Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 4833–4837, May 1995
Medical Sciences

An *n*-allele model for progressive amplification in the *FMR1* locus

(fragile X syndrome/mutation)

A. MORRIS*, N. E. MORTON*, A. COLLINS*, J. MACPHERSON*, D. NELSON†, AND S. SHERMAN‡

*Human Genetics Group, University of Southampton, Level G, Princess Anne Hospital, Southampton SO9 4HA, United Kingdom; †Institute of Molecular Genetics, Baylor College of Medicine, Houston, TX 77030; and ‡Department of Genetics and Molecular Medicine, Emery University, Atlanta, GA 30322

Contributed by N. E. Morton, December 19, 1994

ABSTRACT **An *n*-allele model is developed for the *FMR1* locus, which causes the fragile X syndrome, where *n* is the number of triplet repeats in the first exon. Frequencies in the general population and in index families are used to generate an *n* to *n* + δ transition matrix that predicts specific risks in satisfactory agreement with observation. However, until sequencing distinguishes between stable and unstable alleles with the same value of *n*, it is premature to infer whether allelic frequencies at the *FMR1* locus are at equilibrium or, as some have suggested, are evolving toward higher frequencies of the pathogenic allele.**

The fragile X syndrome [Fra(X)] of mental retardation is the commonest disease of progressive amplification, in which an increasing insert leads to failure of gene expression. A four-allele model accounts for major features of Fra(X) (including gene frequencies, mutation rates, and genetic risks) in terms of triplet repeats in the *FMR1* locus (1). The *N* allele seldom has more than 40 repeats and is stable except for rare mutation to allele *S*, which has more than 40 repeats and converts to an unstable allele *Z* at the rate of about 0.01 per generation. The *Z* allele converts at the rate of 74% in the female germ line to the "full mutation" allele *L* characterized by methylation, late replication, loss of expression of the *FMR1* gene, and the Fra(X) phenotype. This conversion of the maternal X chromosome may occur in the early zygote, and *L* alleles are somatically unstable (2, 3). The *Z* allele typically has more than 60 repeats, while the *L* allele usually has more than 200 (Fig. 1). Poor distinction between *Z* "premutations" with a high risk for Fra(X) offspring and "intermediate" *S* alleles with infinitesimal risk is troublesome in clinical genetics. Although $(CGG)_n$ sequences are more unstable than less monotonous repeats of the same length (4–6), too little sequencing has yet been done to model these qualitative effects. We are concerned here with frequencies and quasicontinuous effects of repeat length.

MATERIALS AND METHODS

From data on fragment lengths in normal populations we estimate frequencies of large values of *n*, assuming a monotonic discrete distribution. Risks for unstable transmission are assumed to have a logistic distribution dependent on fragment size. Parameters were estimated by maximum likelihood (ML), using Newton–Raphson iteration with exact derivatives. Because of evidence that the conventional molecular standard underestimates length of $(CGG)_n$ by one repeat and that the mode of *n* is 30, we added 1 to each value of *n* in samples with the mode at 29 (7).

RESULTS

Frequencies in the General Population. A sample of *T* chromosomes from the general population is expected to have a frequency $Q_S = u/k$ for allele *S*, where the mutation rate from *N* to *S* has been estimated as $u = 24.7 \times 10^{-5}$ per generation and the mutation rate from *S* to *Z* is $k \sim 0.01$ per generation (1). Therefore Q_S is about 0.025. The frequency of alleles *Z* and *L* is much smaller and may be lowered by exclusion of retarded individuals and their close relatives (Table 1).

Table 1. Definitions and estimated frequencies (1)

Allele	Type	Methylation	Instability	Mean gene frequency (Q)
N	Normal	0	0	0.9745
S	Intermediate	0	+	0.0247
Z	Premutation	0	++	0.0005
L	Full mutation	+	+++	0.0003

$Q = (q_m + 2q_f)/3$, where q_m and q_f are frequencies in males and females.

