Genetic fitness in Huntington's Disease and Spinocerebellar Ataxia 1: a population genetics model for CAG repeat expansions

M. FRONTALI[1], G. SABBADINI[1], A. NOVELLETTO[2], C. JODICE[2], F. NASO[3],
M. SPADARO[3], P. GIUNTI[3], A. G. JACOPINI[4], L. VENEZIANO[1], E. MANTUANO[1],
P. MALASPINA[2], L. ULIZZI[5], A. BRICE[6], A. DURR[6] AND L. TERRENATO[2]

[1] *Istituto di Medicina Sperimentale del CNR, Rome, Italy*
[2] *Dipartimento di Biologia Università Tor Vergata, Rome, Italy*
[3] *Istituto di Clinica delle Malattie Nervose e Mentali, Università La Sapienza, Rome, Italy*
[4] *Istituto di Psicologia del CNR, Rome, Italy*
[5] *Dipartimento di Genetica e Biologia Molecolare, Università La Sapienza, Rome, Italy*
[6] *INSERM U289, Hôpital La Salpêtrière, Paris, France*

SUMMARY

An analysis of genetic fitness was performed in Huntington's Disease (HD) and Spinocerebellar Ataxia 1 (SCA1) families. Two partially overlapping samples were used: clinically defined HD and SCA1 patients from families ascertained in definite geographical areas, and molecularly typed carriers of HD and SCA1 mutations (CAG trinucleotide expansions). In both cases, a control group of normal relatives was used. HD and SCA1 patients born before 1915–20 had more children than normal controls. Carriers of HD and SCA1 mutations, all in the low/medium expansion range (37–49 and 47–54 CAG repeats respectively), had a higher number of children than controls up to more recent times (1935–1950). The reproduction of heterozygotes for large expansions could be analysed only in subjects born after 1950 and provided indirect evidence of a lower than normal number of children. The above results fit a model based on a differential fitness according to the degree of expansion. Such a model predicts that 1) up to relatively recently the frequency of alleles in the low/medium range has been maintained or even increased by the increased fitness of their carriers, as well as by new mutations, and 2) the frequency of large expansions, part of which are lost at each generation, is maintained through further expansions of alleles in the low/medium expansion range. The implications of such a model on linkage disequilibrium and the possible spread of these diseases in future generations are discussed.

INTRODUCTION

Expansion of trinucleotide repeat stretches are found to be responsible for an increasing number of diseases. This new type of mutation is unstable (dynamic), being prone during meiosis to either expansion or, less frequently, contraction. Suc-

cessive increases in the number of repeats along generations result in progressive anticipation of age at onset of the disease, as well as in more severe clinical phenotypes (La Spada *et al.*, 1994; Sutherland & Richards, 1995). CAG repeat expansions have, thus far, been found in Huntington's Disease (HD), Spinocerebellar Ataxia 1 (SCA1) and 3 (SCA3), Dentato-Rubro-Pallido-Luysian Atrophy (DRPLA) and Kennedy Disease (KD) (HDCRG, 1993; Orr *et al.*, 1993; Kawaguchi *et al.*, 1994; Koide *et al.*,

Address for correspondence: M. Frontali, Istituto di Medicina Sperimentale del CNR, Viale Marx 15, 00137 Rome, Italy. Tel:. (+39)-6-86090339. Fax: (+39)-6-86090332. E-mail: fronta@biocell.irmkant.rm.cnr.it

1994; La Spada *et al.*, 1991). The corresponding mutations appear to have several characteristics in common: 1) they arise on a polymorphic CAG stretch contained in the gene coding region; 2) they undergo a relatively limited expansion; and 3) the mutated gene product, while almost ubiquitous, affects only specific areas of the central nervous system (CNS) (La Spada *et al.*, 1994). Other dynamic mutations, such as those responsible for Myotonic Dystrophy (DM), Fragile X Mental Retardation (FRAXA) and Friedreich Ataxia, can arise on non-CAG tri-nucleotide stretches placed in non-coding regions of the gene, undergo very large expansions and affect tissues other than CNS (La Spada *et al.*, 1994; Sutherland & Richards, 1995; Campuzano *et al.*, 1996).

The population genetics of dynamic mutations reveals some puzzling features (Mandel, 1994). In diseases with a reduced genetic fitness, such as DM and FRAXA, a strong or complete linkage disequilibrium between the mutation and intra- or extra-genic markers has been found (Imbert *et al.*, 1993; Richards *et al.*, 1992; Oudet *et al.*, 1993). Under the assumption of mutation/selection equilibrium, the expanded alleles lost at each generation should be replaced by new mutations which are unlikely to arise on a restricted set of haplotypes, unless the latter are in disequilibrium with factors predisposing to mutation. Various hypotheses have been proposed to account for these findings. Carey *et al.* (1995) reported data suggesting a segregation distortion favouring the transmission of a subset of normal DM alleles more prone to expansion. Imbert *et al.* (1993) proposed a multi-step mutational model by which DM expansion arises on a subset of normal alleles (19–30 repeats), originated from a jump in a 5-repeat allele carried by one or a few ancestral haplotypes. Finally, the loss of cryptic variations in unstable CGG stretches of the FMR1 gene appears likely to be a predisposing factor to instability (Kunst & Warren, 1994; Eichler *et al.*, 1994).

Contrary to the case of DM and FRAXA, very little is known about the relative genetic fitness of patients with diseases due to CAG expansions,

with the exception of HD. Most studies indicate that relative genetic fitness of HD patients is either normal or increased, although a decrease has also been reported (for a review see Harper, 1991). If relative genetic fitness in HD patients is indeed equal to 1 or greater, non-random association with haplotypes would be easily explained, but an ever increasing frequency of HD should be predicted owing to new mutations. The HD mutation was found to be in linkage disequilibrium with 2 major haplotypes and also to be carried by a plethora of different haplotypes, indicating that new mutations do not constitute as rare an event as was previously thought (MacDonald *et al.*, 1992). Indeed, new HD mutations have been documented by CAG typing (Myers *et al.*, 1993; Goldberg *et al.*, 1993). Several authors have pointed out that haplotypes overrepresented in HD are found in the normal population with repeat numbers at the upper end of the normal range (Squitieri *et al.*, 1994; Almqvist *et al.*, 1995). Rubinsztein *et al.* (1994, 1995) have excluded a segregation distortion for HD and proposed a model based on the absence of natural selection against expanded alleles and the presence of a mutational bias causing the progressive expansion of the CAG sequence, hence a progressive increase in the incidence of the disease. In this context, a detailed evaluation of genetic fitness of hetero-zygotes for expanded alleles may add significantly to the understanding of factors influencing linkage disequilibrium and disease frequency.

