Original Paper

Human Heredity

Hum Hered 2005;60:156–163
DOI: 10.1159/000090035

Received: July 21, 2005
Accepted after revision: September 29, 2005
Published online: December 6, 2005

Population Genetics of Wild-Type CAG Repeats in the *Machado-Joseph Disease* Gene in Portugal

M. Lima[a] M.C. Costa[b,c] R. Montiel[a] A. Ferro[c] C. Santos[d,g] C. Silva[a]
C. Bettencourt[a] A. Sousa[c,e] J. Sequeiros[c,e] P. Coutinho[f] P. Maciel[b,c,e]

[a]Center of Research in Natural Resources (CIRN), University of the Azores, Ponta Delgada, [b]Life and Health Sciences Research Institute, Health Sciences School, University of Minho, Braga, [c]UnIGENe, Institute of Molecular and Cellular Biology (IBMC), University of Porto, Porto, [d]Department of Anthropology, University of Coimbra, Coimbra, [e]Department of Population Studies, Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, and [f]Neurology, Hospital de S. Sebastião, Santa Maria da Feira, Portugal; [g]Department of BABVE, Autonomous University of Barcelona, Barcelona, Spain

Key Words

Machado-Joseph disease · SCA3 · Expansion disorder · Triplet repeats · Population study · Mutational bias · Portugal

Abstract

Objective: To gain insights on the molecular mechanisms of mutation that led to the emergence of expanded alleles in the *MJD* gene, by studying the behavior of wild-type alleles and testing the association of its distribution with the representation of the disease. ***Methods:*** The number of CAG motifs in the *MJD* gene was determined in a representative sample of 1000 unrelated individuals. Associations between the repeat size and the epidemiological representation of MJD were tested. ***Results:*** The allelic profile of the total sample was in the normal range (13–41 repeats), with mode $(CAG)_{23}$. No intermediate alleles were present. Allelic size distribution showed a negative skew. The correlation between the epidemiological representation of MJD in each district and the frequency of small, medium and large normal alleles was not significant. Further correlations performed grouping the districts also failed to produce significant results. ***Conclusions:*** The absence of association between the size of the repeats and the representation of MJD demonstrates that prevalence is not an indirect reflection of the frequency of large normal alleles. Globally the results obtained are in accordance with a model that postulates the occurrence of a few mutations on the basis of most of the MJD cases worldwide.

Introduction

Short-tandem repeats (STRs) are stretches of repetitive DNA composed of repeat units ranging from 2 to 6 bp, abundantly found in eukaryotic genomes, namely in the human genome. The trinucleotide class of STRs is particularly interesting, since in a considerable number of cases pathologically expanded trinucleotide repeats cause hereditary disorders [1–9]. Among such disorders is Machado-Joseph disease (MJD) [MIM109150] [10–11]. MJD is an autosomal dominant neurodegenerative disorder whose gene has been assigned to the long arm of chromosome 14 [10]. Its manifestations include cerebellar ataxia and progressive external ophthalmoplegia, associated in variable degrees with pyramidal signs, extrapyramidal signs, amyotrophy and peripheral neuropathy

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0001–5652/05/0603–0156$22.00/0

Accessible online at:
www.karger.com/hhe

Manuela Lima, PhD
Center of Research in Natural Resources (CIRN), Department of Biology
University of the Azores, Rua Mãe de Deus, 58
PT–9501-801 Ponta Delgada (Portugal)
Tel. +351 96 650 118, Fax +351 96 650 100, E-Mail mlima@notes.uac.pt

[12]. MJD is caused by an expanded CAG repeat within the coding sequence of the *MJD* gene [11], resulting in the lengthening of a polyglutamine tract in the corresponding protein. Overall, 108 extended MJD families have been described in Portugal, distributed in the Azores Islands and in Mainland Portugal [13; Paula Coutinho, personal communication]. Clusters of the disease have been described, namely in the Island of Flores (Azores), where the prevalence of MJD reaches values as high as 1 in 103 [14–15]. In the MJD locus healthy individuals have 12 to 44 CAG units, whereas in MJD patients the repeat copy number in the expanded allele is between 61 and 87 [16]. With the exception of a healthy individual described by Maciel et al. [16] whose family originated from the Tagus valley, there are no reports of intermediate alleles in normal individuals. The extent to which this could be a consequence of the limited number of population studies is unknown. The genotyping of a random sample of 320 individuals from the general population of the Tagus valley failed, however, to detect individuals carrying intermediate alleles [16]. Access to data from the reference laboratory for MJD in Portugal provides an additional confirmation of the absence of individuals carrying intermediate alleles [Jorge Sequeiros, personal communication]. On the other hand, although alleles smaller than 61 repeats are not commonly found in the patient population, a few cases of alleles of intermediate size have nevertheless been identified. Takiyama et al. [17] and van Schaik et al. [18] described two patients with a number of repeats in the normal range on one chromosome and with alleles of 56 and 54 repeats on the other, respectively. Furthermore, van Alfen et al. [19] described four members of a Dutch family that had alleles of intermediate length (53 and 54 repeats) and exhibited an unusual MJD phenotype.

