) observed by Whitfield et al. (1995), after exclusion of known coding sequences in the whole region of the Y chromosome we have screened. Furthermore, the frequency of changes in the segments of sequence analyzed was random, with no evidence of variation in mutation rate by segment. The difference in divergence observed in the X, relative to the Y, DNA seems to indicate clearly a higher male mutation rate during the 5 million years of separate evolution of the two species. This is consistent with the hypothesis of male-driven evolution and other data on mammalian species (Miyata et al. 1990; Shimmin et al. 1993; Drost and Lee 1995; Crow 1997). However, the 3.5 value of the ratio of male-to-female mutation rates, calculated for humans and chimpanzees during this period, is smaller than the 8.6 ratio value we estimated for humans, in a recent study of the U.K. hemophilia B population (P. Green, T. Anagnostopoulos, S. Saad, C. Lewis, F. Gianelli, unpublished data). This may be because the 3.5 ratio value relates to mutations in both humans and chimpanzees, and also because contemporary men are likely to reproduce at older ages than their early ancestors or chimpanzees. An age effect on male mutation rates is expected and is at least in part explained by the age-related increase in the number of cell divisions in the male germ line (Drost and Lee 1995; Crow 1997; Brinkmann et al. 1998).

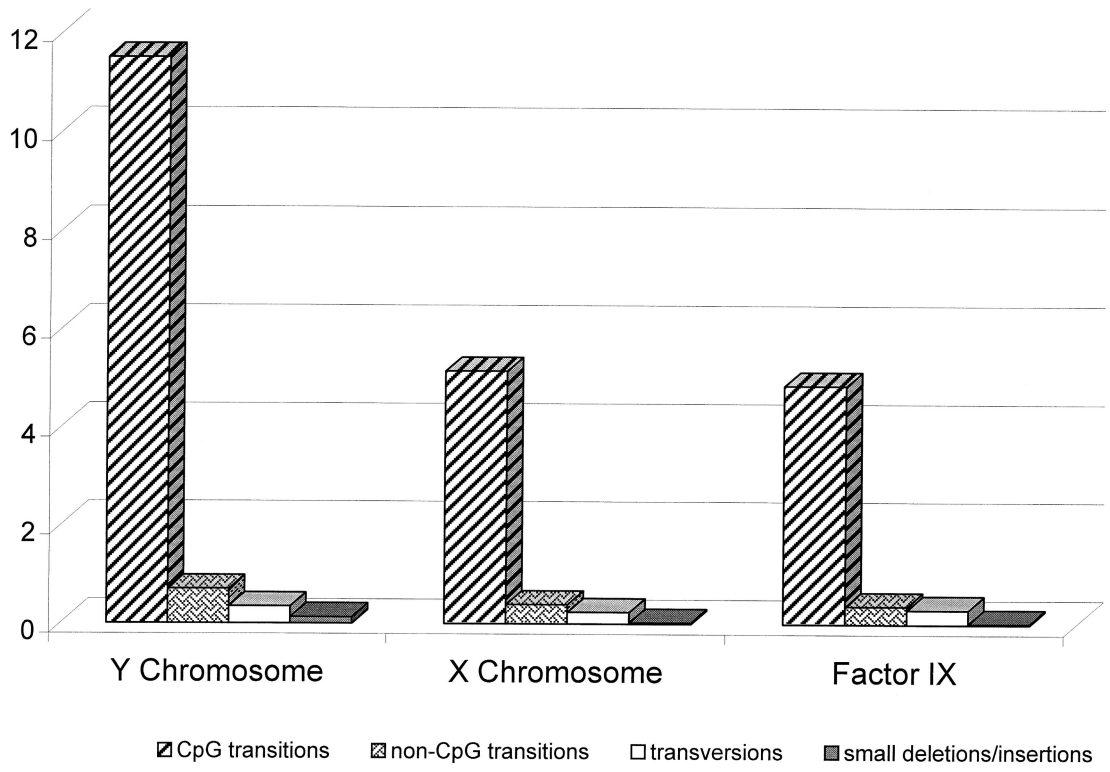

Figure 5 Mutation rates per base per year, relative to different types of changes in the Y and X chromosome of humans and chimpanzees, and the human factor IX gene. The figures on the vertical scale should be multiplied by 10^{-9}.

The type-specific mutation rates for the X chromosome, estimated from the human/chimpanzee sequence divergence, agree with the point mutation rates we estimated for the factor IX gene (fig. 5), using data from our recent study of the U.K. hemophilia B population (P. Green, T. Anagnostopoulos, S. Saad, C. Lewis, F. Gianelli, unpublished data). If we assume a generation interval of 20 years, the latter mutation rates per base per year are 4.86×10^{-9} for transitions at CpG sites, 3.69×10^{-10} for transitions at non-CpG sites, 2.97×10^{-10} for transversions, and 2.39×10^{-11} for small deletions and insertions.