Only sketchy data relevant to these issues are available for SCA1. Non-reduced fertility of patients in a large SCA1 family was reported by Haines *et al.*, 1984. Jodice *et al.* (1994) provided preliminary evidence on reduced fitness of large expansion carriers (> 54 repeats), consistent with the early onset of the disease in these patients. The mutation frequency is not known, but the evidence showing different haplotypes carrying the SCA1 mutation in Italy denotes its recurrent nature (Jodice *et al.*, 1993). Overall, linkage disequilibrium with extra- and intra-genic (absence of cryptic variations in the CAG stretch) markers has been documented (Jodice *et*

al. 1993; Chung *et al.*, 1993; Wakisaka *et al.*, 1995).

The present paper addresses the analysis of genetic fitness of both HD and SCA1 patients, in order to 1) provide data which might be crucial to the understanding of population genetics of these disorders; 2) assess whether the extent of expansion, correlated with phenotypic features, such as age at onset and severity of the disease, might also modify reproductive rates; 3) verify whether these disorders, both due to a CAG expansion, show a similar or a different selective pattern.

METHODS

Absolute fitness was calculated as usual, i.e. considering the total number of children surviving 15 years of age, born to people either older than 45 years of age or dead before that age. The criterion of survival beyond 15 years is unavoidable for the older generations, since in Italy at the beginning of this century, pre-reproductive mortality was still as high as 35%, about half being due to perinatal mortality (Terrenato *et al.*, 1979). The criterion is much less cogent for the younger generations, since in recent years pre-reproductive mortality has dropped to less than 5%, with perinatal mortality accounting for more than 90% of that figure. For these reasons, in some cases (as will be indicated) the total number of liveborn children has been considered instead.

HD Families

Information on living and deceased subjects was available for members of 83 HD families from the Latium regional area (about 5 million inhabitants). All records of patients admitted to neurological, psychiatric and geriatric hospitals in Latium between 1970 and 1990 and, when possible, earlier, were examined and patients with a diagnosis of chorea or choreic syndrome residing in and/or born in the area were selected. Clinical, genetic and sociodemographic data were obtained through inspection of life and clinical records, neurological examination of living

patients, direct interviews with living family members and contacts with the oldest member of each family and/or with the local physician. The family was considered as affected with HD if at least 1 member had an unequivocal diagnosis. Ninety-five percent of family founders were born before 1914 and 64% before 1890. Only 28 of them, however, were included in the analysis, since data were uncomplete for the others.

Overall, data for 319 patients and 328 healthy family members with a 50% prior risk of having inherited the disease were available. Both married and unmarried subjects were included, the proportion of the latter being comparable in the 2 groups (15·8% and 13·4% for patients and normal relatives, respectively). The male to female ratio was 0·99:1 for the patients and 1·02:1 for healthy subjects.

CAG repeat assays (Warner *et al.*, 1993) were performed on 250 subjects from 65 HD families, some of whom belonged to the above group of families and had agreed to donate blood for research, and the rest ascertained through the request of one member for presymptomatic HD testing. The geographical origins of the latter families were scattered throughout Italy.

In order to have an internal control for reproduction within each family, only CAG expansion carriers who had at least one relative homozygote for wildtype alleles in the same generation were considered. Thirty-one families fulfilled these criteria, with overall 154 subjects born before 1951, 110 of which belonged to 20 families from Latium. Eighty-one were carrying one expanded allele and 73 two normal alleles, the threshold between the normal and the expanded range being set at 36 CAG repeats (Novelletto *et al.*, 1994*b*). Eleven subjects of the 81 were asymptomatic expansion carriers, a number not significantly different from the 13 expected on the basis of subject age and of Italian HD life tables (Brambilla *et al.*, 1990). The male to female ratio was 0·88:1 for expansion carriers and 0·74:1 for normals. The smaller number of males reflects the higher percentage of women asking for predictive testing (ECHDCSG, 1993), and of those available for blood donation.

SCA1 families

Historical data starting from around 1800 for 4 very large SCA1 families, all originating from a small area of the Calabrian region of Southern Italy, were collected at the local parishes and from mayoral and noble family archives. Clinical information was obtained from direct examination (living subjects), clinical notes (deceased subjects in recent generations), phenotype description reported in Pisani *et al.* (1964), as well as from the diaries of one noble family. A systematic search for SCA1 families in the area (about 175,000 inhabitants) was completed through contacts with local doctors and inspection of hospital records. Yearly follow-up of the area over 5 years did not reveal any additional affected family. Overall, 77 patients and 96 normal relatives with 50% prior risk were available. The male to female ratio was 1·41:1 for the patients and 0·85:1 for normal subjects, possibly reflecting a differential male emigration according to SCA1 phenotype.

Subjects typed for CAG repeat polymorphism of the SCA1 gene on chromosome 6p23, following Orr *et al.* (1993), were drawn from a sample of 159 subjects from 10 Italian families, including the 4 Calabrian ones and additional six (from the Latium region) ascertained through index cases seen as in/out-patients at the Neurology Clinic of the University La Sapienza in Rome. Using the same criteria reported for HD, 64 subjects from 7 families were selected; 27 subjects carrying a CAG expansion (including 13 asymptomatics) and 37 relatives with normal alleles. Eight additional subjects had their CAG genotype derived from that of their offspring, assuming no back-mutation. Sixty-two of the overall 72 subjects were members of the Calabrian families.

RESULTS

HD

To assess the genetic fitness of HD patients, a first analysis was performed comparing the reproduction of the clinically affected members of HD families to that of their unaffected relatives with 50% risk, subdivided according to their year of birth, in 7 groups (1860–89, 1890–99, 1900–09, 1910–19, 1920–29, 1930–39, 1940–49). No significant differences between males and females were observed in each group (t-test) and data for the 2 sexes were pooled.

The mean number of children for the 2 groups and results of t-tests are reported in Fig. 1. They clearly indicate that, up to about 1920, patients had significantly more children than their normal relatives, with the exception of the decade 1900–1909.

These results appear to be little affected by the typical biases of this kind of data. Family founders are usually more likely to be ascertained if they have a large family size and this might increase the number of offspring of patients in older generations. In our sample of HD families, only a minority of founders were included in the analysis (see Methods section) and most of them were born before 1890. No such bias should be expected after this date. In addition, data were derived from a nearly complete ascertainment of families in the Latium region, and relatively few patients with no children or non-affected descendants might have been lost.

