

Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum

Esther P. Leeflang¹, Lin Zhang¹, Simon Tavaré^{1,2}, Rene Hubert^{1,+}, Jayalakshmi Srinidhi³, Marcy E. MacDonald³, Richard H. Myers⁴, Margot de Young⁵, Nancy S. Wexler⁶, James F. Gusella³ and Norman Arnheim^{1,*}

¹Molecular Biology Program and ²Department of Mathematics, University of Southern California, Los Angeles, CA, 90089, ³The Molecular Neurogenetics Unit, Massachusetts General Hospital, Boston, MA, 02129, ⁴Department of Neurology, Boston University Medical School, Boston MA, 02118, USA, ⁵Association of Friends of Families with Huntington's Disease, Maracaibo, Zulia, Venezuela and ⁶Department of Neurology, Columbia University, New York, NY, 10032, and Hereditary Disease Foundation 1427 7th St. Suite 2, Santa Monica, CA 90401, USA

Received April 10, 1995; Revised and Accepted June 2, 1995

The CAG triplet repeat region of the Huntington's disease gene was amplified in 923 single sperm from three affected and two normal individuals. Average-size alleles (15–18 repeats) showed only three contraction mutations among 475 sperm (0.6%). A 30 repeat normal allele showed an 11% mutation frequency. The mutation frequency of a 36 repeat intermediate allele was 53% with 8% of all gametes having expansions which brought the allele size into the HD disease range (≥ 38 repeats). Disease alleles (38–51 repeats) showed a very high mutation frequency (92–99%). As repeat number increased there was a marked elevation in the frequency of expansions, in the mean number of repeats added per expansion and the size of the largest observed expansion. Contraction frequencies also appeared to increase with allele size but decreased as repeat number exceeded 36. Our sperm typing data are of a discrete nature rather than consisting of smears of PCR product from pooled sperm. This allowed the observed mutation frequency spectra to be compared to the distribution calculated using discrete stochastic models based on current molecular ideas of the expansion process. An excellent fit was found when the model specified that a random number of repeats are added during the progression of the polymerase through the repeated region.

INTRODUCTION

Huntington's disease (HD) is a hereditary neurodegenerative disease, usually with an onset in middle age, and associated with progressive disordered movements, decline in cognitive function and emotional disturbance. It is one of nine triplet repeat diseases (1) which includes fragile X (FRAXA, FRAXE), myotonic dystrophy (DM), spino-bulbar muscular atrophy (SBMA), spinocerebellar ataxia type I (SCA1), dentatorubral-pallidoluysian atrophy (DRPLA), Haw River Syndrome (HRS) and most recently, Machado–Joseph disease (2). In each case, an increase in number of triplets is associated with the disease phenotype, but significant differences exist among the diseases in the number and range of triplet repeats that cause disease, somatic and gonadal mosaicism, and time in development at which mutations occur.

Mosaicism in an individual's sperm DNA relative to lymphocyte DNA has been established in HD (3–5). Heretofore, studies on sperm have only been on purified DNA samples from semen, revealing a radioactive smear of PCR products on acrylamide gels and denoting a heterogeneous triplet repeat length. The extent of triplet expansion could only be evaluated on a gross level, with maximum, minimum and mode values of allele size deduced from densitometric analysis. A problem with this approach is that faint signals derived from a small subset of the sperm population may be obscured by the inherent PCR stutter from more abundant allele sizes. Perhaps more serious is the possibility that data from amplification of semen DNA will be strongly biased towards smaller alleles since they will have a competitive advantage during PCR.

*To whom correspondence should be addressed

⁺Present address: Medical Genetics Program, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA

Single sperm typing has been used to measure germline mutation frequencies at the SBMA and DM loci (6,7). We performed single sperm PCR analysis to examine the allele size variation in sperm cells within and among HD affected and unaffected individuals. We made an exact count of the number of mutant alleles and their sizes, and without the bias mentioned above. We compared stochastic models of the mutation process to the data, an approach which depends on the availability of discrete data on allele size rather than the continuously distributed smears obtained from PCR of semen DNA. Finally, using sperm from a donor carrying an intermediate allele (IA), we were able to estimate the risk of having affected offspring.

Table 1. Single sperm typing data on three HD affected (HD2, HD6 and HD15) and two unaffected individuals (HD 232SP and B)

(CAG) _n	Donor	# Sperm analysed	Mean change	% expanded	% contracted
51	HD2	95	21.71	97	2
49	HD6	94	10.44	95	3
38	HD15	98	4.05	88	4
36	HD 232 SP	81	0.49	42	11
30	B	80	0.04	9	3
18	HD2	123	0.00	0	0
18	HD6	90	-0.01	0	1
18	B	77	0.00	0	0
17	HD 232 SP	82	-0.04	0	2
15	HD15	103	0.00	0	0

The data are listed in rows from largest to smallest number of (CAG)_n triplet repeats as measured in individuals' lymphocytes.

RESULTS

Mutation frequency of HD alleles

We determined the mutation frequency in HD individuals by comparing the size of individual sperm PCR product to the somatic allele size from the same individual obtained by amplifying lymphoblast DNA. Three investigations (3,4,8) concluded that little if any somatic variation exists among different tissues of HD patients. A single report has proposed that different regions of the brain can vary in repeat number but in that study (5) no differences were found among 14 other tissues which included whole testis and spleen.