We have demonstrated a new approach to exploring DNA variation in the human genome. This approach can screen DNA at a rate of 0.25 Mb per person per week, and was used in a 5-Mb region of Xq22 to reveal the haplotypes of men from different populations. The structure found in these data indicates that, with new analytical methods capable of dealing with recombining DNA, the type of data we have obtained could advance greatly the investigation of human genetic history and, in specific cases, could help to distinguish variations important for disease susceptibility from trivial polymorphisms. In addition, we present evidence of a high ratio of male-to-female mutation rates during the 5 million years of divergent evolution of humans and chimpanzees and a fourfold deficit of sequence variation in the Y chromosome of contemporary men. Furthermore, the rates of different types of mutations for the X chromosome, derived from the sequence divergence between humans and chimpanzees, appear to agree with figures for the X-linked factor IX gene in contemporary humans. This suggests that the estimates presented truly reflect the degree of DNA stability in the human germ line.

Acknowledgments

This work was supported by the Medical Research Council. Dr. A. Fensom kindly provided DNA from Ashkenazim, and Dr. E. Kendall provided mapping information on Xq22. We thank Prof. D. J. Balding, Dr. J. C. Whittaker, and Dr. M. E. Weale, for helpful discussions; Brenda Buckle, for tissue culture; and Adrienne Knight, for secretarial help.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

European Molecular Biology Laboratory, http://www.embl-heidelberg.de/ (for sequence information [Z68331,

Z68871, Z69367, Z69722, Z70224, Z70226, Z70227, Z70274, Z73900, Z75745])

Genbank, http://www.ncbi.nlm.nih.gov (for sequence information [X96421, L35265])

References

Akli S, Chelly J, Lacorte J-M, Poenaru L, Kahn A (1991) Seven novel Tay-Sachs mutations detected by chemical mismatch cleavage of PCR-amplified cDNA fragments. Genomics 11: 124–134

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Bentley DR (1996) Genomic sequence information should be released immediately and freely in the public domain. Science 274:533–534

Brauer G (1989) The evolution of modern humans: a comparison of the African and non-African evidence. In: Mellars P, Stringer C (eds) The human revolution: behavioural and biological perspectives on the origin of modern humans. Edinburgh University Press, Edinburgh, pp 123–154

Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, New Jersey

Collins FS, Guyer MS, Charkavarti A (1997) Variations on a theme: cataloguing human DNA sequence variation. Science 278:1580–1581

Cordell HJ, Todd JA (1995) Multifactorial inheritance in type 1 diabetes. Trends Genet 11:499–504

Cotton RG, Rodrigues NR, Campbell RD (1988) Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc Natl Acad Sci USA 85: 4397–4401

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

Dorit Rl, Akashi H, Gilbert W (1995) Absence of polymorphism at the YZF locus on the human Y chromosome. Science 268:1183–1185

Drost JB, Lee WR (1995) Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among Drosophila, mouse, and human. Env Mol Mutagen 25 (supp 26):48–64

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22:521–565

Forrest SM, Dahl HH, Howells DW, Dianzani I, Cotton RGH (1991) Mutation detection in phenylketonuria by using chemical cleavage of mismatch: importance of using probes from both normal and patient samples. Am J Hum Genet 49:175–183

Griffiths RC, Marjoram P (1996) Ancestral inference from samples of DNA sequences with recombination. J Comput Biol 3:479–502

Grompe M, Muzny DM, Caskey CT (1989) Scanning detection of mutations in human ornithine transcarbamylase by chemical mismatch cleavage. Proc Natl Acad Sci USA 86: 5888–5892

Hammer MF (1995) A recent common ancestry for human Y chromosomes. Nature 378:376–378

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, et al (1997) Archaic African *and* Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:772–789

Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. Genetics 145:833–846

Hudson RR (1990) Gene genealogies and the coalescent process. Oxf Surv Evol Biol 7:1–44

Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. Mol Biol Evol 9:138–151

Jobling MA, Tyler-Smith C (1995) Father and sons: the Y chromosome and human evolution. Trends Genet 11: 449–456

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2048

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

Miyata T, Kuma K, Iwabe N, Hayashida H, Yasunaga T (1990) Different rates of evolution in autosome, X chromosome and Y chromosome-linked genes: hypothesis of male-driven molecular evolution. In: Takahata N, Crow J (eds) Population biology of genes and molecules. Baifukan, Tokyo, pp 341–357

Montandon AJ, Green PM, Giannelli F, Bentley DR (1989) Direct detection of point mutation by mismatch analysis: application to haemophilia B. Nucleic Acids Res 17: 3347–3358

Olson MV (1995) A time to sequence. Science 270:394–396

Rowen L, Mahairas G, Hood L (1997) Sequencing the human genome. Science 278:605–607

Rowley G, Saad S, Giannelli F, Green PM (1995) Ultrarapid mutation detection by multiplex, solid-phase chemical cleavage. Genomics 30: 574–582

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

Stoneking M, Soodyall H (1996) Human evolution and the mitochondrial genome. Curr Opin Genet Dev 6:731–736

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, et al (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005

Verpy E, Biasotto M, Brai M, Misiano G, Meo T, Tosi M (1996) Exhaustive mutation scanning by fluorescence-assisted mismatch analysis discloses new genotype-phenotype correlations in angioderma. Am J Hum Genet 59:308–319

Verpy E, Biasotto M, Meo T, Tosi M (1994) Efficient detection of point mutations on color-coded strands of target DNA. Proc Natl Acad Sci USA 91:1873–1877

White TD, Suwa G, Asfaw B (1994) *Australopithecus* [*Ardipithecus*] *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

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, et al (1997) Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183