Full demographic data from successive Italian censuses are reported in Terrenato *et al.* (1979) for subjects born in the period 1871–1920. From these data the mean number of children surviving 15 years in the general population were derived and are reported in Fig. 1. They are very similar to those of the control group, indicating that the use of normal relatives as controls does not underestimate the genetic fitness of the general population (Reed, 1959), at least up to 1920.

Distributions of family size for the generations before 1920 are reported in Table 1. No significant differences were found between patients and their normal relatives by contingency chi^2 test (9·83 d.f. 9 and 13·13 d.f. 8, respectively for 1860–1899 and 1900–1919).

The class with no children deserves special consideration. In fact, HD families in which most of the patients had no offspring have little opportunity of being ascertained through a living
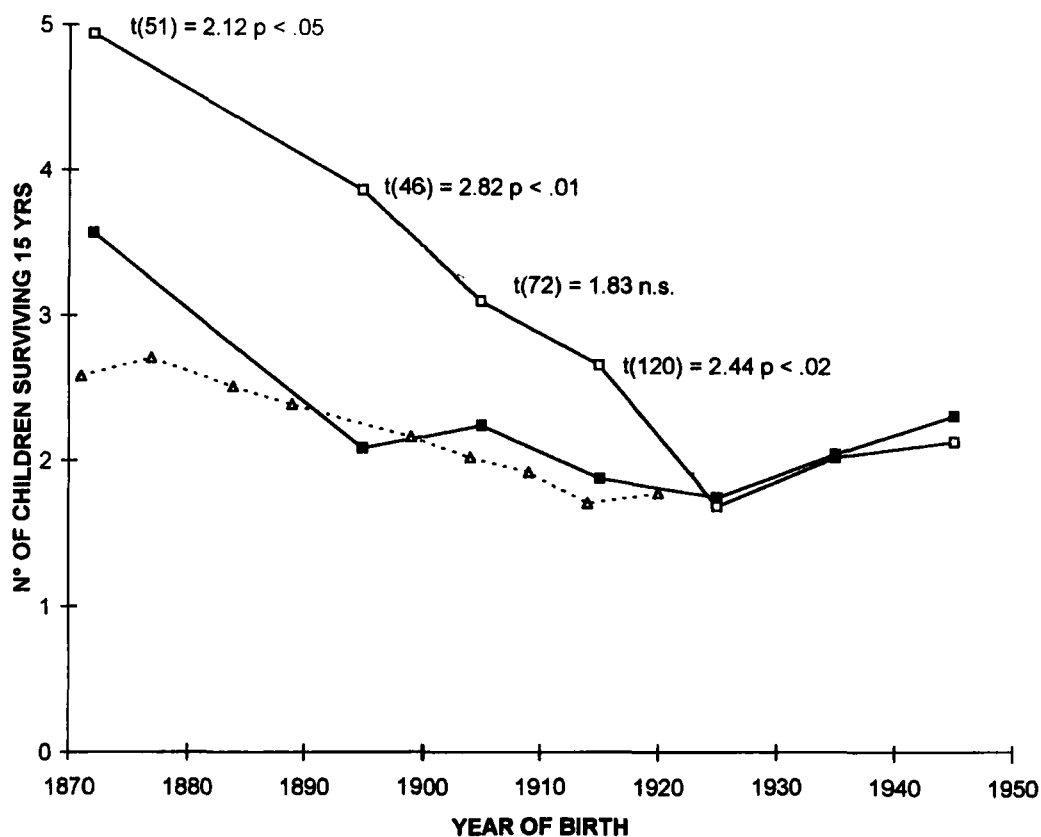
Fig. 1. Mean number of children surviving 15 years born to HD patients (open squares) and to their 50 % at risk relatives (filled squares), subdivided according to date of birth. Results of 2-tailed t tests between the 2 groups are reported. Data for the general population from Terrenato *et al.* (1979) are also reported (open triangle).

Table 1. *Family size distributions of affected and healthy HD family members born before 1920*

1860–1899		Family size									
		0	1	2	3	4	5	6	7	8	⩾ 9	Total
Patients	n	8	5	7	13	11	5	7	6	7	7	76
	%	10·5	6·6	9·2	17·1	14·5	6·6	9·2	7·9	9·2	9·2	100
Normals	n	3	2	5	5	7	1	1	1	0	0	25
	%	12·0	8·0	20·0	20·0	28·0	4·0	4·0	4·0	0	0	100
1900–1919											
Patients	n	15	12	22	18	18	8	5	3	1	1	103
	%	14·6	11·6	21·3	17·5	17·5	7·8	4·8	2·9	1·0	1·0	100
Normals	n	23	20	22	7	12	6	3	1	0	0	94
	%	24·4	21·3	23·4	7·4	12·8	6·4	3·2	1·1	0	0	100