Table 1 shows the results of typing 923 sperm, from three individuals with HD and two normal individuals representing nine different allele sizes, using a sperm typing protocol modified for analysis of trinucleotide repeat disease alleles (7). Among the diseased individuals, 316 normal alleles were detected, while 287 HD alleles were observed, which is not statistically different ($P > 0.20$) from the expected 1:1 segregation. Assuming no meiotic segregation distortion, our PCR conditions are capable of amplifying large and small alleles with almost equal efficiency. The lack of any meiotic segregation distortion was confirmed when data on 522 sperm coamplified for the tightly linked dinucleotide repeat marker D4S127 revealed 1:1 segregation (data not shown).

Among the 287 sperm carrying HD chromosomes, 96% differed in size from the donors' somatic HD allele. Of these, 267 were expansions and nine were contractions. Representative allele-sizing data are shown in Figure 1. Two thirds of all the samples were studied to distinguish between alterations in the number of CAG and CCG repeats since the latter are polymorphic in the human population (9,10) and some mutations might be the result of CCG repeat expansions or

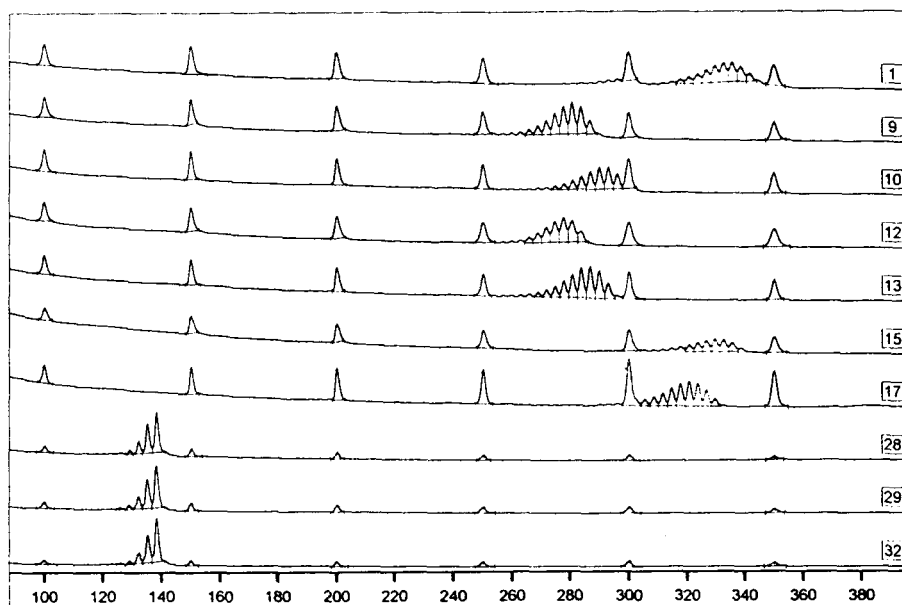


Figure 1. Representative measurements of triplet repeat lengths from individual HD2 using fluorescein-labeled single sperm PCR products analyzed on the A.L.F.TM Fragment Manager (Pharmacia). The top seven traces show HD alleles, reflecting a heterogeneous triplet repeat length. The bottom three traces show normal alleles which reflect stability in triplet repeat number. Each lane shows an internal size marker ladder from 100–350 bp, in increments of 50 bp. Peak find settings were as follows: peak width 5, peak height 1.0, cluster sensitivity 10.

contractions. However, no example of a CCG mutation was detected (data not shown).

The mutation frequency of normal alleles

An intermediate allele (36 CAG repeats) from an unaffected individual in his fifth decade showed 42% expansions and 11% contractions for an overall mutation frequency of 53%. Sixteen percent were mutations with ≥ 38 repeats. Mutations of an allele with 30 repeats were less frequent (9% expansions and 2.5% contractions). No mutations into the HD range were detected among these 80 sperm. We observed seven expansions (+1 repeat) and two contractions (-1 and -3 repeats). Representative data for two expansion events for this donor (B) are shown in Figure 2. Among 475 sperm from 5 alleles in the 15-18 repeat range, only three small contractions (1 or 2 repeats) were observed.

The size distribution of mutations

The CAG repeat number distribution in single sperm for the diseased and one intermediate allele donors are shown in Figure 3. There is a significant change in the size distribution of the expansion mutations with increasing allele size. It can be seen to progress from an apparently normal distribution around the somatic DNA size in the case of the 36 repeat allele (Fig. 3d), to a markedly more uniform distribution of expanded alleles up to twice the size of the somatic DNA in the case of the 49 and 51 repeat CAG tracts (Fig. 3a and 3b). Contraction sizes were generally limited to six repeats or less.

It is well known that PCR stutter occurs during the amplification of microsatellite repeats. One potential problem is that *Taq* polymerase errors during the many amplification cycles required for single sperm analysis could produce PCR products with an allele size different from that of the input sperm DNA molecule. It is easy to imagine how contraction mutations might result from *Taq* artifacts since smaller PCR products would have a selective amplification advantage. Control experiments involving amplification of normal alleles at the SBMA

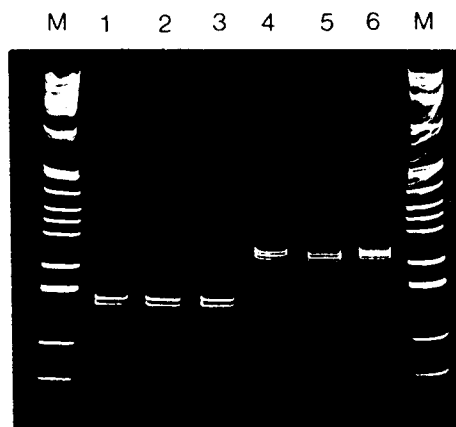


Figure 2. Single sperm PCR products of the triplet repeat region from individual 'B'. Lanes 1-3 show the (CAG)₁₈ allele, lanes 4-6 show the (CAG)₃₀ repeat allele. Lanes 4 and 6 each contain one 3 bp expansion mutation which occurred in the CAG repeat tract (CCG data not shown). M denotes a pBR322 *Msp*I digest marker lane. The size of the somatic alleles was determined by analysis of lymphocyte DNA.