The tail of the distribution of the control sample shows a gradual decline in frequency with many empty cells (Table 2). To obtain a reasonable representation of the tail we must fit a theoretical distribution $f(m)$, where $m = n - n_0 + 1$ for $m = 1, \ldots \infty$. Here *n* is the number of repeats and $n_0, n_0 + 1, \ldots$ are the repeat numbers in the tail, without implying any relation between n_0 and the thresholds for alleles *S* and *Z*. Since the sample size *M* in the tail is small, analysis is confined to distributions with only a single parameter *q*. Denoting the observed frequency by $g(m)$ and $\Sigma g(m) = M$, the *t*th Newton–Raphson iteration is

$$U = \sum_m g(m)\left[\frac{\partial \ln f(m)}{\partial q}\right]$$

$$K = \sum g(m)\left[-\frac{\partial^2 \ln f(m)}{\partial q^2}\right] \qquad [1]$$

$$q^{(t)} = q^{(t-1)} + U/K\big|_q^{(t-1)}.$$

At convergence the likelihood ratio χ^2 is $2\Sigma g(m) \ln[g(m)/Mf(m)]$. The large proportion of cells with small expectations makes the degrees of freedom approximately equal to one less than the number of values of *m* preceding the terminal vector with $g(m) = 0$. Given *r* estimates of *q*, a pooled value is obtained by taking $\bar{q} = \Sigma qK/\Sigma K$, and heterogeneity is tested by $\chi_{r-1}{}^2 = \Sigma U^2/K - (\Sigma U)^2/\Sigma K$.

We considered two discrete frequency distributions $f(m)$ for tails ($m = 1, 2, \ldots \infty$; $0 < q < 1$), the logarithmic and the truncate geometric.

$$\text{Logarithmic: } \left[\frac{-1}{\ln(1-q)}\right]\frac{q^m}{m}.$$

$$\text{Truncate geometric: } q^{m-1}(1-q).$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Fra(X), fragile X syndrome; ML, maximum likelihood.

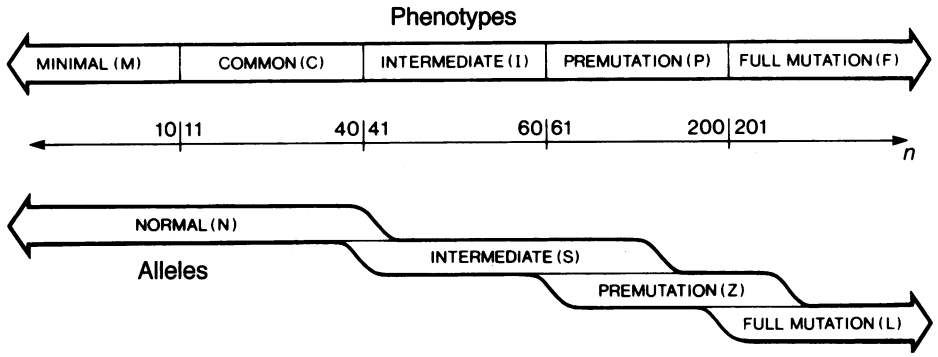

FIG. 1. Relation between alleles and phenotypes (1, 8).

Both fit well, but extrapolation to large alleles is risky (Table 3). Either density may be used to infer other properties of the *FMR1* locus, estimating the distribution of n in the general population as

$$f(n) = \begin{cases} g(n)/T \text{ for } n < n_0 \\ Mf(m)/T \text{ for } n \geq n_0, \end{cases} \quad [2]$$

where $g(n)$, T, and M are taken from the Wessex sample for which $n_0 = 35$ (8).