The biological basis of repeat expansion is believed to be the instability of the repeats during meiotic and mitotic cell divisions [20–22]. In expansion disorders a tendency for the expanded alleles to further increase in size in successive generations has been reported, particularly in male transmissions [23]. In MJD, few contractions have been described for the expanded trinucleotide repeats, and a bias in favor of expansions seems to exist [24], although to a smaller degree than what is seen in Huntington disease (HD) [25].

Once expanded above a certain threshold the trinucleotide repeat alleles are implicated in the genesis of diseases and thus should theoretically impose a reduction in Darwinian fitness of their carriers that should lead, over an evolutionary timescale, to the self-extinction of expanded alleles. In accordance with this assumption, it is necessary to explain how the expanded alleles are maintained in human populations throughout time. It was postulated that a mutational bias in favor of expansions existed in trinucleotide repeat loci, suggesting that the large majority of new mutations at these loci originated from the upper end of the normal allele distribution, that would provide a reservoir from which expanded alleles were generated [25, 26]. Support put forward for this model came from the reported association between the prevalence of trinucleotide repeat diseases in specific populations and the frequency in these disease loci of alleles with longer repeats [21, 27–30]. More recently, however, several authors have demonstrated the heterogeneity in the mutational mechanisms underlying the different repeat loci and recognized that CAG repetitive regions with similar characteristics and pathogenicity show very different dynamic patterns [31, 32]. An independent analysis of every gene is thus mandatory; insights derived from the possible association between the epidemiological representation of expansion disorders and patterns of allele distribution in normal populations should also contribute to further define the singularity of each locus. The ability to correlate the prevalence of expansion disorders with the frequency of large normal alleles in the respective loci directly depends on the accuracy of the epidemiological profile of these disorders. The Portuguese population is extensively characterized at the epidemiological level, resulting from a systematic nationwide survey of MJD that has been going since 1993 [33]. Furthermore, extensive genealogical information concerning the affected families is available [13, 34, 35]. Given the accurate knowledge of the epidemiology of MJD, the Portuguese population provides an adequate background to test several aspects related with the relation between the behavior of the $(CAG)_n$ tract of the *MJD* gene in the normal population and the representation of the disease.

With the purpose of understanding the dynamics of the MJD locus, and of gaining insights into the molecular mechanisms of mutation that lead to the emergence of expanded alleles in the *MJD* gene, we studied the polymorphism of wild-type MJD alleles in a large representative sample of the Portuguese population. The presence of alleles of intermediate size was ascertained and the existence of a geographic sub-structuring of the MJD alleles was investigated. Finally, the relationship between the frequency of different sizes of the $(CAG)_n$ tract and the epidemiological representation of the disease was analyzed.

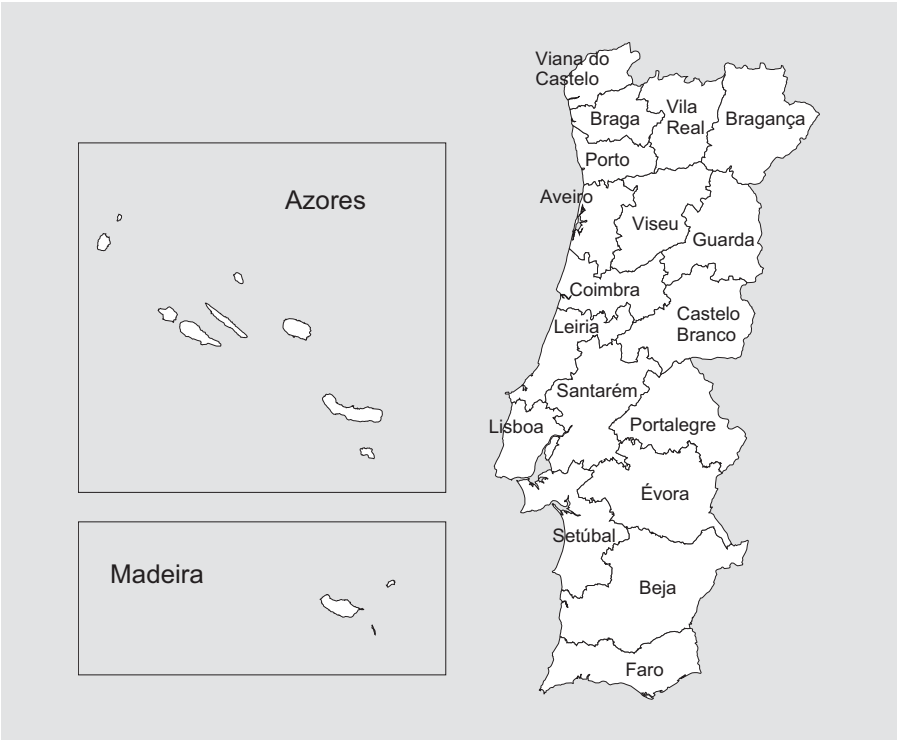

Fig. 1. Geographic delimitation of the 20 Portuguese districts. Total sample of 1,000 individuals included 50 samples per district (map source: http://portugal.veraki.pt/distritos/Fpage.php).