locus in single sperm showed however that the observed contraction mutations were not artifacts of PCR (6). However, the probability of an artifact would be expected to depend upon the number of repeats in the input DNA template. The vast majority of the mutations we observed at the HD locus were large expansions. Although it is difficult to imagine selective pressures that would allow molecules expanded *in vitro* to take over the population of PCR product, we nevertheless performed control PCR experiments to examine this question.

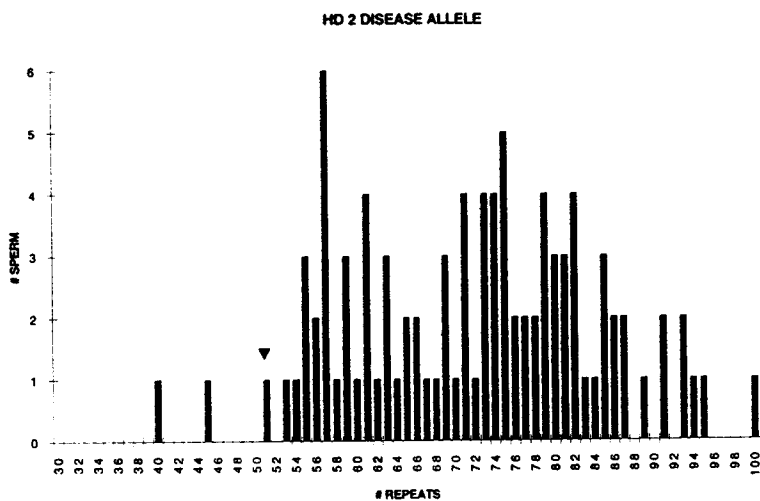
Analysis of HD sperm from donor HD2 (51 repeats) showed a mean change of +21 repeats. Single molecule dilution experiments were carried out on somatic DNA from this donor. Among the 30 disease-length molecules examined, we detected seven events with a change of -1 repeats and three events with a change of +1 repeats for a mean change of -0.13 repeats. We do not know whether the variation in allele size is due to PCR artifact, whether there is some variation in allele size among individual lymphoblast cells or whether the variation seen is due to the inherent error in the exact measurement of molecular weight using the A.L.F.TM. Regardless, the variation in repeat number seen in the somatic DNA of HD2 is insignificant relative to the large variation in repeat number observed among the sperm cells of this donor. Thus, the occurrence of PCR artifacts would have little if any effect on the estimates of mutation frequency and size distribution. Single molecule dilution studies on somatic DNA from donor HD 232 SP (36 repeats) showed that PCR artifacts would also have little if any impact on our estimate of the mutation frequency spectrum of the smaller alleles. Among 35 molecules tested we detected one event with +1 repeat change and one event with -1 repeat change. Again, the exact source of this variation is not known.

DISCUSSION

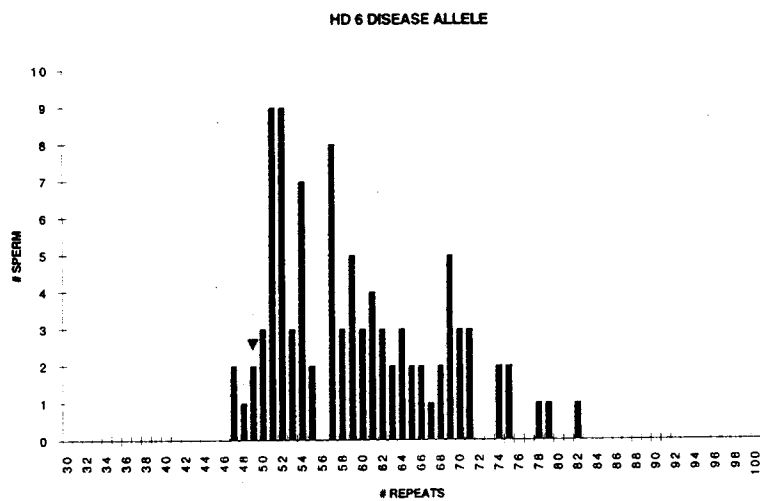
Comparison of sperm typing data to family transmissions

The single sperm typing data on disease alleles can be compared with HD paternal transmissions studied in families. In five reports (3,8,11-13) exact repeat number changes could be determined in 141 paternal transmissions. Of these, 72% showed a contraction or expansion mutation. The single sperm data for the three HD alleles we studied gave a 96% average mutation frequency. Expansions account for 97% of the sperm mutations and 91% (range: 83-100%) of the mutations detected in paternal transmissions (3,8,11,12). An average of 5.6 repeats was added for each expansion observed in the pedigrees (3,8,11-13) (range: 3.0 to 9.0) while 12.1 repeats was the average size of an expansion detected in sperm from HD2, HD6 and HD15. The differences between the sperm and family data could result from the small number of HD sperm donors studied and that two of them had an HD allele size significantly larger than the average paternal HD allele size in the families. There may also be differences among the populations studied (3,8,11-13). We note that the sperm data, overall, corresponds best to the family data (3) (83% mutation frequency; 83% of mutations were expansions and an average of nine repeats were added per expansion) which includes a large number of Venezuelan pedigrees, the source of our sperm donors. Some of the differences between the sperm and family data could also reflect selection against sperm with very large alleles in