Transition Probabilities. We wish to estimate the matrix $T_{n(n+\delta)}$ of transition from n to $n + \delta$, with sequence effects neglected for lack of detailed information. $T_{n(n+\delta)}$ is square, asymmetric, and row stochastic ($\Sigma_\delta T_{n(n+\delta)} = 1$ for all n). We assume that only two risk categories need be considered, from Z alleles in females (Z_f) and the remainder ($\overline{Z}_f$) that includes N and S alleles in both sexes and Z alleles in males. For each category the possible outcomes E_i are grouped into four classes ($i = 1, \ldots 4$) corresponding to $\delta = 0$, $\delta < 0$, $0 < \delta < 200 - n$, and $\delta \geq 200 - n$. The last event includes expansions of Z_f and $\overline{Z}_f$ alleles to premutations with more than 199 repeats which are assigned to $n + \delta = 200$ as well as expansions to the full mutation which occur only in Z_f and are assigned to $n + \delta = 201$ with $T_{201(201)} = 1$. Rare instances of $\delta \leq -n$ are assigned to $n + \delta = 1$. Therefore $T_{n(n+\delta)}$ is of order 201 and each element (for $n \leq 200$) is of the form

$$P(E_i|n) = P(Z_f|n)P(E_i|Z_f, n) + [1 - P(Z_f|n)]P(E_i|\overline{Z}_f, n). \quad [3]$$

To estimate these elements we take

$$P(Z_f|n) = P(Z_f|Z)P(Z|f, n), \quad [4]$$

where $P(Z_f|Z)$ is the proportion of Z alleles in females, or $6/(9 - y) = 0.73$ (ref. 1, Table 2), and $P(Z|f, n)$ is the probability that an allele with n repeats drawn randomly from a female is Z. These probabilities reflect transmission of Z alleles from fathers exclusively to daughters. In the next sections we derive and estimate conditional probabilities.

Meiotic Instability. Fra(X) families are enriched for conversion of Z to L. To allow for this we delete probands and let $l(n)$ be the residual number of full mutations among $t(n)$ transmissions from Z carrier females to nonprobands (Table 4). Then the frequency of conversion of Z to L is a sigmoid function that will be approximated by

$$z(n) = E[l(n)/t(n)] = 1/(1 + e^{a - b[\ln(n)]}). \quad [5]$$

Pooling the three Fra(X) samples, the ML estimates are $a = 65.155 \pm 6.782$ and $b = 14.955 \pm 1.558$ with $\chi^2 = 18.17$, whereas the linear logistic gave $\chi^2 = 23.51$.

One study (9) recorded the number of unstable transmissions $u(n)$ among observed transmissions $t(n)$ in the normal population. We assume that this can be approximated in both sexes by

$$s(n) = E[u(n)/t(n)] = 1/(1 + e^{w - xn}). \quad [6]$$

Table 2. Analysis of large numbers of repeats in control chromosomes

Statistic	Macpherson *et al.* (8)		Reiss *et al.* (9)		Brown *et al.* (10)		S. L. Sherman (personal communication)		Fu *et al.* (11)	
	n	$g(n)$	n	$g(n)$	n	$g(n)$	n	$g(n)$	n	$g(n)$
	35	2	47	3	42	4	41	3	40	6
	36	2	48	2	43	4	43	1	41	5
	37	1	49	4	44	1	45	1	42	1
	39	1	50	1	47	1	46	1	44	1
	40	1	51	1	48	1	48	1	45	2
	43	1	52	1	49	1	52	1	46	1
	49	1	53	1	52	2	54	1	47	1
	50	1	76	1			57	1	55	1
q for logarithmic	0.946		0.931		0.908		0.955		0.892	
q for truncate geometric	0.833		0.803		0.759		0.853		0.731	
χ^2 for logarithmic	14.71		16.08		15.57		18.51		13.54	
χ^2 for truncate geometric	13.56		17.06		15.17		17.33		14.29	
Degrees of freedom	15		29		10		16		15	
Sample size (T)	188		1538		570		263		492	
Tail size (M)	10		14		14		10		18	
$P(n > 60)$ for logarithmic	0.00190		0.00071		0.00068		0.00292		0.00044	
$P(n > 60)$ for truncate geometric	0.00046		0.00042		0.00017		0.00158		0.00005	

Table 3. Comparison of tail distributions for control chromosomes

Statistic	df	Logarithmic	Truncate geometric
Total χ^2	85	78.41	77.41
χ^2 for pooled $g(m)$	29	34.69	29.89
Heterogeneity χ^2	4	7.88	8.42
$\bar{q}$	—	0.934	0.804
$\Sigma_r K$	—	375.1	1315.2
$P(n > 60)$*	—	0.00115	0.00045

*Weighted by m.