Material and Methods

Molecular Characterization of the CAG Repetitive Tract

For screening *MJD* $(CAG)_n$ sizes in normal chromosomes, anonymous samples were chosen at random from a Medical Genetics Center, where Guthrie cards are routinely used to screen newborns for genetic disorders (such as phenylketonuria and congenital hypothyroidism). The study was previously approved by the Ethics Commission of this center. In an attempt to obtain a control population which would be representative of the whole country, 1,000 anonymous and randomly selected samples corresponding to individuals born in each of the 20 districts in Portugal (50 individuals per district) were retained (fig. 1). DNA from the peripheral blood was extracted from the filter paper blots using Chelex in a final concentration of 0.8%. Amplification of the CAG repeat-containing fragment of the *MJD* gene was performed by PCR, using previously described conditions [11]. The size of the PCR products was determined by denaturing 6% polyacrylamide gel electrophoresis, in parallel with a M13 sequence ladder, and visualized by autoradiography.

Statistical Analyses

Analyses were initially performed by global sample and by district. Districts were then grouped according to: (1) geographic localization, and (2) epidemiological representation of MJD. Geographic localization allowed five groups to be defined: North (districts of Viana do Castelo, Braga, Vila Real, Porto, Bragança, Aveiro and Viseu), Center (Guarda, Castelo Branco, Coimbra, Leiria, Santarém and Portalegre), South (Lisboa, Évora, Setúbal, Beja and Faro), Madeira Island and Azores Islands (fig. 1). Epidemiologic representation of MJD was established using the district of origin of the MJD families. To define the district of origin of the Portuguese MJD families we accessed information contained in the patients' ascending genealogies. These genealogies allowed the most remote individual in each family for whom there was indication of an 'affected' status to be identified. Based on this information the patients were distributed according to districts (this measure was termed 'epidemiological representation' of MJD and not 'prevalence' of MJD, since to calculate prevalence the information used is usually relative to the patients' residence). Districts were then classified into 3 different groups: group 1 includes all districts which genealogical data could not identify as being the place of origin of any MJD family (Aveiro, Beja, Faro, Madeira Island, Porto, Viana do Castelo, Vila Real and Viseu); group 2 includes districts with values of epidemiological representation lower than 1/12,000 (Braga, Bragança, Castelo Branco, Coimbra, Évora, Leiria, Lisboa and Setúbal), and group 3 includes districts with values of epidemiological representation higher than 1/12,000 (Azores Islands, Guarda, Portalegre and Santarém).

The mean, variance range and skewness were determined for the distribution of the wild-type MJD alleles in the whole sample using SPSS 9.0.0 [36]. Population conformity with the Hardy-Weinberg equilibrium (HWE) expectations was tested using the Guo and Thompson [37] exact test with the Genepop package [38]. Bonferroni corrections were applied for multiple testing. Unbiased estimates of Nei's heterozigosity [39] were calculated with the Arlequin package [40]. Population differentiation was tested using the Fisher exact test as implemented in the Arlequin package [40]. Differences in alleles between districts and between groups of districts were evaluated using the Fisher exact test, as implemented in the

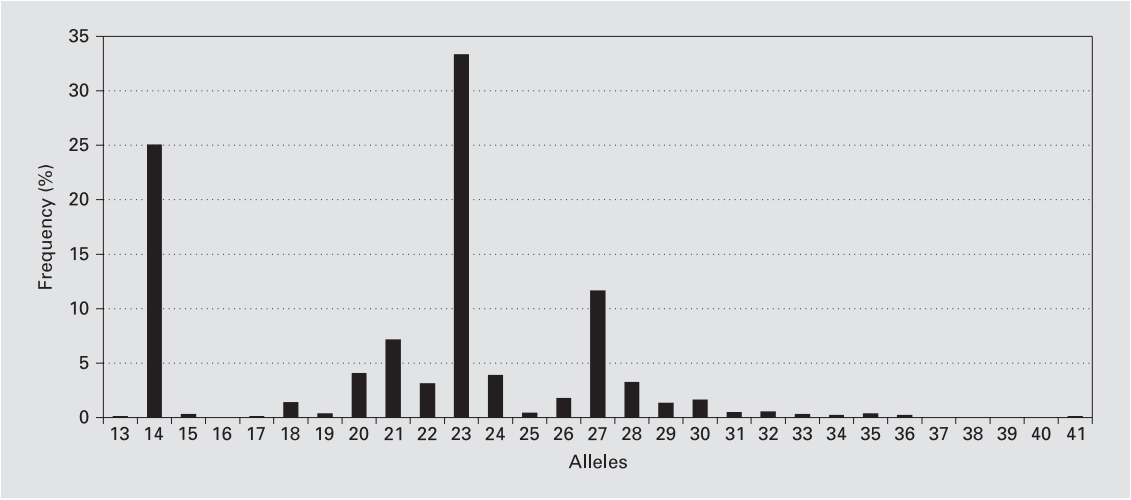

Fig. 2. Distribution of allelic frequencies of CAG repeats at the MJD locus in the Portuguese population (1,844 normal chromosomes).

