

Asian Origin for the Worldwide-Spread Mutational Event in Machado-Joseph Disease

Sandra Martins, PhD; Francesc Calafell, PhD; Claudia Gaspar, PhD; Virginia C. N. Wong, MD, PhD; Isabel Silveira, PhD; Garth A. Nicholson, MD, PhD; Ewout R. Brunt, MD; Lisbeth Tranebjaerg, MD, PhD; Giovanni Stevanin, PhD; Mingli Hsieh, PhD; Bing-wen Soong, MD, PhD; Leal Loureiro, MD; Alexandra Dürr, MD, PhD; Shoji Tsuji, MD, PhD; Mitsunori Watanabe, MD, PhD; Laura B. Jardim, MD, PhD; Paola Giunti, MD, PhD; Olaf Riess, MD; Laura P. W. Ranum, PhD; Alexis Brice, MD; Guy A. Rouleau, MD, PhD; Paula Coutinho, MD, PhD; António Amorim, PhD; Jorge Sequeiros, MD, PhD

Background: Machado-Joseph disease is the most frequent dominant ataxia worldwide. Despite its frequency and presence in many populations, only 2 founder mutations have been suggested to explain its current geographic distribution.

Objectives: To trace back in history the main mutational events in Machado-Joseph disease, we aimed to assess ancestral haplotypes and population backgrounds, to date the mutations, and to trace the routes and time of introduction of the founder haplotypes in different populations.

Design, Setting, and Participants: We studied 264 families with Machado-Joseph disease from 20 different populations. Six intragenic single-nucleotide polymorphisms were used to determine ancestral mutational events; 4 flanking short tandem repeats were used to construct extended haplotypes and measure accumulation of genetic diversity over time within each lineage.

Results: The worldwide-spread lineage, TTACAC, had its highest diversity in the Japanese population, where we identified the ancestral short tandem repeat-based haplotype. Accumulated variability suggested a postneolithic mutation, about 5774 ± 1116 years old, with more recent introductions in North America, Germany, France, Portugal, and Brazil. As to the second mutational event, in the GTGGCA lineage, only 7 families (of 71 families) did not have Portuguese ancestry, although gene diversity was again smaller in Portuguese families (0.44) than in non-Portuguese families (0.93).

Conclusions: The worldwide-spread mutation may have first occurred in Asia and later been diffused throughout Europe, with a founder effect accounting for its high prevalence in Portugal; the other Machado-Joseph disease lineage is more recent, about 1416 ± 434 years old, and its dispersion may be explained mainly by recent Portuguese emigration.

Arch Neurol. 2007;64(10):1502-1508

MACHADO-JOSEPH DISEASE (MJD) is an autosomal dominant ataxia first described in families of Portuguese-Azorean extraction (Machado, Thomas, and Joseph families) in the United States.¹⁻³ Since then, MJD has been reported in many ethnic backgrounds and is known to be the most common spinocerebellar ataxia worldwide. Its relative frequency among the spinocerebellar ataxias is higher in Portugal (49%),⁴ China (49%),⁵ Brazil (44%⁶; 92% in Rio Grande do Sul, Brazil⁷), the Netherlands (44%),⁸ Japan (43%),⁹ and Germany (42%)¹⁰; its relative frequency is lower in France (33%),¹¹ the United States (21%),¹² and Australia (12%)¹³; and it is rare in the United Kingdom (5%),¹⁴ India (3%),¹⁵ and Italy (1%).¹⁶

Machado-Joseph disease is a neurodegenerative disorder characterized by ataxia and limitation of eye movements, com-

bined with more severe spasticity and dystonia in patients with earlier onset and with peripheral neuropathy in patients with later onset.¹⁷ The gene responsible, the ataxin 3 gene (ATXN3), includes a CAG repeat in exon 10. Although exceptions have been reported, normal alleles range from 12 to 44 CAG repeats, whereas patients with MJD most often have 1 expanded allele with 61 to 87 repeats.¹⁸

In our previous study, only 4 MJD intragenic haplotypes were found, but 2 (ACA or GGC) were present in 94% of all of the families.¹⁹ Among patients originating from the Azores, we found island-specific haplotypes: ACA in Flores and GGC in São Miguel; both haplotypes were found in mainland Portugal. The ACA haplotype was shared by 72% of all of the families worldwide, suggesting a founder mutation, although its origin and diffusion remained unresolved. The hypothesis was

Author Affiliations are listed at the end of this article.