Table 4 gives $w = 10.863 \pm 3.891$ and $x = 0.177 \pm 0.076$ (Fig. 2).

Several studies have attempted to estimate the magnitude of transitions in the normal population (3, 8–10, 11, 13). However, changes of only one or two repeats may be confounded with errors of estimation and are least likely to be detected in stable N alleles. We therefore excluded the sole observation (14) from alleles with less than 41 repeats, where the bias would be expected to be greatest (Table 5).

Excluding full mutations ($Z \rightarrow L$), a transition from n to $n + \delta$ for $\delta \neq 0$ and $d = |\delta|$ can be decomposed into two probabilities. On the assumption that half of transitions are to smaller sizes when $d = 1$, but the frequency becomes negligible as d increases, the first probability will be approximated by

$$P(\delta < 0 | d \neq 0) = 1/(1 + e^{\beta(d-1)}). \quad [7]$$

The data of Table 5 give $\beta = 0.295 \pm 0.138$. The conditional frequency of negative transitions is estimated to be 0.4 at $d = 3$ and 0.1 at $d = 9$. The second probability distribution is $P(d|d > 0)$. The variance is too great for a Poisson distribution (Table 5), and so we assume a truncate geometric or

$$P(d|d \neq 0, n) = p^{d-1}(1 - p), \quad [8]$$

where $p = 1/(1 + e^{s-tn})$. From Table 5 we estimate $s = 1.296 \pm 0.875$, $t = 0.060 \pm 0.012$. In principle, separate estimates should be made for each sex, but the numbers and differences in Table 5 are too small to warrant this at present.

We need one further pair of estimates, for $P(Z|f, n) = f(u, v)$. Observations $h(n)$ on the distribution $k(n)$ of Z alleles in mothers of Fra(X) probands are informative but propagate errors in $f(n)$ and $z(n)$. We therefore combined $f(n)$ in Wessex (Table 2) with $k(n)$ in Table 6, taking

$$k(n) = f(n)P(Z|f, n)z(n)/\Sigma f(n)P(Z|f, n) \quad [9]$$

and imposing constraints on Q_z and y as a trinomial pseudosample of size N with likelihood $r^{NQ_z y} s^{NQ_z(1-y)} t^{N(1-Q_z)}$, where $r = \Sigma f(n)P(Z|f, n)z(n)$, $s = \Sigma f(n)P(Z|f, n)[1 - z(n)]$, and $t = \Sigma f(n)[1 - P(Z|f, n)]$. We took $N = 10^6$, $Q_z = 0.0005$, and $y = 0.74$. The function $P(Z|f, n) = 1/(1 + e^{u-vn})$ gave the same likelihood with the truncate geometric and logarithmic distributions for $f(n)$, but with different estimates. Replacement of n by n^2 or $\ln(n)$ did not alter the likelihood. We therefore retained the linear estimates, which were $q = 0.944$, $u = 27.356$, $v = 0.355$ for the logarithmic and $q = 0.901$, $u = 24.787$, $v = 0.306$ for the truncate geometric distribution (Fig. 3). Although these and previous estimates are not exact, we shall use them as parameters for lack of better alternatives. A direct estimate of $P(Z|f, n)$ can be made when the sequences that determine the Z alleles have been defined.