affected descendent in recent generations. In principle, such a bias could not be excluded in our sample, although the rather substantial proportion of HD subjects in class 0 does not support a strong selection against them. However, in order to verify if even a mild selection could have affected our results, it was assumed that HD subjects without children would include: a) juvenile cases (onset below 20 years) under the very conservative assumption that all of them have no offspring, and b) cases with later onset but no children for the same reasons as the control group (non-marriage, infant mortality, etc.). The percentage of the former cases was assumed to be equal to 4 %, i.e. the frequency of patients with juvenile onset in our, as well as in other samples of HD families (Frontali *et al.*, 1990; Walker *et al.*, 1981) and that of the latter
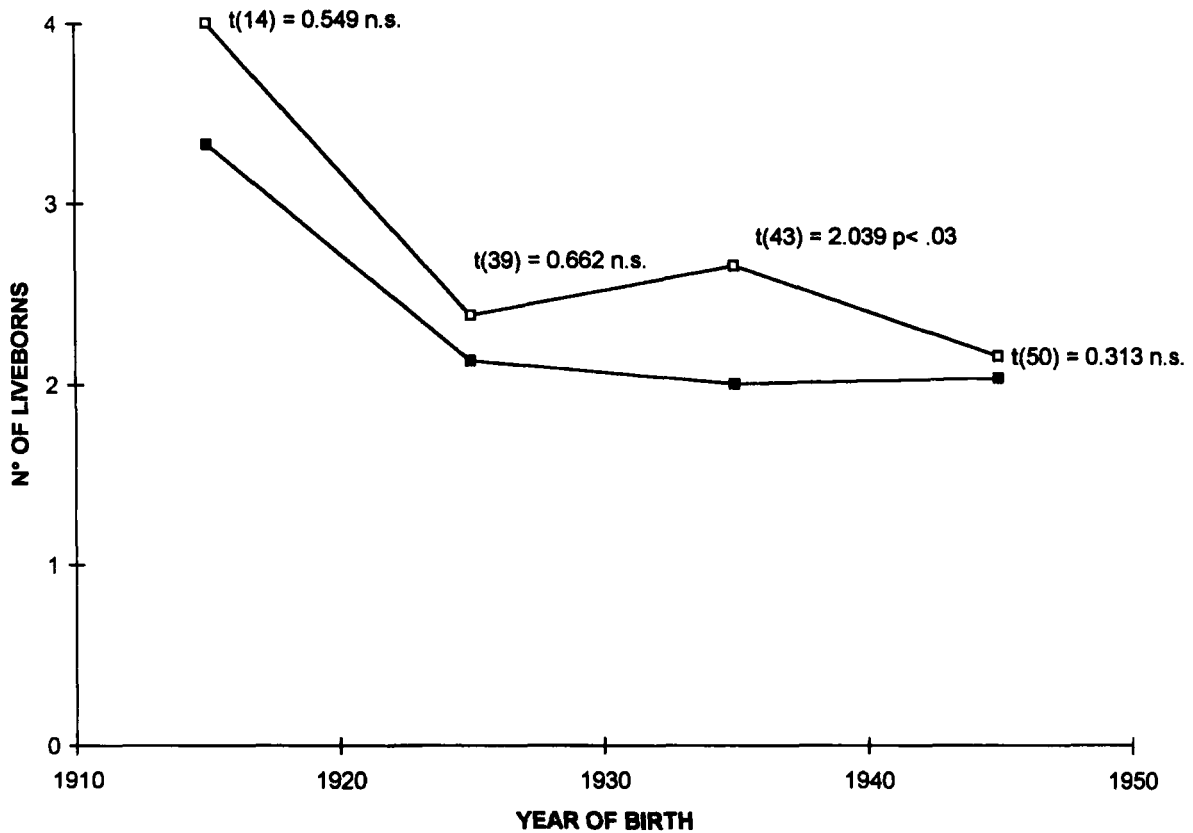
Fig. 2. Mean number of liveborn children born to heterozygotes for the HD mutation, independently of their clinical status (open squares) and to their relatives with a wildtype genotype (filled squares), subdivided according to their date of birth. Results of 2-tailed t test between the 2 groups are reported.

equal to the frequency in the control group, for which no bias should be present. Applying this correction, the mean number of children born to HD patients was recalculated, adding to class 0 as many subjects as needed to reach the expected rate. Under these conditions, the means are still 1·42 and 1·19 times greater than those of normal relatives, in the first and second time periods respectively (1800–1899: mean ± S.E. 2·92 ± 0·35, for the control group; mean ± S.E. 4·15 ± 0·33 [$n_{tot} = 81$, $n_0 = 13$] for patient group; 1900–1919: mean ± S.E. = 2·0 ± 0·18 for control group; mean ± S.E. 2·38 ± 0·20 [$n_{tot} = 122$, $n_0 = 34$] for patient group).

Among the possible explanations for similar mean numbers of children born to patients and to their normal relatives after 1920 (see also Discussion) is the admixture, in the control group, of homozygotes for wildtype alleles and of asymptomatic heterozygotes for the HD gene, the proportion of whom progressively increases

along generations due to the late onset of the disease. A relatively low rate of asymptomatic carrier is expected in older generations, being due only to the death of subjects before the onset of the disease (mean age at onset 42·73 ± 1·72, as reported by Frontali et al., 1990), while in more recent generations, the number of living subjects who have not yet manifested the disease is expected to be higher. If asymptomatic carriers have more children, their admixture with normals would tend to raise the mean number of children and to blur the difference with the affected group.

The assessment of the number of CAG repeats provides the opportunity to compare homozygotes for wildtype alleles and carriers of the HD mutation, independently of their clinical status. From a sample of 250 typed subjects, 154 fulfilled the criteria reported in the Methods section. Seventy-three were carriers of a CAG repeat expansion ranging from 37 to 49, and 81

Table 2. *Family size distributions of CAG typed HD subjects and controls in the period 1910–1939*

(CAG)n		Family size								
		0	1	2	3	4	5	6	⩾7	Total
37–49	n	3	7	17	14	7	4	2	1	55
	%	5·4	12·7	19·3	25·4	12·7	7·3	3·6	1·8	100·0
<35	n	2	10	19	10	4	2	0	0	47
	%	4·2	21·3	40·4	21·3	8·5	4·2	0	0	100·0

Table 3. *Reproduction of carriers of the HD mutation with different degrees of expansion and of their normal relatives born in the period 1951–1972*

(CAG)n	No. subjects	Mean age	Phenotype	Mean No. of children $\pm$ S.E.
<35	47	35·0±5·9	Normal	1·23±0·16
37–49	36	34·9±5·8	70% normal 30% affected	0·83±0·18
⩾50	6	30·5±6·2	Deceased or end stage of HD	0·67±0·42

were normals (17 to 34 repeats). No carriers of larger expansions were present in this sample nor carriers of intermediate alleles (35 or 36 repeats). The total number of children was considered, instead of the number of children surviving 15 years, since subjects were born in a more recent period of time (1910–1950). No significant difference between males and females was obtained.

The mean number of children as a function of parental decade of birth in the 2 groups is reported in Fig. 2. The carriers of expanded alleles have larger families than the control group, in all the time intervals considered. However, separate t-tests provided significant results only for the period 1930–39. An overall comparison in the period 1910–1939 showed a difference between the 2 groups at the significance threshold (mean±S.E. 2·78±0·23 for expansion carriers, 2·21±1·14 for carriers of normal alleles, t = 1·93, d.f. 100, p=0·05, 2-tailed test).

The family size distributions for the 2 groups in the period 1910–1939 is reported in Table 2. As for data reported in Table 1, chi² between the 2 distributions is not significant (5·398 d.f. 6). Nevertheless, a bias against carriers of expanded alleles without children cannot be excluded, although in this case the proportion of subjects in

class 0 (mostly due to the first 2 decades 1910–1919 and 1920–1929) is higher among heterozygotes than normals. However, the mean number of children born to expansion carriers is still 1·21 times higher than that of normals when the frequency of class 0 among heterozygotes is brought to that of normals plus an additional 4% due to juvenile cases (mean±S.E 2·68±0·23, [n_{tot} = 57, n_0 = 5] for the group of expansion carriers; 2·21±0·25 for the group with wild genotypes).