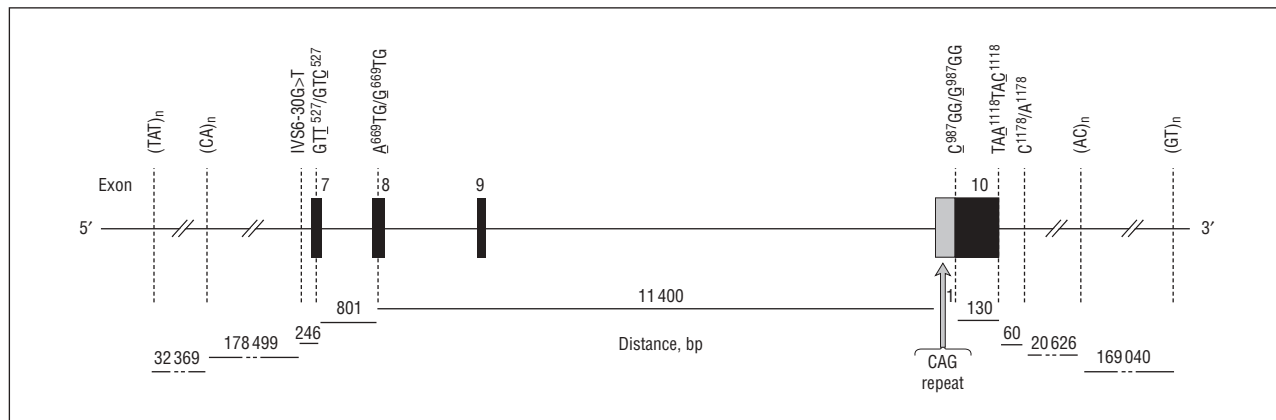


Figure 1. Schematic map of the polymorphisms studied within the ataxin 3 gene (*ATXN3*). Exons 7 through 10 are indicated with filled boxes and the CAG repeat in exon 10 is dotted. Distances are given in base pairs (bp).

raised that the original event occurred in mainland Portugal, spread to the archipelago during its colonization, and from there spread to Brazil, the United States, and Canada, while the Portuguese seafare could explain its presence in Japan, China, and India. Nevertheless, the possibility that the mutational event associated with ACA occurred in other populations and was brought to the Azores and mainland Portugal could not be excluded.

Here, we intend to clarify these questions, determining the origins, age, and spread of the 2 mutational events through more extensive haplotype analyses.

METHODS

FAMILIES WITH MJD AND CONTROLS

We studied 264 families with MJD, from Portugal (104 families), Brazil (37 families), France (29 families), Japan (27 families), North America (23 families), Germany (14 families), India (4 families), Australia (4 families), Spain (3 families), the Netherlands (2 families), Norway (2 families), Yemen (2 families), Taiwan (2 families), England (1 family), the West Indies (1 family), Belgium (1 family), Algeria (1 family), Somalia (1 family), the Ivory Coast (1 family), Cambodia (1 family), and unknown origin (4 families). Family origin was defined as the family's most remote known ancestry in the affected line (families classified here as North American had no known previous origin; among the Australian families, 3 claimed to have English, Scottish, and Chinese ancestry, but this was rather uncertain). We genotyped healthy individuals as controls from Europe (277 phase-known chromosomes from Portugal), Asia (100 unrelated chromosomes from China), and Africa (26 phase-known and 28 unrelated chromosomes from Angola and Mozambique). Peripheral blood samples were collected after informed consent was obtained, and genomic DNA was extracted by standard procedures.