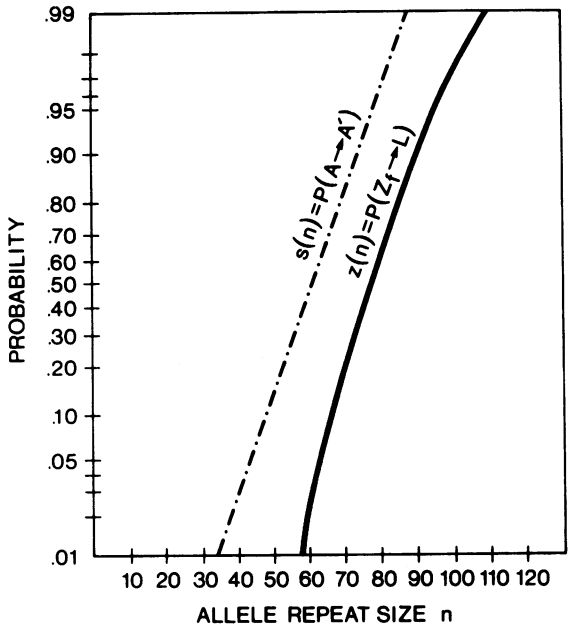

FIG. 2. Instability of repeat number.

Tests of Predictions. The three samples in Table 2 with $n_0 > 40$ support 41 as the lower limit of S alleles. They give 25/943 = 0.0265 as the pooled estimate of $Q_s + Q_z$, in close agreement with the preliminary estimate of 0.0252 in Table 1. An alternative expression is $\Sigma_{n=41}^{200} f(n)$, which also agrees well (Table 7). There is enforced agreement with $\Sigma f(n)P(Z|f, n) = Q_z = 0.0005$. The estimate of $\Sigma f(n)P(Z|f, n)z(n)/Q_z = y = 0.74$ is nearly but not perfectly satisfied.

Unconstrained predictions are of greater weight. The conversion rate from N to S is

$$u = \sum_{n=1}^{40} f(n) \sum_{n+\delta=41}^{200} T_{n(n+\delta)}/Q_N \quad [10]$$

Table 4. Instability of meiotic transmissions

Control chromosomes [Reiss *et al.* (9)]			Fra(X) families [Heitz *et al.* (12)]			Fra(X) families [Fu *et al.* (11)]			Fra(X) families [S. L. Sherman (personal communication)]		
n	Total $t(n)$	Unstable $u(n)$	n	Total $t(n)$	Full mutation $l(n)$	n	Total $t(n)$	Full mutation $l(n)$	n	Total $t(n)$	Full mutation $l(n)$
19	12	0	70	20	2	67	5	0	60	1	0
24	12	0	80	16	9	71	3	1	70	2	2
29	18	0	97	50	45	74	6	4	76	2	1
34	7	0	113	24	24	78	2	2	86	5	4
39	11	0	130	21	21	81	3	2	106	14	14
44	3	0				84	8	7			
49	11	2				87	3	2			
61	3	1				91	3	3			
						94	11	11			

Table 5. Transitions in unstable transmissions ($S \rightarrow S'$, Z and $Z \rightarrow Z'$)

Ref.	Sex	n	δ	Ref.	Sex	n	δ
9	f	49	+1	11	f	74	+40
10	f	52	−5	11	f	74	+97
11	f	53	+21	3	f	75	+13
11	f	55	−2	3	f	75	+26
11	f	55	+3	3	f	75	+96
11	f	55	+4	11	f	81	+20
11	f	55	+4	11	f	84	+10
11	f	55	+6	11	f	87	+40
11	f	60	−5	11	f	87	+107
11	f	67	+7	13	f	91	+25
11	f	67	+7	9	m	47	+6
11	f	67	+14	10	m	56	+4
11	f	67	+20	10	m	56	+12
11	f	67	+44	11	m	67	+4
10	f	68	+27	11	m	67	+17
11	f	71	−4	11	m	87	+14
11	f	71	+33	11	m	117	+47

and the conversion rate from S to Z is

$$k = \sum_{n=41}^{200} f(n)[1 - P(Z|f, n)] \sum_{n+\delta=41}^{200} T_{n(n+\delta)}P(Z|f, n)/Q_s. \quad [11]$$