With the above data, only the difference between carriers of the CAG expansion in the low-medium range (37–49) and normals could be assessed. Subjects born before 1951 and carrying large expansions, and, hence, with earlier onset and shorter disease duration, are unlikely to be still alive in the present decade (Novelletto *et al.*, 1994*a*).

In order to estimate approximately the rate of reproduction of heterzygotes for large expansions, data from 89 typed subjects born from 1951 to 1972 were analyzed. The sample was subdivided into 3 groups, according to the larger CAG allele carried by each subject: <35, 37–49, and 50 or more repeats. The mean number of children for the 3 groups, reported in Table 3, is not significantly different at one-way ANOVA (F = 1·767, d.f. 2/88). It should be considered,
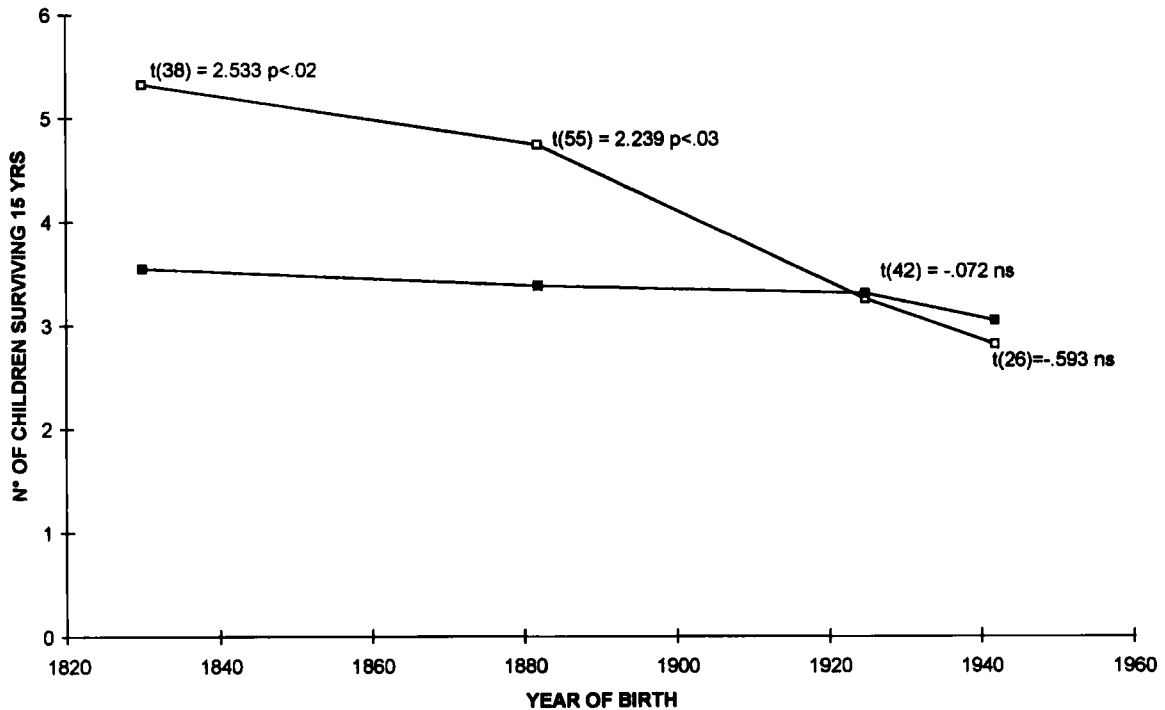
Fig. 3. Mean number of children surviving 15 years born to SCA1 patients (open squares) and to their 50 % at risk relatives (filled squares), subdivided according to date of birth. Results of 2-tailed t tests between the 2 groups are reported.

however, that the 6 subjects with 50 CAG repeats or more had completed reproduction, being either deceased or in the last stages of HD, while in the other 2 groups the number of children is bound to increase before the end of the reproductive period.

SCA1

The mean number of children surviving 15 years born to SCA1 patients and to their 50 % at-risk relatives is reported in Fig. 3. The two groups were subdivided according to the parents' date of birth in 4 time intervals (1790–1869, 1870–1914, 1915–1934 and 1935–1950). Data are similar to those found for HD, SCA1 patients having significantly more children than the control group in the first 2 time intervals, i.e. up to about 1915. The mean number of children of the normal relatives was higher than that found in the Italian general population (Fig. 1), owing to the higher reproductive rate in Southern Italian regions (Ulizzi et al., 1979). The results should be little affected by the inclusion of family

founders. In fact, all subjects belonged to 4 large Calabrian families originating in a 10 km^2 area, as a consequence of a founder effect. All families carry the same disease haplotype, not present in the Italian population (Jodice et al., 1993). Although a common ancestor could not be reconstructed from historical records, and some family branches might have been lost, the inclusion of the 4 family founders born before 1870 might have only slightly biased the first period of time considered (1790–1869).

The distributions of family size for the period 1790–1914, reported in Table 4, are significantly different (chi^2 16·26 d.f. 8 p < 0·05). An ascertainment bias against patients with no offspring should be conservatively hypothesized in this sample. Assuming that patients in class 0 had similar rates to those of control group and an additional 5 %, due to juvenile cases (Jodice et al., 1994) their mean number of children is still 1·30 times greater than normals (mean ± S.E. 4·47 ± 0·15, [n_{tot} = 47, n_0 = 5] for SCA1 subjects; 3·45 ± 0·29 for normal relatives).