IDENTIFICATION OF MJD LINEAGES

To analyze independent mutational events, we extended haplotypes, including $(CAG)_n$, the previously studied single-nucleotide polymorphisms (SNPs) $A^{669}TG/G^{669}TG$ (rs1048755), $C^{987}GG/G^{987}GG$ (rs12895357), and TAA^{1118}/TAC^{1118} (rs7158733), and the 3 additional SNPs $IVS6-30G>T$ (rs12590497), GTT^{527}/GTC^{527} (rs1130166), and C^{1178}/A^{1178} (rs3092822) (**Figure 1**).

HAPLOTYPE ANALYSES

Extended haplotypes were based on 4 flanking short tandem repeats (STRs), $(TAT)_n$, $(CA)_n$, $(AC)_n$, and $(GT)_n$ (Figure 1). Amplification and genotyping conditions were performed as previously described.²⁰ PHASE software version 2.0 (<http://www.stat.washington.edu/stephens/software.html>) was used to reconstruct haplotypes from genotypic data when the phase could not be directly inferred by family structure or by allele-specific amplification. Allele frequencies and phase-known haplotypes were taken into account, but only haplotype pairs with a probability greater than 0.6 were used for further analyses.

To estimate genetic distances among populations, we performed pairwise analyses with Arlequin software version 2.000 (<http://anthro.unige.ch/software/arlequin/software>), using the sum of square size difference (R_{ST}) as a measure of distance; R_{ST} is an analogue of F_{ST} suited for STR haplotype comparisons because it explicitly considers the number of single mutation steps between alleles. The resulting matrix of inter-population values was plotted with multidimensional scaling using Statistica version 6 software (StatSoft, Tulsa, Oklahoma).

ESTIMATION OF AGE OF MUTATIONS

We estimated the age of the 2 mutations from the variation accumulated in their ancestral haplotypes. This method was modified to include both recombination (c) and mutation (μ) rates in the generation of variation.^{21,22} The probability of change per generation (ϵ) was given by $\epsilon = 1 - [(1-c)(1-\mu)]$, and the average of mutation and recombination events (λ) equals ϵt , where t is the number of generations.

RESULTS

MJD LINEAGES

The TTACAC lineage was shared by 19 of 20 populations, whereas GTGGCA was observed only in Portugal and 7 non-Portuguese families, which were from North America (4 families), Spain (1 family), the West Indies (1 family), and unknown origin (1 family). In the Portuguese families, both haplotypes were frequent but their distribution was heterogeneous: all of the 12 families from the northeastern mainland and virtually all of the families from Flores shared the TTACAC haplotype; in São

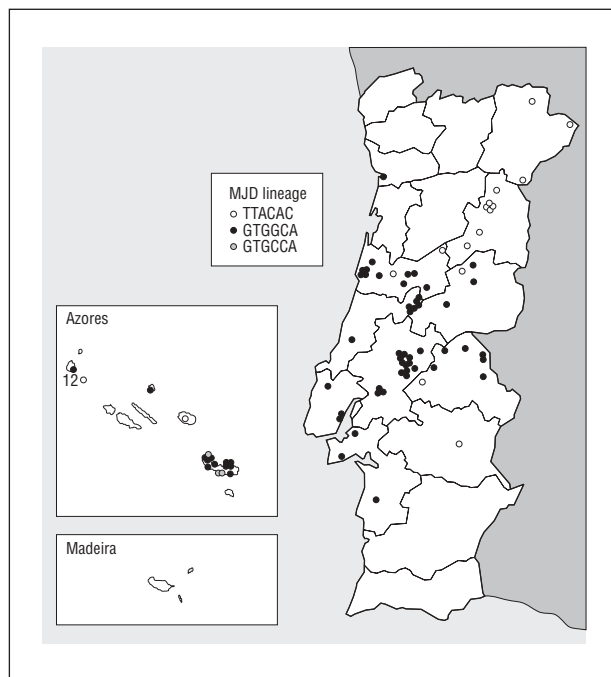


Figure 2. Geographical distribution of Portuguese families with Machado-Joseph disease (MJD). Three single-nucleotide polymorphism-based haplotypes are heterogeneously distributed in mainland Portugal, Flores, and São Miguel. Each circle represents 1 family; 12 indicates that 12 families with MJD carrying the TTACAC lineage were assessed from the island of Flores.