STRUC program [38]. This program was also used to test the intra- and inter-group differentiation, considering the distribution of all alleles. The range of the wild-type alleles was divided into three classes, based on their overall distribution: small (less than 22 repeats), medium (between 22 and 29 repeats) and large (30 or more repeats). All correlations between epidemiological and genetic data were performed using SPSS 9.0.0 [36]. Power analyses were performed using Power and Precision v. 2.0 [41]. In all the analyses a significance level of 0.05 was used.

Results

Distribution of the Wild-Type CAG Repeats in the Total Sample

Twenty-four allelic variants were found for the CAG repeat-containing segment of the *MJD* gene. Repeat numbers were all in the normal range, varying from 13 to 41 units (fig. 2). The mean was 21.5 ± 5.0. No alleles of intermediate size were found. Alleles with 23 repeats (33.3%) were the most frequent. Besides allele 23, alleles 14 (24.9%) and 27 (11.6%) were clearly more frequent than the remaining. The observed distribution is in accordance with previous reports of the $(CAG)_n$ length in the Portuguese population [16]. Allelic size distribution was not normal (Kolmogorov-Smirnov, $Z = 8.964$; $p < 0.001$), showing a negative skew (–0.185 ± 0.057), which indicates the presence of a tail below the mode. In fact, nearly 70% of the chromosomes have a repeat size that equals or is inferior to the mode.

All districts exhibited conformity with HWE (table 1). Observed heterozygosity values were always higher than 67% (table 1). These values are in accordance with what has been reported for wild-type alleles in other triplet disease loci [42, 43]. The results obtained for gene diversity together with those provided by Hardy-Weinberg calculations indicate that the $(CAG)_n$ tract of the *MJD* gene can be used as a marker in population genetic studies.

Genetic Differentiation

The test of population differentiation based on allele distribution, and performed including all districts, revealed the absence of significant differences between districts ($p = 0.196$). This result shows the genetic homogeneity of Portugal as a whole, in which the CAG repetitive tract of the *MJD* gene is involved as a genetic system. The evaluation of differences between pairs of districts detected that only 17 out of 180 pairs compared showed differentiation; Guarda, Lisboa, Setúbal and Aveiro were the most differentiated districts. When the test was performed allele by allele, an analysis of differentiation between the districts revealed significant differences only for allele 18 ($p = 0.029$). An analysis of intra-group variation considering the groups defined in accordance with the geographic localization of the districts did not detect any differences within each group ($p = 0.351$ for the North, $p = 0.829$ for the Center, $p = 0.117$ for the South). In all three cases, the power computed is higher than 0.99 ($\alpha = 0.05$, two-tailed). A test of differentiation between the five groups considered revealed that these were not

Table 1. Number of chromosomes analyzed, observed and expected heterozygosity, p values for the exact test of population conformity with Hardy-Weinberg (HW) equilibrium expectations and frequency of grouped allele sizes by district and for the total sample analyzed

District	Number of chromosomes	Observed/expected heterozygosity	HW p value	Grouped allele sizes, %		
				small	medium	large
Aveiro	80	0.800/0.815	0.1456	42.7	54.2	3.1
Azores	96	0.729/0.845	0.1166	37.5	58.8	3.8
Beja	90	0.933/0.863	0.1314	33.3	60.0	6.7
Braga	100	0.820/0.835	0.9658	41.0	56.0	3.0
Bragança	86	0.907/0.847	0.6444	37.2	54.7	8.1
C. Branco	82	0.805/0.764	0.5703	36.6	62.2	1.2
Coimbra	98	0.673/0.759	0.1583	30.6	68.4	1.0
Faro	76	0.763/0.819	0.5355	41.3	56.5	2.2
Evora	92	0.717/0.803	0.0373*	36.8	57.9	5.3
Guarda	100	0.840/0.779	0.2960	32.0	65.0	3.0
Leiria	94	0.851/0.821	0.9187	37.2	58.5	4.3
Lisboa	98	0.918/0.749	0.2578	38.8	58.2	3.1
Portalegre	98	0.755/0.756	0.7581	47.9	48.9	3.2
Porto	94	0.787/0.791	0.2486	35.7	62.2	2.0
Madeira	94	0.766/0.834	0.098	30.9	63.8	5.3
Santarem	92	0.804/0.777	0.8219	40.2	58.7	1.1
Setubal	94	0.872/0.850	0.3363	34.0	59.6	6.4
Viana Castelo	96	0.813/0.805	0.1368	41.7	56.3	2.1
Vila Real	86	0.814/0.828	0.0373*	39.5	58.1	2.3
Viseu	98	0.755/0.798	0.4123	45.9	51.0	3.1
Portugal	1,844	0.800/0.801	0.5324	38.1	58.5	3.5