The total number of children born to subjects

Table 4. *Family size distributions of SCA1 patients and their normal relatives in the period 1790–1914*

(CAG)n		0	1	2	3	4	5	6	7	8	9	10	Total
						Family size							
47–54	n	0	1	5	8	5	6	6	4	3	3	1	42
	%	0	2·4	11·9	19·1	11·9	14·3	14·3	9·5	7·1	7·1	2·4	100·0
< 38	n	3	10	9	6	9	7	5	5	1	0	0	55
	%	5·4	18·2	16·4	10·9	16·4	12·7	9·1	9·1	1·8	0	0	100·0

Table 5. *Comparison between total number of children born to carriers of expanded SCA1 alleles and to normal relatives as a function of date of birth in the period 1890–1950*

	Number of repeats						
	< 38			47–54			
Date of birth	Mean	S.E.	No.	Mean	S.E	No.	t-test
1890–1934	2·58	0·43	19	4·55	0·75	11	t = 2·453 p < 0·03
1935–1950	2·55	0·28	22	2·00	0·29	20	t = 1·320 n.s.

typed for CAG repeats is reported in Table 5. Thirty-one subjects were carriers of one expanded allele ranging from 47 to 54 triplet repeats and 41 carried two normal alleles (37 being the largest allele). No carriers of intermediate alleles (38 to 46 repeats) or of very large expansions (> 54 repeats) were present. Due to its relatively small size, the sample was subdivided in only 2 time intervals for date of birth (1890–1934 and 1935–1950). A significant increase in the number of children born to expansion carriers was found for the first of the 2 intervals. When correcting for possible selection against patients with no offspring the mean number of children born to heterozygotes (mean ± S.E. 3·85 ± 0·58 n_{tot} = 13 n_0 = 2) is still 1·49 times greater than that of normals.

In addition, a French sample of 113 subjects from 12 families genotyped for the CAG alleles was screened according to the same criteria used for the Italian samples. Thirty subjects were selected, of whom 19 were carriers of one expanded allele, ranging from 42 to 54 repeats and 11 homozygotes for wildtype alleles with 34 repeats or less. No intermediate alleles (35 to 41 repeats) were found. The sample was too small to perform analyses in different time intervals. Since 83 % of the sample was born between 1935 and 1950, only this interval was taken into consideration. The mean number of children was

2·40 ± 1·55 and 1·63 ± 1·51, for expansion carriers and normals, respectively. Although the difference is not significant at t-test, the mean for expansion carriers is 1·47 times greater than that of the control group.

An analysis of carriers of larger expansions was possible in the Italian sample for only 64 subjects born in the period 1951–1972. The sample was subdivided into 3 groups, according to the largest CAG allele carried by each subject (< 38 n = 36, 47–54 n = 20, > 54 n = 8). Results were similar to those obtained for HD (mean ± S.D. 0·81 ± 0·18, 0·90 ± 0·23, 0·88 ± 0·52 for normals, low/medium, and large expansion group, respectively). No significant difference was present among the 3 groups at one-way ANOVA (F = 0·565 d.f. 2/61), but the reproduction of the 8 subjects with more than 54 repeats should be considered as complete, given that they are either deceased or at the end stage of the disease, while the other 2 groups still have time to increase their family size.

DISCUSSION

Genetic fitness was analyzed in partially overlapping samples: on one hand, clinically defined HD and SCA1 patients from families ascertained in definite geographical areas, and on the other, molecularly typed carriers of HD or

SCA1 mutations partly from the above and partly from additional families. In both cases normal relatives, respectively clinically and molecularly assessed, were used as controls.

In the first set of data, both HD and SCA1 patients born before 1915–20 appear to have an increased genetic fitness, while those born after that date have a rate of reproduction similar to that of the control group. Two factors might be responsible for the vanishing difference in more recent generations. First, the demographic transition, i.e. the progressive reduction of the mean number of children, appears to reach a plateau in Italy around the Second World War, when subjects born in the period 1915–20 are reproducing. As the age at first pregnancy and the time elapsed between successive offspring progressively increase, less time is left for patients to reproduce before the onset of severe physical or mental impairment. Second, the control group is an admixture of carriers of normal genotypes and an increasing number of asymptomatic carriers of CAG expansions along generations. As already mentioned in the Results section, this could blur the difference with the affected group in recent generations.

The second set of data shows that when carriers of HD and SCA1 expansions in the low/medium range are compared to relatives with a normal genotype, independently of their clinical status, they show an increase in the number of children up to more recent generations. The possibility of detecting a difference in reproductive rate for a longer period of time than in the first set of data is most likely due to the exclusion of asymptomatic heterozygotes from the control group and their pooling with symptomatic ones. However, this holds true only for carriers of alleles in the low/medium range of expansion (37–49 for HD and 47–54 for SCA1), whose fitness could be correctly assessed. Heterozygotes for larger expansions born before 1951 are not likely to be still alive and available for blood sampling in the present days, due to their early onset and short disease duration: they were found only among subjects younger than 45 years.

Carriers of larger expansions are expected to have very few children, if any, on the basis of the reported inverse correlation between the number of CAG repeats and the age at onset. In the Italian series, carriers of 50 or more repeats have an HD onset between 4 and 35 years (Novelletto et al., 1994a and additional unpublished data) and that of SCA1 is below 30 years (Jodice et al., 1994). Indirect support of this expectation is provided by data from subjects born in the period 1951–1972 (Table 3), showing that carriers of large expansions, all without the opportunity for further reproduction, due to death or severe invalidity, have the same number of children as carriers of smaller expansions (most of whom are still asymptomatic) who are bound to increase their family size before their reproductive life is completed.

These results suggest a model that takes into account the differential fitness according to the CAG repeat length for both HD and SCA1 heterozygotes. Under this model, most of the alleles with the largest expansions are lost at each generation and are replaced by those in the low/medium expansion range, as a consequence of instability. The pool of the latter is likely to be maintained or even expanded until relatively recently, not only by new mutations but also by the increased fitness of their carriers. While the first source generates expansions likely to be found on random haplotypes, the second maintains the disease mutation on ancestral chromosomes. The maintenance of the expanded allele frequency therefore would depend not only on the balance between selection and new mutations but even more on the differential fitness of expansion carriers, associated with a continuous shift from the pool of low/medium expansion alleles toward that of large expansions.

Such a hypothesis might explain previous findings of a relative genetic fitness of HD patients equal or greater than one, in spite of the fact that juvenile cases, varying from 1 to 12% of HD patients (Harper, 1991), are expected to have a decreased reproduction. In addition, it accounts for the heterogeneous set of HD chromosomes reported in the literature, showing

that few haplotypes in linkage disequilibrium with the mutation coexist with many different haplotypes most likely carrying recent mutations (MacDonald *et al.*, 1992; Squitieri *et al.*, 1994)). According to the present model, haplotypes in linkage disequilibrium with the HD expansion are observed mostly as a result of common ancestry, although recurrent *de novo* expansions on a subset of normal chromosomes with 'predisposing' haplotypes may contribute to this pool. The present results also suggest that reproduction of HD heterozygotes tends to be restricted by recent sociodemographic constraints more than that of their normal relatives. Assuming a steady rate of new mutations and instability, the magnitude of this phenomenon would predict a reduction of HD frequency in future generations.