Our problem is to estimate $T_{n(n+\delta)}$ from these data. If the geometric distribution could be trusted for $\delta = 0$ we would have $P(\delta = 0|Z, n) = 1 - p$. However, this extrapolation from the truncate distribution is questionable and violates our estimate of $P(\delta = 0|n)$ by Eq. **6**. We therefore take

$$P(\delta = 0|Z_f, n) = \frac{[1 - s(n)](1 - p_f)}{(1 - p_f)P(Z_f|n) + (1 - p_m)P(\overline{Z}_f|n)}$$

and [12]

$$P(\delta = 0|\overline{Z}_f, n) = \frac{[1 - s(n)](1 - p_m)}{(1 - p_f)P(Z_f|n)+(1-p_m)P(\overline{Z}_f|n)},$$

which satisfy $P(\delta = 0|n) = 1 - s(n)$.

To conserve our estimate of $z(n)$ we take

$$P(\delta \geq 200 - n|Z_f, n) = z(n) + [s(n) - z(n)] \sum_{d=200}^{\infty} p(\delta > 0|d \neq 0)P(d|d \neq 0, n)$$

Table 6. Distribution $h(n)$ in progeny-tested Z_f alleles

Heitz *et al.* (12)		S. L. Sherman (personal communication)		Fu *et al.* (11)	
n	$h(n)$	n	$h(n)$	n	$h(n)$
70	4	70	2	67	1
80	20	76	1	71	2
97	68	86	4	74	5
113	35	106	14	78	3
130	18			81	3
140	2			84	8
156	1			87	3
				91	4
				94	8
				101	4
				111	2
				114	1

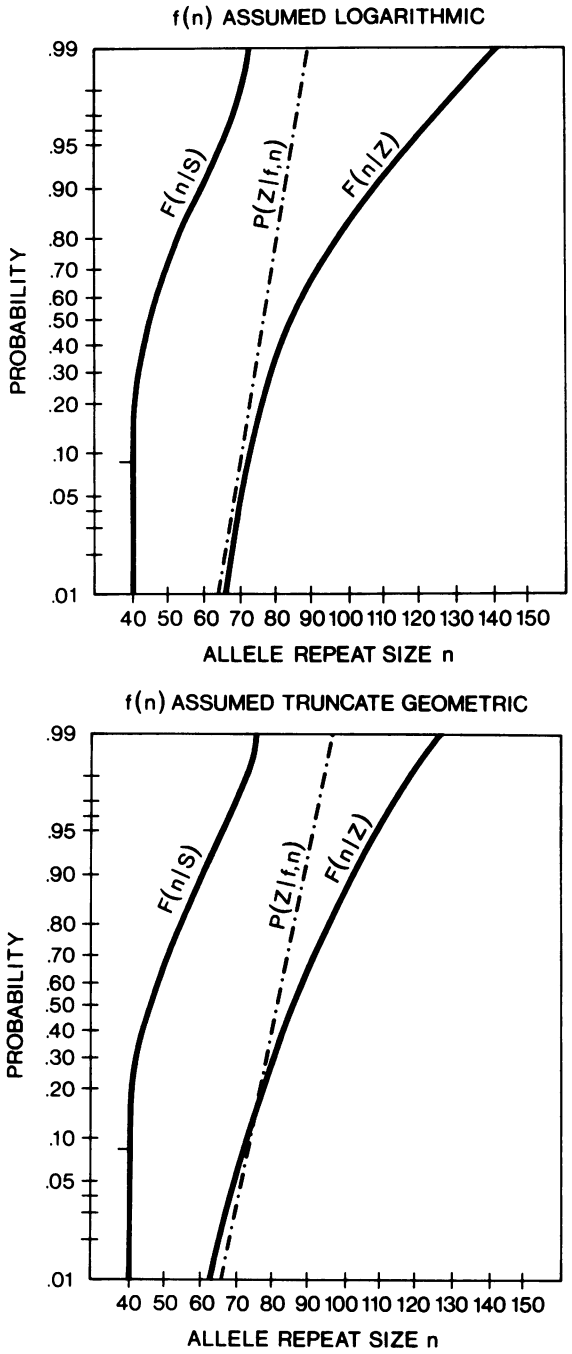

FIG. 3. Predictions of the model.

$$P(\delta \geq 200 - n|\overline{Z}_f, n) = 0 + s(n) \sum_{d=200}^{\infty} P(\delta > 0|d \neq 0)P(d|d \neq 0, n).$$

The first term in each of these expressions is the contribution to $T_{n(201)}$; the remainder is the contribution to $T_{n(200)}$.