* After Bonferroni correction $p > 0.05$.

homogeneous ($p = 0.022$), indicating the presence of a geographical sub-structure in terms of the allelic frequencies. However, an analysis of the 5 groups, performed allele by allele, produced p values lower than 0.05 only for alleles 18, 20 and 23. To exclude the possibility that the absence of significant results in the allele-by-allele comparison could be due to a type II error, we performed a power calculation for each allele. The average power for large alleles (30 or more repeats) was of 0.76, with extremes of 0.64 and 0.87 ($\alpha = 0.05$, two-tailed). Moreover, we compared the allele distribution between districts grouping the alleles, no significant differences being detected (exact test, $p = 0.662$; power = 0.92, $\alpha = 0.05$, two-tailed). These results reinforce the idea that this sub-structuring is not related with the frequency of large sized alleles.

The differentiation tests conducted between pairs of geographic groups indicated that the populations from the Center show significant differences when compared with all the other groups, with the exception of the Azores Islands, which did not show statistically significant differences in any of the differentiation tests performed. The differences detected between Center-North, Center-South and Center-Madeira, are probably due to the distribution of alleles 18, 20 and 23. In fact, alleles 20 and 23 have higher frequencies in the Center when compared to the other groups, whereas allele 18 is relatively underrepresented in this region in contrast to what would be expected. In the North and South groups alleles 20 and 23 show lower frequencies than expected, whereas the opposite happens with allele 18. Additionally, the Center of Portugal shows the lowest heterozygosity (gene diversity) value (table 1).

Distribution of Repeats and Epidemiological Representation of MJD

A correlation test between the epidemiological representation of MJD observed in the different districts and the frequency of the 3 classes of alleles (small, intermediate and large) in the populations from those districts (table 1) was performed. There were no significant correlations between the values of representation of the disease and the frequency of small, medium or large alleles (Spearman's rho, $r = -0.179$, $p = 0.450$; $r = 0.271$, $p = 0.247$; $r = -0.417$, $p = 0.068$, respectively). The same analysis was performed considering the average value of epidemiological representation of MJD observed in each of the 5 geographical groups (North, Center, South, Madei-

ra and Azores) and the frequency of the 3 classes of wild-type alleles. Again, there were no significant correlations between these two variables (Spearman's rho, r = –0.300, p = 0.624; r = 0.300, p = 0.624; r = –0.600, p = 0.285, respectively). Analysis performed computed a power of 0.95 (α = 0.05, two-tailed) for this association. Further analyses were performed to fully assess the relationship between the epidemiology of MJD and the number of CAG motifs in the *MJD* gene. The districts were grouped according to their representation of MJD. The relation between the epidemiological representation in each of these groups and the frequency of small, medium and large alleles was tested. The association obtained was not significant (chi-square test, $\chi^2 = 6.115$, d.f. = 4, p = 0.191; Kendall Rank Correlation Test, tau-b = 0.018, p = 0.4143).

Discussion

The question of how the trinucleotide repeat loci, currently known to be associated with several human diseases, have evolved into their current lengths and degrees of polymorphism has been the subject of much debate. Various mechanisms have been proposed, but the processes that give rise to expanded alleles that cause disease in the human population remain mostly unexplained. In MJD, several arguments can be put forward to support one or a few mutational events on the basis of cases worldwide, including patients in the island of Flores within the Azores and a small region in the center part of north Portugal. Haplotype studies have shown that most of the mutant chromosomes worldwide share a common haplotype. A second haplotype is present in a subgroup of Portuguese patients on the Azorean Island of S. Miguel and on the mainland of Portugal, as well as in a few Japanese, North American, Brazilian and Spanish families. Two additional haplotypes have been identified, corresponding to a very small proportion of families, which may actually derive from the main haplotypes as a result of mutation of microsatellites [44]. Furthermore, studies associating intragenic haplotypes with intergenerational instability of repeat length in the normal chromosomes of Portuguese origin did not support the assumption of a subset of expansion-prone normal alleles [45]. As the evidence previously reported did not support the hypothesis of multiple mutational events (the mutated/expanded alleles being recurrently generated from the upper-tail of the normal variation), in order to understand the origins of the MJD mutation it was important to consider the information provided by the normal range variation. Availability of an extensively characterized population at an epidemiological level offered a unique opportunity to perform a series of tests, aiming to understand the dynamics of the *MJD* locus.

The widespread allelic distribution of the wild-type CAG alleles in our population (average of 21.46 ± 5.03, coefficient of variation of 23.44%) is similar to that presented by Andres et al. [32] for Europeans (average of 21.69 ± 5.17, coefficient of variation of 23.84%).