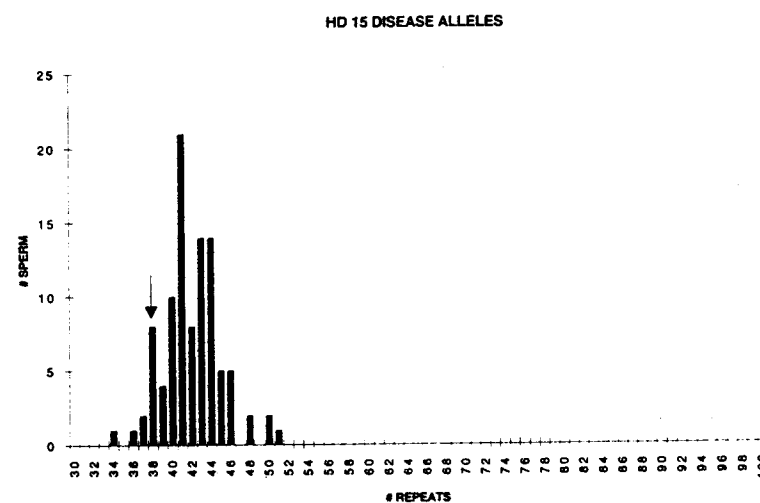
a



b



c



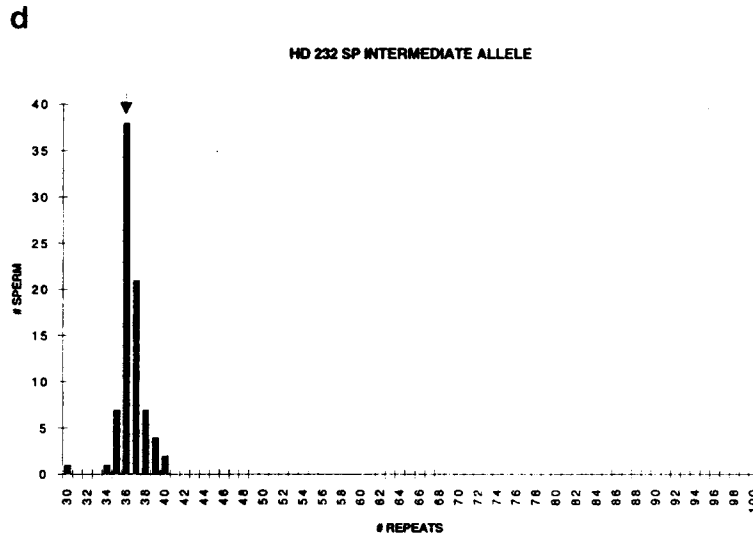


Figure 3. Distribution of single sperm length mutations for three diseased individuals [(a-c), HD 2, 6 and 15 respectively] and one carrying an intermediate allele [(d), HD 232 SP]. Arrows denote the length of the donors HD (or intermediate) allele in somatic cells determined by analysis of lymphoblast DNA.

Table 2. Expansion size variance and parameter estimates using the model for trinucleotide repeat expansions

(CAG) _n	Mean size of expansion	Variance of size of expansions (observed)	Variance of size of expansion (fitted)	T (± s.e.)	α (± s.e.)	p (± s.e.)
51	21.33	134.59	157.1	6.71 (1.08)	0.131 (0.021)	0.24 (0.03)
49	10.84	64.92	64.1	4.41 (0.69)	0.090 (0.014)	0.29 (0.03)
38	4.31	7.44	7.8	10.72 (3.72)	0.282 (0.098)	0.71 (0.07)
36	0.76	1.06	1.1	3.90 (2.51)	0.108 (0.070)	0.84 (0.09)

terms of survival or ability to carry out fertilization, or post-zygotic selection.

Alleles in the 33–37 repeat size range have been found in parents of HD offspring in families with no previous history of the disease. These so called 'intermediate alleles' are thought to be a source of new HD mutations (8,13,14–16). Among the 163 sperm analyzed from the individual carrying a 36 and 17 repeat allele, 8% of all the gametes had expanded into the HD size range. Additional studies on other intermediate alleles are necessary. It is important clinically to assess the risk of having a child carrying an HD allele in other males found to carry intermediate alleles. The degree to which the risk may vary among chromosomes carrying the same allele size, but with different haplotypes (17) in the Huntington gene region, must be specifically addressed. Such studies could lead to information on potential *cis*-acting elements that affect mutation frequency (18,19).

Comparison of the sperm typing data on alleles in the 15–18 repeat range to the results of family studies are difficult, given the low mutation frequency of these events. Our estimate of the mutation frequency is 0.6%, whereas a recent estimate based on pedigree analysis was 0.2% (12). The reliability of these estimates is uncertain considering the small number of mutations observed. In addition, sperm typing measures the mutation frequency of specifically chosen alleles whereas

in family studies, mutation frequency estimates represent transmissions encompassing the full distribution of alleles in the population.

Comparison of mutation characteristics at the HD and SBMA loci

Single sperm data from one SBMA individual with 47 repeats shows an 81% mutation frequency with 66% expansions and 15% contractions (7). Most of the mutation events were +1 or 2 repeats. The largest expansion seen in this SBMA individual was +11. The sperm data on the HD allele closest in size (49 repeats) shows a 95% expansion and a 3% contraction frequency with an average 10.8 repeats added per expansion. The largest expansion was +33 repeats. Even the 38 repeat HD allele has a higher mutation frequency, a greater average change in repeat number per mutation and a broader distribution in mutant allele size than the 47 repeat SBMA allele. Alleles at the SBMA locus in the high normal range (28–31 repeats) show only a 0.4% expansion frequency (7) compared to the 30 repeat allele we describe in this paper which has a 9% expansion frequency.