Miguel, 10 families had the GTGGCA haplotype (also predominant in the central mainland) and 3 had the GTGCCA haplotype (**Figure 2**).

POPULATIONS OF ORIGIN AND SPREAD OF MJD MUTATIONS

For TTACAC, we found the highest molecular diversity among Japanese families, followed (in this order) by North American, French, German, Portuguese, and Brazilian families (**Table 1**). This suggests that the worldwide spread of the MJD lineage is more ancient in Japan than in any other populations studied. Further introductions would have occurred in North America, then throughout Europe and later in Brazil, according to lower diversity among these populations.

A less clear picture emerged for the origin of the GTGGCA lineage. The vast majority of families shared a Portuguese origin, harboring 7 of 11 STR-based haplotypes (**Figure 3**); however, haplotypes in non-Portuguese families were more distant from each other, showing higher diversity (gene diversity \pm SD, 0.93 ± 0.12) when compared with Portuguese haplotypes, which were separated by a single mutation or recombination and displayed much less diversity (gene diversity \pm SD, 0.44 ± 0.08).

ANCESTRAL STR-BASED HAPLOTYPES

The most parsimonious relationships among flanking haplotypes are presented as phylogenetic networks (Figure 3 and **Figure 4**). The probability of mutation vs recombination was evaluated, considering the number of stepwise mutations required, observed intermediate haplotypes, and

Table 1. Gene Diversity Based on 4 Flanking Short Tandem Repeats Found in Families With Machado-Joseph Disease of the TTACAC Lineage From Different Origins

Origin of Families	Gene Diversity \pm SD
Japanese (n=22)	0.87 \pm 0.05
North American (n=12)	0.82 \pm 0.07
French (n=25)	0.59 \pm 0.11
German (n=13)	0.53 \pm 0.15
Portuguese (n=33)	0.33 \pm 0.10
Brazilian (n=29)	0.32 \pm 0.11

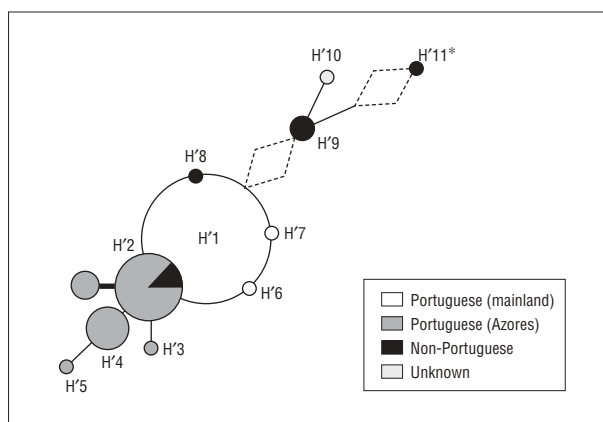


Figure 3. Phylogenetic network showing the most parsimonious relationships among flanking short tandem repeat-based haplotypes in families carrying the GTGGCA lineage. Haplotype H'11* was reconstructed by PHASE software version 2.0. A thicker line indicates the recurrent mutation in C⁹⁸⁷GG/G⁹⁸⁷GG; all of the 3 families with the GTGGCA haplotype shared the same flanking short tandem repeat alleles as H'2. Circle and line sizes are proportional to the number of families and stepwise mutations, respectively. Dashed diamonds indicate recombinations.

allele and haplotype frequency in controls (eFigure). To determine ancestral or founder haplotypes, we considered the following: (1) frequency in each population, (2) number of unchanged alleles, and (3) number of families with haplotypes just 1 step away from that putative ancestral haplotype. According to these criteria, the mutation in TTACAC seems to have occurred in the H1 (8-25-14-19) haplotype (**Table 2**). Introduction in North American, French, Portuguese, and Brazilian populations occurred through the H12 (11-21-14-15) haplotype, whereas the 1-step-away H13 (10-21-14-15) haplotype was introduced in Germany. In the United States, the founder haplotype H12 reflects the European contribution, whereas phylogenetically more distant haplotypes suggest another, more ancient introduction of the same lineage; whether this occurred directly from the ancestral background or another genetic background is to be determined. Although there were no MJD haplotypes common to North American and Japanese families, intermediate haplotypes may have been lost over time. Assuming the TTACAC lineage spread worldwide from East Asia, it is intriguing that we could not find MJD haplotypes of Japanese origin as founders among North American families or in higher frequency in European families. Other populations, however (not included in this analysis owing to low numbers of families), carried some