The fact that no intermediate alleles were found in this control population of Portuguese origin corroborates the findings of Maciel et al. [16] and is relevant for the molecular diagnosis of MJD. Furthermore, the absence of intermediate or expanded alleles in the normal population is relevant in order to understand the mutation pattern in the *MJD* gene, which contrasts with what has been observed for the Huntington disease gene, for which both intermediate and expanded alleles were identified in a similar sized sample of the Portuguese population [46]. The presence of distinct dynamics in relation to loci with very similar molecular characteristics is in accordance with what has been reported by Andres et al. [32], based on the allelic profile of 8 distinct CAG repeat loci, including the *MJD* gene. Andres et al. [32] acknowledge the widespread distribution of the wild-type MJD alleles, and attribute it to a high mutation rate following an unrestricted model without any bias to expansion or contraction. With respect to HD, a constrained mutation model with a bias toward expansion could explain the observed allelic frequencies better [32].

In the Portuguese population, a negative skew for the allele size distribution of the $(CAG)_n$ tract in the *MJD* gene was detected. This finding contrasts with what was reported by [47], who described a positive skew (implying an excess of normal alleles with longer repeats) for the allele size distribution of the MJD alleles in the Japanese population.

No allelic differentiation was detected between districts presenting distinct values of epidemiological representation of MJD. Furthermore, differences between geographical groups were not due to upper-tail alleles. Overall, none of the analyses conducted in this study was able to detect the existence of an association between the representation of MJD and the frequency of large alleles within the normal range.

Understanding the survival mechanisms of deleterious genes such as the *MJD* gene requires quantifying the selective pressures that it has been submitted to over the generations. In this context, phenotypic variability is a

crucial factor for quantifying selective pressure. For MJD, detailed fitness studies based on the genealogical reconstruction of the Portuguese affected families demonstrated that selection in the MJD locus only occurs relative to the early onset of more severe forms [48]. Since there is a correlation between early onset forms of the disease and larger expansions, larger alleles will also be selected against. Furthermore, these studies also demonstrated that patients with milder forms of the disease and onset after their reproductive period (>40 years) presented no differences in the fitness values in comparison to the controls [34, 35, 48, 49]. We can therefore imagine a mechanism for maintaining the *MJD* gene in the populations that postulates the existence of a few founding chromosomes from which all families originated. Selection against early onset (more severe) forms with larger expanded alleles would prevent the successive increase of allele size over the generations, whereas a normal fitness value for later onset (milder) forms of the disease would keep the remaining expanded alleles within the population.

In conclusion, our results obtained for MJD in a large sample of a population in which extensive epidemiological studies have been performed demonstrate that the epidemiological representation of the disease is not an indirect reflection of the frequency of the large normal alleles. These results, associated with the absence of intermediate alleles in the normal population and with the negative asymmetry of the distribution of the normal alleles, are in accordance with a model that postulates the presence of a few mutational events from which most of the MJD cases worldwide have originated.

Acknowledgments

This work was partially supported by the Regional Government of the Azores (Regional Project on Machado-Joseph disease). Cristina Santos (SFRH/BD/723/2000), Carmo Costa (SFRH/BD/9759/2003) and Anabela Ferro (SFRH/BD/1288/2000) are PhD grantees of Fundação para a Ciência e a Tecnologia (FCT). Rafael Montiel is a Post-Doc grantee of FCT (SFRH/BPD/13256/2003).

References

1 La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH: Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 1991;352:77–79.

2 Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, Sohn R, Zemelmen B, Snell RG, Rundle SA, Crow S, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, Johnson K, Harper PS, Shaw DJ, Housman DE: Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell 1992;68:799–808.

3 Orr HT, Chung MY, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LP, Zoghbi HY: Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet 1993;4:221–226.

4 The Huntington's Disease Collaborative research Group: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 1993;72:971–983.

5 Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, Kondo R, Ishikawa A, Hayashi T, Saito M, Tomoda A, Miike T, Naito H, Ikuta F, Tsuji S: Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 1994;6:9–13.

6 Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunkes A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S: Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet 1996;14:269–276.

7 David G, Abbas N, Stevanin G, Durr A, Yvert G, Cancel G, Weber C, Imbert G, Saudou F, Antoniou E, Drabkin H, Gemmill R, Giunti P, Benomar A, Wood N, Ruberg M, Agid Y, Mandel JL, Brice A: Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nat Genet 1997;17:65–70.

8 Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, Dobyns WB, Subramony SH, Zoghbi HY, Lee CC: Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. Nat Genet 1997;15:62–69.

9 Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, Ranum LP: An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat Genet 1999;21:379–384.

10 Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, Shimazaki H, Soutome M, Endo K, Ohta S, Kagawa Y, Kanazawa I, Mizuno Y, Yoshida M, Yuasa T, Horikawa Y, Oyanagi K, Nagai H, Kondo T, Inuzuka T, Onodera O, Tsuji S: The gene for Machado-Joseph disease maps to human chromosome 14q. Nat Genet 1993;4:300–304.