These differences show that factors in addition to overall repeat size must be important contributors to the measured triplet expansion frequency in sperm. Interrupted repeat tracts, as seen in SCA1 (20), DM (21) and FraxA (22,23) cannot

contribute to this difference since no interruptions of the CAG stretches have been reported for SBMA or HD. It remains to be determined whether differences exist between the loci in terms of selection against large alleles during gametogenesis or the molecular mechanism itself (3), including the possibility of *cis*-acting elements (18).

Studies on normal alleles at the SBMA locus showed that an increase in contraction frequency accompanies an increase in expansion mutations. As allele size increases from 21 (normal) to 47 (SBMA) repeats, the contraction frequency goes up seven fold while the expansion frequency increases 200 fold (6,7). At the huntingtin locus the contraction frequency of the 36 repeat allele appears higher than that of smaller alleles. However, HD alleles with expansion frequencies approaching 100% have reduced contraction frequencies relative to the 36 repeat allele. Additional data on contraction frequencies as a function of repeat size are needed.

A model for the CAG repeat mutation process

It is clear from Figures 3A–3D that the mutation frequency spectrum of individuals with different somatic allele sizes vary significantly. Because we have been able to generate discrete (count) data on the distribution of allele sizes in sperm, we can compare our data to mutation frequency spectra generated from discrete stochastic models of the expansion process based on current molecular ideas on triplet repeat expansion.

A number of models may account for trinucleotide repeat instability. Both contraction and expansion events can result from unequal reciprocal recombination between homologs or sister chromatids. In either case, equal numbers of contractions and expansions should be observed, an expectation that is not fulfilled by the HD or SBMA data, although non-reciprocal recombination mechanisms (24) cannot be excluded. Instability due to replication slippage is a more likely possibility (22,23,25). Recent experimental evidence on dinucleotide repeat mutations in yeast and studies on colon cancer also support this type of mechanism (26–29).

The proximal cause of a slippage event, the number of repeats added or deleted per slippage event and the number of slippage events that take place as the polymerase traverses the repeated region is unknown. Repeated DNA may be capable of secondary structure formation (30,31) (Petruska, Arnheim and Goodman, unpublished data). These structures could lead to a block and subsequent stalling of polymerase elongation and be accompanied by a backwards slippage of the newly synthesized strand on the repeated template followed by further nucleotide incorporation. If the advance of the polymerase through the repeated region was punctuated by such episodes, sizeable expansions could be created even though the number of repeats added per slippage event was small. It is likely that with increasing repeat number more secondary structures will be encountered and the complexity of these structures could increase. A higher mutation rate and an increase in the mean number of repeats added per expansion event could result.

We propose that as the polymerase advances through a repeated region containing L repeats there is a probability $1-\alpha$ that it successfully replicates a given trinucleotide without being blocked. The number of triplets for which a block occurs therefore has a binomial distribution with parameters L and α , which we approximate by a Poisson distribution with mean

$T = L\alpha$. When a block is encountered, the original triplet is copied and a random number of additional repeats (N) are added by the stutter event. Our stochastic model ignores the rare contraction events since they account for less than 4% of the mutations of HD alleles. We suppose that N has a geometric distribution with parameter p , so that N can take any value from 0,1,2... The mean number of additional repeats added per stutter event is $(1-p)/p$. The probability p can be thought of as the chance that the stutter event results only in the copying of the original triplet with no additional repeats added. We expect that p will decrease as the number of repeat units in an allele increases due to increased complexity of secondary structures.

Under our model, the probability distribution of the number of repeats added is given by the distribution of a Poisson sum of geometrically distributed random variables, known as the Pólya-Aeppli law (32). Each HD and intermediate somatic allele has its own values of the parameters $T = L\alpha$ and p . These are estimated from the sperm typing data shown in Figure 3 using the method of maximum likelihood. A recursive algorithm for calculating the required probability distributions was used (32).

The estimated values for T , p and α and their approximate standard errors are shown in Table 2. A detailed study of the size distribution of the observed expansions with the distribution of the sizes predicted by our model, using standard χ^2 goodness-of-fit tests, shows that the fit of the model is satisfactory in all cases. (None of the χ^2 tests for lack of fit of observed to expected frequencies was significant at the 5% level). We note that, based on our single molecule dilution experiments of somatic DNA, our model ignores the small effect that possible PCR artifacts could have on the observed mutation frequency and spectra estimates.

The average number of additional repeats added per stutter event $[(1-p)/p]$ was found to be directly proportional to repeat size: 0.2 (36 allele), 0.4 (38 allele), 2.5 (49 allele) and 3.2 (51 allele). We take this to mean that the larger the repeat number, the more complex the blocks are on average. It is this increased complexity that results, on average, in more repeats added with each slippage event. It is interesting to note in Table 2 that there is no statistically significant difference among allele sizes in the probability α . Thus the chance that a block is encountered over the range of 36 to 51 repeats appears to remain the same per triplet replication, while the average total number of blocks encountered of course increases with allele size, as does the number of repeats added per slippage event.