A similar situation emerges for SCA1, although fewer population genetics data are available in the literature. Non-reduced fertility rates of SCA1 patients up to 1919 is reported by Haines *et al.* (1984), with juvenile cases accounting for a substantial proportion of the patients in the sample. In addition, linkage disequilibrium with extragenic markers has been documented in Italy in a small set of families, as well as the presence of different haplotypes not related to each other (Jodice *et al.*, 1993)

The evidence in favour of a differential fitness found in HD and SCA1 patients poses the question of whether the same might hold true for other diseases due to CAG expansion, such as SCA3, KD and DRPLA. In any case, the present data suggest a population genetics model which is completely different from those proposed for diseases such as DM or FRAXA (Mandel *et al.* 1994).

At present, one can only speculate about causes of a higher fitness in carriers of alleles in the low-medium expansion range. These include possible effects of CAG expansions on reproductive physiolology or on sexual drive and behaviour, and psychological mechanisms compensating for the stress of initial subclinical signs of the disease through family-raising. In any case, the change in reproduction associated with

CAG repeat expansion appears to be independent of the clinical manifestations of the disease and of subjects' mental state. In fact, the group of expansion carriers showing a fitness higher than normal included a substantial proportion of asymptomatic subjects. Moreover, results were similar in HD, a disorder with progressive dementia, and SCA1, whose clinical picture rarely includes mental decay.

REFERENCES

ALMQVIST E., SPENCE N., NICHOL K., ANDREW S. E., VESA J., PELTONEN L., ANVRET M., GOTO J., KANAZAWA I., GOLDBERG Y. P. & HAYDEN M. R. (1995). Ancestral differences in the distribution of the Δ2642 glutamic acid polymorphism is associated with varying CAG repeat length on normal chromosomes: insights into the genetic evolution of Huntington disease. *Hum. Mol. Genet.* **4**, 207–214.

BRAMBILLA C., FRONTALI M., MALASPINA P. & ROSSI C. (1990). On the estimation of the age at onset distribution in Huntington's chorea using EM algorithm. *Ann. Hum. Genet.* **54**, 225–233.

CAMPUZANO V., MONTERMINI L., MOLTO' M. D., PIANESE L., COSSEE M., CAVALCANTI F., MONROS E., RODIUS F., DUCLOS F., MONTICELLI A., ZATA F., CANIZARES J., KOUTNIKOVA H., BIDICHANDANI S.I., GELLERA C., BRICE A., TROUOILLAS P., DE MICHELE G., FILLA A., DE FRUTOS R., PALAU F., PATEL P. I., DI DONATO S., MANDEL J-L., COCOZZA S., KOENIG M. & PANDOLFO M. (1996) Friedrich's Ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**, 1423–1427.

CAREY N., JOHNSON K., NOKELAINEN P., PELTONEN L., SAVONTAUS M. L., JUVONEN V., ANVRET M., GRANDELL U., CHOTAI K., ROBERTSON E., MIDDLETON-PRICE H., MALCOM S. (1995). Meiotic drive and myotonic dystrophy. *Nature Genet.* **10**, 132–133.

CHUNG M., RANUM L. P. WP., DUVICK L. A., SERVADIO A., ZOGHBI H. Y. & ORR H. T. (1993). Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type 1. *Nature Genet.* **5**, 254–258.

EICHLER E. E., HOLDEN J. J. A., POPOVICH B. W., REISS A. L., SNOW K., THIBODEAU S. N., RICHARDS C. S., WARD P. A. & NELSON D. L. (1994). Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature Genet.* **8**, 88–94.

EUROPEAN COMMUNITY HUNTINGTON'S DISEASE COLLABORATIVE STUDY GROUP. (1993). Ethical and social issues in presymptomatic testing for Huntington's disease: a European Community collaborative study. *J. Med. Genet.* **30**, 1028–1035.

FRONTALI M., MALASPINA P., ROSSI C., JACOPINI A. G., VIVONA G., PERGOLA M. S., PALENA A. & NOVELLETTO

A. (1990). Epidemiological and linkage studies on Huntington's disease in Italy. *Hum. Genet.* **85**, 165–170.

GOLDBERG Y. P., KREMER B., ANDREW S. E., THEILMANN J., GRAHAM R. K., SQUITIERI F., TELENIUS H., ADAM S., SAJOO A., STARR E., HEILBERG A., WOLFF G. & HAYDEN M. R. (1993). Molecular analysis of new nutations for Huntington's disease: intermediate alleles and sex of origin effects. *Nature Genet.* **5**, 174–179.

HAINES J. L., SCHUT L. J., WEITKAMP L. R., THAYER M. & ANDERSON V. E. (1984). Spinocerebellar ataxia in a large kindred: age at onset, reproduction and genetic linkage studies. *Neurology* **34**, 1542–1548.

HARPER P. S. (1991). *Huntington's Disease.* London: Saunders, pp. 294–298.

HUNTINGTON 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.

IMBERT G., KRETZ C., JOHNSON K. & MANDEL J-L. (1993). Origin of expansion mutation in myotonic dystrophy. *Nature Genet.* **4**, 72–76.

JODICE C., FRONTALI M., PERSICHETTI F., NOVELLETTO A., PANDOLFO M., SPADARO M., GIUNTI P., SCHINAIA G., LULLI P., MALASPINA P., PLASMATI R., TOLA R., ANTONELLI A., DI DONATO S., MOROCUTTI C., WEISSENBACH J., CANN H. M. & TERRENATO L. (1993). The gene for spinal cerebellar ataxia 1 (SCA1) is flanked by two closely linked highly polymorphic microsatellite loci. *Hum. Mol. Genet.* **2**, 1383–1387.

JODICE C., MALASPINA P., PERSICHETTI F., NOVELLETTO A., SPADARO M., GIUNTI P., MOROCUTTI C., TERRENATO L., HARDING A. E. & FRONTALI M. (1994). Effect of trinucleotide repeat length and parental sex on phenotypic variation in Spinocerebellar Ataxia 1. *Am. J. Hum. Genet.* **54**, 959–965.

KAWAGUCHI Y., OKAMOTO T., TANIWAKI M., AIZAWA M., INOUE M., KATAYAMA S., KAWAKAMI H., NAKAMURA S., NISHIMURA M., AKIGUCHI I., KIMURA J., NARUMIYA S. & KAKIZUKA A (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32·1. *Nature Genet.* **8**, 221–228.