11 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: CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet 1994;8:221–228.

12 Coutinho P, Andrade C: Autosomal dominant system degeneration in Portuguese families of the Azores Islands. Neurology 1978;28:703–709.

13 Lima M, Mayer FM, Coutinho P, Abade A: Origins of a mutation: Population genetics of Machado-Joseph disease in the Azores (Portugal). Hum Biol 1998;70:1011–1023.

14 Lima M, Mayer F, Coutinho P, Abade A: Prevalence, geographic distribution, and genealogical investigation of Machado-Joseph disease in the Azores (Portugal). Hum Biol 1997;69:383–391.

15 Lima M, Kay T, Vasconcelos J, Mota-Vieira L, Gonzalez C, Peixoto A, Abade A, MacLeod P, Graca R, Santos J: Disease knowledge and attitudes toward predictive testing and prenatal diagnosis in families with Machado-Joseph disease from the Azores Islands (Portugal). Community Genet 2001;4:36–42.

16 Maciel P, Costa MC, Ferro A, Rousseau M, Santos CS, Gaspar C, Barros J, Rouleau GA, Coutinho P, Sequeiros J: Improvement in the molecular diagnosis of Machado-Joseph disease. Arch Neurol 2001;58:1821–1827.

17 Takiyama Y, Sakoe K, Nakano I, Nishizawa M: Machado-Joseph disease: cerebellar ataxia and autonomic dysfunction in a patient with the shortest known expanded allele (56 CAG repeat units) of the MJD1 gene. Neurology 1997;49:604–606.

18 van Schaik IN, Jobsis GJ, Vermeulen M, Keizers H, Bolhuis PA, de Visser M: Machado-Joseph disease presenting as severe asymmetric proximal neuropathy. J Neurol Neurosurg Psychiatry 1997;63:534–536.

19 van Alfen N, Sinke RJ, Zwarts MJ, Gabreels-Festen A, Praamstra P, Kremer BP, Horstink MW: Intermediate CAG repeat lengths (53,54) for MJD/SCA3 are associated with an abnormal phenotype. Ann Neurol 2001;49:805–808.

20 Takiyama Y, Igarashi S, Rogaeva EA, Endo K, Rogaev EI, Tanaka H, Sherrington R, Sanpei K, Liang Y, Saito M, Tsuda T, Takano H, Ikeda M, Lin C, Chi H, Kennedy JL, Lang AE, Wherrett JR, Segawa M, Nomura Y, Yuasa T, Weissenbach J, Yoshida M, Nishizawa M, Kidd KK, Tsuji S, St George-Hyslop PH: Evidence for inter-generational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. Hum Mol Genet 1995;4:1137–1146.

21 Deka R, Chakraborty R: Trinucletide repeats, genetic instability and variation in the human genome; in Papiha SS, Deka R, Chakraborty R (eds): Genomic Diversity: Applications in Human Population Genetics. New York, Kluwer Academic/Plenum Publishers, 1999, pp 53–64.

22 Parniewski P, Staczec P: Molecular Mechanisms of TRS instability; in Timchenko L (ed): Triplet Repeat Diseases of the Nervous System. New York, Kluwer Academic, 2002, pp 1–25.

23 Tsuji S: Unstable expansion of triplet repeats as a new disease. Jpn J Hum Genet 1996;41:279–290.

24 Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P, Radvany J, Dawson DM, Sudarsky L, Guimarães J, Loureiro JE, Nezarati MM, Corwin LI, Lopes-Cendes I, Rooke K, Rosenberg R, MacLeod P, Farrer LA, Sequeiros J, Rouleau GA: Correlation between CAG repeat length and clinical features in Machado-Joseph disease. Am J Hum Genet 1995;57:54–61.

25 Rubinsztein DC, Amos W, Leggo J, Goodburn S, Ramesar RS, Old J, Bontrop R, McMahon R, Barton DE, Ferguson-Smith MA: Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. Nat Genet 1994;7:525–530.

26 Amos W, Sawcer SJ, Feakes RW, Rubinsztein DC: Microsatellites show mutational bias and heterozygote instability. Nat Genet 1996;13:390–391.

27 Davies J, Yamagata H, Shelbourne P, Buxton J, Ogihara T, Nokelainen P, Nakagawa M, Williamson R, Johnson K, Miki T: Comparison of the myotonic dystrophy associated CTG repeat in European and Japanese populations. J Med Genet 1992;29:766–769.

28 Goldman A, Ramsay M, Jenkins T: Absence of myotonic dystrophy in southern African Negroids is associated with a significantly lower number of CTG trinucleotide repeats. J Med Genet 1994;31:37–40.

29 Watkins WS, Mamshad M, Jorde LB: Population Genetics of trinucleotide repeat polymorphisms. Hum Mol Genet 1995;4:1485–1491.

30 Zerylnick C, Torroni A, Sherman SL, Warren ST: Normal variation at the myotonic dystrophy locis in global human populations. Am J Hum Genet 1995;56:123–130.