We tested our data against a variation of the above model. In this model the number of repeats that can be added when a slippage event occurs was uniformly distributed over the number of template triplets that had already been replicated. Under this model the number of repeats that can be added is, on average, greater as replication of the repeated region nears completion since more template repeats are available to anneal with the 3' end of the nascent strand in a slippage event. For values of the parameters that make the average predicted expansion size equal to the observed average size, the variance in the size distribution for each allele is much greater than observed in the data and suggests that this model does not adequately describe the data.

In conclusion, single sperm as well as single molecule and small pool (33) PCR studies have the potential to provide

detailed information on mutation frequency spectra for individual allele sizes. Discrete count data based on large sample sizes can be useful for modeling the molecular details of the trinucleotide repeat mutation process and assist in formulating hypotheses to be tested by biochemical, biophysical and genetic studies.

Preimplantation disease diagnosis

Since the PCR protocol used here can amplify alleles with 100 triplet repeats from single sperm cells, and analysis can be done on agarose gels without radioactivity, this method may be applied to preimplantation genetic diagnosis. It must first be demonstrated that the efficiency of amplification of the normal allele is not so high as to preclude observation of the much larger HD allele for diploid cells. That this is likely is supported by the detection of both alleles in PCR studies on mixtures of 20 sperm from affected individuals.

MATERIALS AND METHODS

Single sperm isolation and preparation

Single sperm cells were isolated (34) into a 96-well microtiter dish using a FacStar Plus Fluorescence Activated Cell Sorter, lysed for 10 min at 65°C in 5.0 µl 200 mM KOH, 50 mM DTT, and neutralized (7) with 5.2 µl 200 mM Tricine, pH 4.93 for a final pH of 8.7.

PCR

Two rounds of PCR were designed to amplify the HD triplet repeat regions using a nested strategy and primers

IT1 (CGACCCTGGAAAAGCTGATGAA),

IT2 (CTTTGGTGGTGCAGCGGCTCT),

IT3 (GAAGGCCTTCGAGTCCCTCAAGT),

and non fluoresceinated or fluorescein labeled

IT4 ([F]-GGCGGCTGAGGAAGCTGAGGA).

An 8-allele polymorphic microsatellite marker 200 kb telomeric to IT15 (35) was coamplified to detect contamination, to detect meiotic segregation and to identify the original chromosome on which the mutation occurred.

To distinguish between CAG and CCG length mutations (9,10,36,37) primers ITCCG (CAGCAGCAGCAGCAGCAACAGCC) and non-fluorescein labeled IT4 were used in separate second round PCR reactions.

All PCRs were performed in 50 µl volume, overlaid with 40 µl light mineral oil, in 96-well microtiter plates (Falcon) on MJ Research thermal cyclers. All second round PCRs employed 2 µl of first round product as template.

First round PCR

Final concentrations for first round PCR were 100 µM each of dATP, dCTP, dTTP and 7-deaza dGTP, 10% (v:v) glycerol, 3% (V:V) DMSO, 1 U rTth DNA Polymerase (Perkin-Elmer), 1.2 mM Mg(OAc)₂, 0.5 µM each of IT1 and IT2 and 0.05 µM each of D4S127 primers P1 and P2 (38). 80 mM KOAc was added for a final K⁺ concentration of 100 mM. The final pH was 8.7.

The thermal cycling profile for first round PCR was 94°C 5 min initial denaturation: 26 cycles of: 94°C 30 s denaturation, 60°C 5 min for the first eight cycles and 3 min for the next 18 cycles and a 72°C 5 min final extension.

Second round PCR amplification encompassing both (CAG)_n and (CCG)_n

Final concentrations for this second round PCR were 100 µM of each dNTP, 10% (V:V) glycerol, 3% (v:v) DMSO, 1 U *Taq* DNA Polymerase (Promega), 50 mM KCl, 10mM Tris-HCl (pH 8.3), 2.5 mM MgCl₂, 100 µg/ml gelatin and 0.5 µM each of IT3 and F-IT4. The thermal cycling profile was 94°C 5 min initial denaturation, 20 or 21 cycles of: 94°C 30 s denaturation, 68°C 1.5 min annealing 72°C 1 min extension and a 72°C 5 min final extension.

Second round PCR amplification encompassing (CCG)_n only

Final concentrations for this second round PCR were as above, except using 2.5% (V:V) DMSO, no glycerol, and 0.2 µM of each primer ITCCG and IT4. The thermal cycling profile was 94°C 5 min initial denaturation, 21 cycles

of: 94°C 10 s denaturation, 60°C 1 min annealing and extension and a 72°C 5 min final extension.

Second round PCR amplification of the microsatellite D4S127

Final concentrations for this second round PCR were as for the amplification of the HD triplet region incorporating both (CAG)_n and (CCG)_n, except using 0.5 µM each of D4S127 P2 and P3 primer (38).

Twenty one rounds of thermal cycling (denaturation, 94°C, 10 s, annealing and extension, 60°C, 1 min) were preceded by a 5 min 94°C denaturation step, and followed by a 5 min final extension at 72°C.

Analysis of PCR product lengths

Ten microlitres of PCR product was first analyzed on agarose gels for presence of a product. Alleles greater than 30 repeats were then run on a A.L.F. ReadyMix Gel (6% acrylamide, 7M urea, 0.6×TBE), 0.5 mm wide short plate gels (approx. 11.5 cm long) using the A.L.F.TM Fragment Manager (Pharmacia) with internal standards. 1.2 µl PCR product and 0.6 fmol of each internal molecular weight standard, in 42% formamide were denatured at 95°C for 2 min prior to loading, and run at 600V, 50°C. Allele size was defined as the peak with the greatest area.