We constructed $T_{n(n+\delta)}$ from these results, tentatively taking $p_f = p_m = p$ as noted above. The estimates of k by Eq. **11** and of u by Eq. **10** are in surprisingly good agreement with prediction (Table 7), considering low precision of the transition matrix for small values of n, possible bias in Table 5 toward reporting of large expansions, and arbitrary distinction between N and S alleles. The cumulative distribution of n among S alleles is

Table 7. Tests of predictions

Frequency or rate	Predicted value		
	Ref. 1	Truncate geometric	Logarithmic
$Q_s + Q_z$	0.0252	0.0285	0.0138
Q_z	0.0005	0.0005	0.0005
y	0.74	0.72	0.71
k	0.006–0.049	0.041	0.043
u	24.7×10^{-5}	19.1×10^{-5}	19.3×10^{-5}

$$F(n|S) = \frac{\sum_{41}^{n} f(n)[1 - P(Z|f, n)]}{\sum_{41}^{200} f(n)[1 - P(Z|f, n)]}. \quad [13]$$

Although most S alleles have n less than 60, a fraction are predicted to fall between 60 and 70, with a few larger values (Fig. 3). This conclusion does not depend on the assumed tail distribution $f(n)$ and is consistent with considerable amplification before conversion to Z in a proportion of cases (perhaps by loss of TGG and AGG triplets), but a critical test depends on large alleles not ascertained through Fra(X) probands. Such alleles are rare and at present difficult to sequence.

DISCUSSION

Much of mathematical genetics is now concerned with genetic epidemiology, often directed toward phenotypes determined by multiple mendelian genes. Progressive amplification (together with mitochondrial inheritance, germinal mosaicism, parental imprinting, and uniparental disomy) typifies a complementary class of nonmendelian monogenic disorders. The challenge to genetic epidemiology they currently provide will not disappear with advances in molecular biology unless the risk associated with each sequence is either 0 or 1, and even in that unlikely event risk will have to be determined by observations in families. The number and location of AGG and TGG triplets within the CGG-rich exon 1 of *FMR1* will contribute to risk estimates by discriminating between S and Z alleles of the same length. If risk categories (however defined) are ordered from smallest to largest, the ordinal position can replace n in the theory we have given here and the parameters must be estimated again. Even if our understanding of transition probabilities were fundamentally altered, risks would still be determined by $f(n)$ and $T_{n(n+\delta)}$. There is need for better estimates of these functions in rigorously defined and larger samples.

From these estimates the evolutionary dynamics of *FMR1* alleles can be inferred, but there are several uncertainties. L alleles are quickly eliminated and so constitute an absorbing barrier in the Markov chain. Are small alleles also selected against? The modal number of repeats is about 30 in primates and 10 in other mammals, but individuals with few repeats are rare (15). This may be related to failure of the CGG-binding protein 1 to bind to sequences with fewer than six repeats (16). Since the repeat sequence is transcribed (but not translated), a possible effect on *FMR1* expression cannot be dismissed. Expression studies in individuals with extremely small values of n will be awaited with interest.

A second uncertainty arises from evidence suggesting that decuplet slippage may be favored for the N allele. In the general population there are modes near $n = 20$, 30, and 40 (8). Only one increase of allele N has been reported, from 30 to 39 repeats in an X chromosome received from the father (14). Decuplet slippage would tend to stabilize the distribution of n. An allele that has undergone one change may be at greater risk for a subsequent one, requiring that n be generalized to include sequence-specific risks. Until these problems are studied more closely, the stability of this polymorphism is *sub judice* (5, 17, 18).