KOIDE R., IKEUCHI T., ONODERA O., TANAKA H., IGARASHI S., TAKAHASI H., KONDO R., ISHIKAWA A., HAYASHI T., SAITO M., TOMODA A., MILKA T., NAITO H., IKUTA F. & TSIYI S. (1994). Unstable expansion of a CAG repeat in hereditary dentatorubro-pallidoluysian atrophy (DRPLA). *Nature Genet.* **6**, 9–13.

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.

LA SPADA A. R., WILSON E. M., LUBAHN D. B., HARDING A. E. & FISHBECK H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**, 77–79.

LA SPADA A. R., PAULSON H. L. & FISCHBECK K. H. (1994). Trinucleotide repeat expansion in neurological disease. *Ann. Neurol.* **36**, 814–822.

MACDONALD M. E., NOVELLETTO A., LIN C., TAGLE D., BARNES G., BATES G., TAYLOR S., ALLITTO B., ALTHERR M., MYERS R., LEHARACH H., COLLINS F. S., WASMUTH J. J., FRONTALI M. & GUSELLA J. F. (1992).

The Huntington's Disease candidate region exhibits many different haplotypes. *Nature Genet.* **1**, 99–103.

MANDEL J-L. (1994). Trinucleotide diseases on the rise. *Nature Genet.* **7**, 453–455.

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., SCHWARR M., WOLFF G., BIRD E. D., VONSATTEL J. P. G. & GUSELLA J. F. (1993). *De novo* expansion of a (CAG)n repeat in sporadic Huntington's disease. *Nature Genet.* **5**, 168–173.

NOVELLETTO A., PERSICHETTI F., SABBADINI G., MANDICH P., BELLONE E., AJMAR F., PERGOLA M., DEL SENNO L., MACDONALD M. E., GUSELLA J. F. & FRONTALI M. (1994a). Analysis of the trinucleotide repeat expansion in Italian families affected with Huntington disease. *Hum. Mol. Genet.* **3**, 93–98.

NOVELLETTO A., PERSICHETTI F., SABBADINI G., MANDICH P., BELLONE E., AJMAR F., SQUITIERI F., CAMPANELLA G., BOZZA A., MACDONALD M. E., GUSELLA J. F. & FRONTALI M. (1994b). Polymorphism analysis of the huntingtin gene in Italian families affected with Huntington disease. *Hum. Mol. Genet.* **3**, 1129–1132.

ORR H. T., CHUNG M., BANFI S., KWIATKOWSKI T. J. JR., SERVADIO A., BEAUDET A. L., MCCALL A. E., DUVICK L. A., RANUM L. P. W. & ZOGHBI H. Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genet.* **4**, 221–226.

OUDET C., MORNET E., SERRE J. L., THOMAS F., LENTES-ZENGERLING S., KRETZ C., DELUCHAT C., TEJADA I., BOUE' J., BOUE' A. & MANDEL J-L. (1993). Linkage disequilibrium between the Fragile X mutation and two closely linked CA repeats suggest that fragile X chromosomes are derived from a small number of founder chromosomes. *Am. J. Hum. Genet.* **52**, 297–304.

PISANI D., PAPALIA A., BARBIERI N. F. & ROMBOLA F. (1964). Eredoatassia Spino-Cerebellare (Studio genetico e clinico). *Neuropsichiatria* **30**, 237–310.

REED T. E. (1959). The definition of relative fitness of individuals with specific genetic traits. *Am. J. Hum. Genet.* **2**, 37–155.

RICHARDS R. I., HOLMAN K., FRIEND K., KREMER E., HILLEN D., STAPLES A., BROWN W. T., GOONEWARDENA P., TARLETON J., SCHWARTZ C. & SUTHERLAND G. (1992). Fragile X syndrome: evidence of a founder chromosome. *Nature Genet.* **1**, 257–260.

RUBINSZTEIN D. C., AMOS W., LEGGO J., GOODBURN S., RAMESAR R. S., OLD J., BONTROP R., MCMAHON R., BARTON D. E. & FERGUSON SMITH D. (1994). Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. *Nature Genet.* **7**, 525–530.

RUBINSZTEIN D. C., AMOS W., LEGGO J., GOODBURN S., JAIN S., LI S-H., MARGOLIS R. L., ROSS C. A. & FERGUSON-SMITH M. A. (1995). Microsatellite evolution – evidence for directionality and variation in rate between species. *Nature Genet.* **10**, 337–343.

SQUITIERI F., ANDREW S. E., GOLDBERG Y. P., KREMER B., SPENCE N., ZEISLER J., NICHOL K., THEILMANN J., GREENBERG J., GOTO J., KANAZAWA I., VESA J., PELTONEN L., ALMQVIST E., ANVRET M., TELENIUS H., LIN B., NAPOLITANO G., MORGAN K. & HAYDEN M. R. (1994). DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG

expansion and reasons for geographic variations of prevalence. *Hum. Mol. Genet.* **3**, 2103–2114.

SUTHERLAND G. R. & RICHARDS R. I. (1995). Simple tandem DNA repeats and human genetic disease. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3636–3641.

TERRENATO L., ULIZZI L. & SAN MARTINI A. (1979). The effects of demographic transition on the opportunity for selection: changes during the last century in Italy. *Ann. Hum. Genet.* **42**, 391–399.

ULIZZI L., SAN MARTINI A. & TERRENATO L. (1979). Changes of selection opportunities with a changing environment: regional heterogeneity in Italy. *Am. Hum. Genet.* **43**, 135–139.

WAKISAKA A., SASAKI H., TAKADA A., FUKAZAWA T., SUZUKI Y., HAMADA T., IWABUCHI K., TASHIRO K. & YOSHIKI T. (1995). Spinocerebellar ataxia 1 (SCA1) in the Japanese in Hokkaido may derive from a single common ancestry. *J. Med. Genet.* **32**, 590–592.

WALKER D. A., HARPER P. S., WELLS C. E. C., TYLER A., DAVIES K. & NEWCOMBE R. G. (1981) Huntington's Corea in South Wales. A genetic and epidemiologic study. *Clin. Genet.* **19**, 213–221.

WARNER J. P., BARRON L. H. & BROCK D. J. H. (1993). A new polymerase chain reaction (PCR) assay for the trinucleotide repeat that is unstable and expanded in Huntington's disease. *Mol. Cell. Probes* **7**, 235–239.