Alleles in the 15–30 repeat size range were analyzed on either 1.0 mm or 1.5 mm thick, 8 cm long, 0.5× or 1×TBE, nondenaturing, 8% polyacrylamide gels run at 120–160V (1–3 h) and stained in 0.5 µg/ml ethidium bromide.

ACKNOWLEDGEMENTS

The authors thank Nancy Huyn and Laura Macias for technical assistance and Robert G.Korneluk for materials. This work was supported by grants R37 GM37645 (NA), NS16367 (JFG), NS22031 (NW) and GM36232 (ST) from the National Institutes of Health and grant DMS90–05833 (ST) from the National Science Foundation.

REFERENCES

- Sutherland, G.R., and Richards, R.I. (1995) Simple tandem DNA repeats and human genetic disease. *Proc. Natl Acad. Sci. USA* **92**, 3636–3641.
- Kawaguchi, Y., Okamoto, T., Tanikawa, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., Narumiya, S., and Kakizuka, A. (1994) CAG expansion in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nature Genet.* **8**, 221–228.
- Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., Folstein, S., Ross, C., Franz, M., Abbott, M., Gray, J., Conneally, P., Young, A., Penney, J., Hollingsworth, Z., Shoulson, I., Lazzarini, A., Falek, A., Koroshetz, W., Sax, D., Bird, E., Vonsattel, J., Bonilla, E., Alvir, J., Bickham Conde, J., Cha, J.-H., Dure, L., Gomez, F., Ramos, M., Sanchez-Ramos, J., Snodgrass, S., de Young, M., Wexler, N., Moscovitz, C., Penchaszadeh, G., MacFarlane, H., Anderson, M., Jenkins, B., Srinidhi, J., Barnes, G., Gusella, J., and MacDonald, M. (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genet.* **4**, 387–392.
- MacDonald, M.E., Barnes, G., Srinidhi, J., Duyao, M.P., Ambrose, C.M., Myers, R.H., Gray, J., Conneally, P.M., Young, A., Penney, J., Shoulson, I., Hollingsworth, Z., Koroshetz, W., Bird, E., Vonsattel, J.P., Bonilla, E., Moscovitz, C., Penchaszadeh, G., Brzustowicz, L., Alvir, J., Bickham Conde, J., Cha, J.-H., Dure, L., Gomez, F., Ramos-Arroyo, M., Sanchez-Ramos, J., Snodgrass, S.R., de Young, M., Wexler, N.S., MacFarlane, J., Anderson, M.A., Jenkins, B., and Gusella, J.F. (1993) Gametic but not somatic instability of CAG repeat length in Huntington's disease. *J. Med. Genet.* **30**, 982–986.
- Telenius, H., Kremer, B., Goldberg, Y.P., Theilmann, J., Andrew, S.E., Zeisler, J., Adam, S., Greenberg, C., Ives, E.J., Clarke, L.A., and Hayden, M.R. (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nature Genet.* **6**, 409–414.
- Zhang, L., LeeFlang, E.P., Yu, J., and Arnheim, N. (1994) Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene. *Nature Genet.* **7**, 531–535.
- Zhang, L., Fischbeck, K., and Arnheim, N. (1995) CAG repeat mutations in sperm from a patient with Kennedy's disease. *Hum. Mol. Genet.* **4**, 303–305.
- Zuhlke, C., Riess, O., Bockel, B., Lange, H., and Thies, U. (1993) Mitotic stability and meiotic variability of the (CAG)_n repeat in the Huntington disease gene. *Hum. Mol. Genet.* **2**, 2063–2067.