Models with 4 alleles (1) or 10 alleles (19) have now been generalized to n alleles with acceptable fit to empirical data. It remains to extend them to sequences and haplotypic associations and to investigate evolutionary dynamics. Our approach is applicable to progressive amplification at other loci and therefore (if data were available) to all repetitive sequences in nuclear DNA.

1. Morton, N. E. & Macpherson, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 4215–4217.
2. Reyniers, E., Vits, L., De Boulle, K., Van Roy, B., Van Velzen, D., de Graaff, E., Verkerk, A. J. M. H., Jorens, H. Z. J., Darby, J. K., Oostra, B. & Willems, P. J. (1993) *Nat. Genet.* **4,** 143–146.
3. Kruyer, H., Mila, M., Glover, G., Carbonell, P., Ballesta, F. & Estevill, X. (1994) *Am. J. Hum. Genet.* **54,** 437–442.
4. Chung, M., Ranum, L. P. W., Duvick, L. A., Servadio, A., Zoghbi, H. Y. & Orr, H. T. (1993) *Nat. Genet.* **5,** 254–258.
5. Kunst, C. B. & Warren, S. T. (1994) *Cell* **77,** 853–861.
6. Hirst, M. C., Grewal, P. K. & Davies, K. E. (1994) *Hum. Mol. Genet.* **3,** 1553–1560.
7. Fisch, G. S., Nelson, D. L., Snow, K., Thibodeau, S. N., Chalifour, M., White, B. N. & Holden, J. J. A. (1994) *Am. J. Med. Genet.* **51,** 339–345.
8. Macpherson, J. N., Bullman, H., Youings, S. A. & Jacobs, P. A. (1994) *Hum. Mol. Genet.* **3,** 399–405.
9. Reiss, A. L., Nelson, D. L., Kazazian, H. H., Krebs, C. M., Boehm, C. D., Abrams, M. T. & McAughan, A. (1994) *Hum. Mol. Genet.* **3,** 393–398.
10. Brown, W. T., Houck, G. E., Jeziorowska, A., Levinson, F. N., Ding, X., Dobkin, C., Zhong, N., Henderson, J., Brooks, S. S. & Jenkins, A. C. (1993) *J. Am. Med. Assoc.* **270,** 1569–1575.
11. Fu, Y.-H., Kuhl, D. P. A., Pissuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J. M. H., Holden, J. J. A., Fenwick, R. G., Warren, S. T., Oostra, B. A., Nelson, D. L. & Caskey, C. T. (1991) *Cell* **67,** 1047–1058.
12. Heitz, D., Devys, D., Imbert, G., Kretz, C. & Mandel, J.-L. (1992) *J. Med. Genet.* **29,** 794–801.
13. Snow, K., Doud, L. K., Hagerman, R., Pergolizzi, R. G., Erster, S. H. & Thibodeau, S. N. (1993) *Am. J. Hum. Genet.* **53,** 1217–1228.
14. Macpherson, J. N., Curtis, G., Crolla, J. A., Dennis, N., Migeon, B., Hirst, M. C., Grewal, P. K., Davies, K. E. & Jacobs, P. A. (1995) *J. Med. Genet.*, in press.
15. Deelen, W., Bakker, C., Halley, D. J. J. & Oostra, B. A. (1994) *Am. J. Med. Genet.* **51,** 513–516.
16. Richards, R. I., Holman, K., Yu, S. & Sutherland, G. R. (1993) *Hum. Mol. Genet.* **2,** 1429–1435.
17. Mandel, J. L. (1994) *Nat. Genet.* **7,** 453–455.
18. Rubinsztein, D. C., Amos, W., Leggo, J., Goodburn, S., Rajhumar, S., Ramesar, R. S., Old, J., Bontrop, R., McMahon, R., Barton, D. E. & Ferguson-Smith, M. A. (1994) *Nat. Genet.* **7,** 525–530.
19. Kolehmainen, K. (1994) *Am. J. Med. Genet.* **51,** 428–435.