31 Bertranpetit J, Calafell F: Genome versus population understanding in human genetic studies; in Donnelly P, Foley R (eds): Genes, Fossils and Behavior: An Integrated Approach. Amsterdam, IOS Press, 2001, pp 49–62.

32 Andres A, Lao O, Soldevila M, Calafell F, Bertranpetit J: Dynamics of CAG repeat loci revealed by the analysis of their variability. Hum Mutat 2002;21:61–70.

33 Silva MC, Coutinho P, Pinheiro CD, Neves JM, Serrano P: Hereditary ataxias and spastic paraplegias: Methodological aspects of a prevalence study in Portugal. J Clin Epidemiol 1997;50:1377–1384.

34 Abade A, Santos C, Lima M, Freire-Gonçalves A, Carvalho S, Coutinho P: Reproductive success in families with Machado-Joseph disease from the center of Portugal; in Varela TA (ed): Investigaciones en Biodiversidad Humana. Santiago de Compostela, Universidade de Santiago de Compostela, 2000, pp 17–24.

35 Santos C, Abade A, Lima M, Freire-Gonçalves A, Carvalho S, Coutinho P: Machado-Joseph disease in Portugal: the beginning of a long 'puzzle'; in Varela TA (ed): Investigaciones en Biodiversidad Humana. Santiago de Compostela: Universidade de Santiago de Compostela, 2000, pp 944–950.

36 SPSS Inc. SPSS for windows-release 9.0.0. Chicago, 1989–1999.

37 Guo SW, Thompson EA: Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics 1992;48:361–372.

38 Raymond M, Rousset F: GENEPOP 1.2: Population genetics software for exact tests and ecumenicism. J Hered 1995;86:248–249.

39 Nei M: Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 1978;89:583–590.

40 Schneider S, Roessli D, Excoffier L: ARLEQUIN: A software for population genetics data analysis 2.000. University of Geneva, Dept of Anthropology, Genetics and Biometry Lab, 2000.

41 Borenstein M, Rothstein H, Cohen J, Schoenfeld D, Berlin J: Power and Precision 2.00. Biostat, USA, 2000.

42 Leone O, Muglia M, Gabriele AL, Annesi G, Conforti FL, Imbrogno E, Imbrogno L, Brancati C: Analysis of the (CAG)n repeat at the IT15 locus in a population from Calabria (southern Italy). Hum Biol 1997;69:653–662.

43 Gyurus P, Molnar J, Melegh B, Toth G, Morava E, Kosztolanyi G, Mehes K: Trinucleotide repeat polymorphism at five disease loci in mixed Hungarian population. Am J Med Genet 1999;87:245–250.

44 Gaspar C, Lopes-Cendes I, Hayes S, Goto J, Arvidsson K, Dias A, Silveira I, Maciel P, Coutinho P, Lima M, Zhou YX, Soong BW, Watanable M, Giunti P, Stevanin G, Riess O, Sasaki H, Hsieh M, Nicholson GA, Brunt E, Higgins JJ, Lauritzen M, Tranebjaerg L, Volpini V, Wood N, Ranum L, Tsuji S, Brice A, Sequeiros J, Rouleau GA: Ancestral origins of the Machado-Joseph disease mutation: A worldwide haplotype study. Am J Hum Genet 2001;68:523–528.

45 Maciel P, Gaspar C, Guimarães L, Goto J, Lopes-Cendes I, Hayes S, Arvidsson K, Dias A, Sequeiros J, Sousa A, Rouleau GA: Study of three intragenic polymorphisms in the Machado-Joseph disease gene (MJD1) in relation to genetic instability of the (CAG)n tract. Eur J Hum Genet 1999;7:147–156.

46 Costa MC, Guimarães L, Ferreirinha F, Sousa A, Maciel P, Sequeiros J: Study of the normal CAG tract at the Huntington disease locus in the Portuguese population. Eur J Hum Genet 2002;10:263–264.

47 Takano H, Cancel G, Ikeuchi T, Lorenzetti D, Mawad R, Stevanin G, Didierjean O, Durr A, Oyake M, Shimohata T, Sasaki R, Koide R, Igarashi S, Hayashi S, Takiyama Y, Nishizawa M, Tanaka H, Zoghbi H, Brice A, Tsuji S: Close associations between prevalences of dominantly inherited spinocerebellar ataxias with CAG-repeat expansions and frequencies of large normal CAG alleles in Japanese and Caucasian populations. Am J Hum Genet 1998;63:1060–1066.

48 Lima M, Smith MT, Silva C, Abade A, Mayer FM, Coutinho P: Natural selection at the MJD locus: Phenotypic diversity, survival and fertility among Machado-Joseph disease patients from the Azores. J Biosoc Sci 2001;33:361–373.

49 Lima M: Machado-Joseph disease in the Azores: Epidemiological, Biodemographic and Genetic study. PhD thesis (in Portuguese). Ponta Delgada, University of the Azores, 1996.