9. Rubinsztein, D.C., Barton, D.E., Davison, C.B.C., and Ferguson-Smith, M.A. (1993) Analysis of the huntington gene reveals a trinucleotide-length polymorphism in the region of the gene that contains two CCG-rich stretches and a correlation between decreased age of onset of Huntington's disease and CAG repeat number. *Hum. Mol. Genet.* **2**, 1713-1715.
10. Andrew, S.E., Goldberg, Y.P., Theilmann, J., Zeisler, J., and Hayden, M.R. (1994) A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene, implications for diagnostic accuracy and predictive testing. *Hum. Mol. Genet.* **3**, 65-67.
11. Trotter, Y., Biancalana, V., and Mandel, J.-L. (1994) Instability of CAG repeats in Huntington's disease, relation to parental transmission and age of onset. *J. Med. Genet.* **31**, 377-382.
12. Novelletto, A., Persichetti, F., Sabbadini, G., Mandich, P., Bellone, E., Ajmar, F., Pergola, M., Del Senno, L., MacDonald, M.E., Gusella, J.F., and Frontali, M. (1994) Analysis of the trinucleotide repeat expansion in Italian families affected with Huntington disease. *Hum. Mol. Genet.* **3**, 93-98.
13. Legius, E., Cuppens, H., Dierick, H., Van Zandt, K., Dom, R., Fryns, J.-P., Evers-Kiebooms, G., Decruyenaere, M., Demyttenaere, K., Marynen, P., and Cassiman, J.-J. (1994) Limited expansion of the (CAG)_n repeat of the Huntington Gene: A premutation? *Eur. J. Hum. Genet.* **2**, 44-50.
14. Myers, R.H., MacDonald, M.E., Koroshetz, W.J., Duyao, M.P., Ambrose, C.M., Taylor, S.A.M., Barnes, G., Srinidhi, J., Lin, C.S., Whaley, W.L., Lazzarini, A.M., Schwarz, M., Wolff, G., Bird, E.D., Vonsattel, J.-P.G., and Gusella, J.F. (1993) De novo expansion of a (CAG)_n repeat in Huntington's disease. *Nature Genet.* **5**, 168-173.
15. Goldberg, Y.P., Kremer, B., Andrew, S.E., Theilmann, J., Graham, R.K., Squitieri, F., Telenius, H., Adam, S., Sajoo, A., Starr, E., Heiberg, A., Wolff, G., and Hayden, M.R. (1993b) Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. *Nature Genet.* **5**, 174-179.
16. Goldberg, Y.P., Andrew, S.E., Theilmann, J., Kremer, B., Squitieri, F., Telenius, H., Brown, J.D., and Hayden, M.R. (1993a) Familial predisposition to recurrent mutations causing Huntington's disease, genetic risk to sibs of sporadic cases. *J. Med. Genet.* **30**, 987-90.
17. MacDonald, M.E., Novelletto, A., Lin, C., Tagle, D., Barnes, G., Bates, G., Taylor, S., Allitto, B., Altherr, M., Myers, R., Lehrach, H., Collins, F.S., Wasmuth, J.J., Frontali, M., and Gusella, J.F. (1992) The Huntington's disease candidate region exhibits many different haplotypes. *Nature Genet.* **1**, 99-103.
18. Monckton, D.G., Neumann, R., Guram, T., Fretwell, N., Tamaki, K., MacLeod, A., and Jeffreys, A.J. (1994). Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nature Genet.* **8**, 162-169.
19. Telenius, H., Almqvist, E., Kremer, B., Spence, N., Squitieri, F., Nichol, K., Grandell, U., Starr, E., Benjamin, C., Castaldo, I., Calabrese, O., Anvert, M., Goldberg, Y.P., and Hayden, M.R. (1995) Somatic mosaicism in sperm is associated with intergenerational (CAG)_n changes in Huntington disease. *Hum. Mol. Genet.* **4**, 189-195.
20. Chung, M.-Y., Ranum, L.P.W., Duvick, L.A., Servadio, A., Zoghbi, H.Y., and Orr, H.T. (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nature Genet.* **5**, 254-258.
21. Leeflang, E.P., and Arnheim, N. (1995) A novel repeat structure at the myotonic dystrophy locus in a 37 repeat allele with unexpectedly high stability. *Hum. Mol. Genet.* **4**, 135-136.
22. Eichler, E.E., Holden, J.J.A., Popovich, B.W., Reiss, A.L., Snow, K., Thibodeau, S.N., Richards, C.S., Ward, P.A., and Nelson, D.L. (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature Genet.* **8**, 88-94.
23. Kunst, C.B., Warren, S.T. (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* **77**, 853-861.
24. Jansen, G., Willems, P., Coerwinkel, M., Nillesen, W., Smeets, H., Vits, L., Howeler, C., Brunner, H., and Wieringa, B. (1994) Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in (CTG)_n repeat variation and selection against extreme expansion in sperm. *Am. J. Hum. Genet.* **54**, 575-585.
25. Richards, R.I., and Sutherland, G.R. (1994) Simple repeat DNA is not replicated simply. *Nature Genet.* **6**, 114-116.
26. Aaltonen, L.A., Peltomaki, P., Leach, F.S., Sistonen, P., Pylkkanen, L., Mecklin, J.-P., Jarvinen, H., Powell, S.M., Jen, J., Hamilton, S.R., Petersen, G.M., Kinzler, K.W., Vogelstein, B., and de la Chapelle, A. (1993) Clues to the pathogenesis of familial colorectal cancer. *Science* **260**, 812-816.
27. Peltomaki, P., Aaltonen, L.A., Sistonen, P., Pylkkanen, L., Mecklin, J.-P., Jarvinen, H., Green, J.S., Jass, J.R., Weber, J.L., Leach, F.S., Petersen, G.M., Hamilton, S.R., de la Chapelle, A., and Vogelstein, B. (1993) Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **260**, 810-812.
28. Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**, 558-561.
29. Strand, M., Prolla, T.A., Liskay, R.M., and Petes, T.D. (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274-276.
30. Sinden, R.R. and Wells, R.D. (1992) DNA structure, mutations, and human genetic disease. *Current Opinion in Biotechnology* **3**, 612-622.
31. Mitas, M., Yi, A., Dill, J., Kamp, T.J., Chambers, E.J., and Haworth, I.S. (1995). Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)₁₅. *Nucleic Acids Res.* **23**, 1050-1059.
32. Johnson, N.L. and Kotz, S. (1969) *Discrete Distributions*. Wiley, New York, p. 197.
33. Monckton, D.G., Wong, L.-J. C.A.T., and Casky, C.T. (1995) Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analysis. *Hum. Mol. Genet.* **4**, 1-8.
34. Leeflang, E.P., Hubert, R., Schmitt, K., Zhang, L., and Arnheim, N. (1994) Single Sperm Typing. *Current Protocols in Human Genetics* **1**, 1.6.1-1.6.15.
35. Taylor, S.A.M., Barnes, G.T., MacDonald, M.E., and Gusella, J.F. (1992) A dinucleotide repeat polymorphism at the D4S127 locus. *Hum. Mol. Genet.* **1**, 142.
36. Huntington's Disease Collaborative Research Group. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971-983.
37. Rubinsztein, D.C., Leggo, J., Barton, D.E., and Ferguson-Smith, M.A. (1993) Site of (CCG) polymorphism in the HD gene. *Nature Genet.* **5**, 214-215.
38. Hubert, R., MacDonald, M., Gusella, J., and Arnheim, N. (1994) High resolution localization of recombination hot spots using sperm typing. *Nature Genet.* **7**, 420-424.