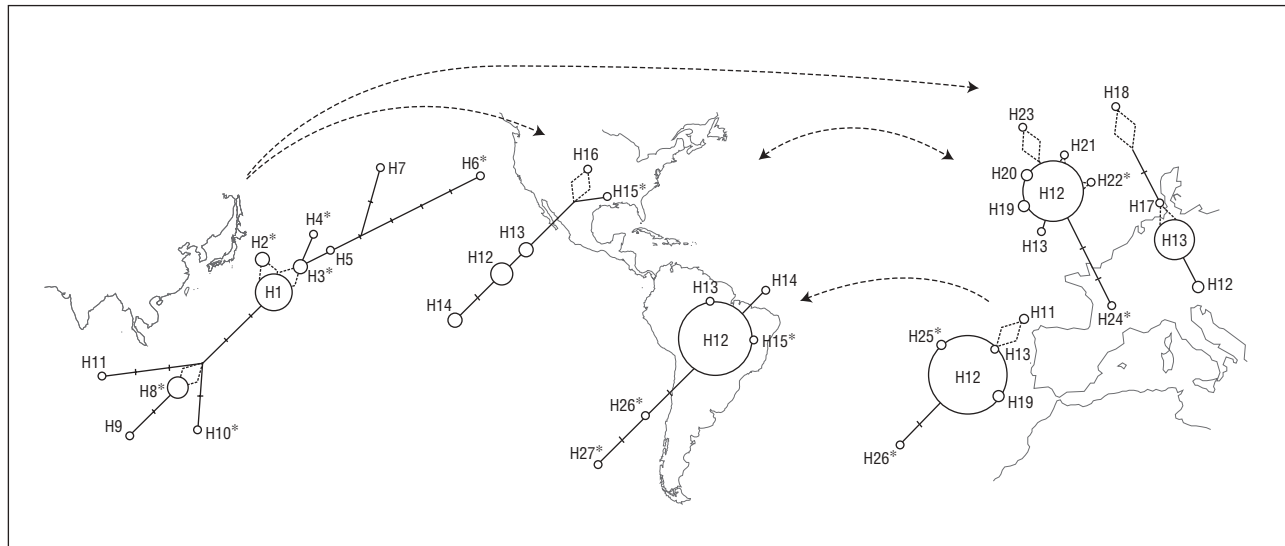


Figure 4. Principal routes of the worldwide-spread mutation in lineage TTACAC. Close to the respective country of origin is the phylogenetic network for Japanese, North American, Brazilian, Portuguese, French, and German families. Haplotypes with asterisks were reconstructed by PHASE software version 2.0. Circle and line sizes are proportional to the number of families and stepwise mutations, respectively. Dashed diamonds indicate recombinations.

haplotypes from the ancestral Japanese background, such as those from India (H7 haplotype), the Ivory Coast (H7 haplotype), and Yemen (H3 haplotype), in addition to new haplotypes, such as the H28 (11-21-12-19) and H29 (10-23-18-19) haplotypes in India, the H30 (10-20-14-15) and H31 (13-23-14-15) haplotypes in Taiwan, the H32 (10-22-14-15) haplotype in Cambodia, and the H33 (16-21-14-13) haplotype in Belgium. It is noteworthy that Indian families, closer to the Japanese families in the multidimensional scaling representation (**Figure 5**), shared a complete haplotype and many STR alleles with the putative ancestral population.

As for GTGGCA, the H'1 (10-28-18-19) haplotype seems to be the ancestral STR background, as it is shared by 65% of the families, whereas the rare haplotypes H'8, H'9, H'10, and H'11 could have originated and migrated from mainland Portugal to Spain, North America, and the West Indies. On the other hand, haplotypes H'3, H'4, and H'5, exclusively found in the Azores, derived from the second most frequent haplotype H'2 (16-22-18-19), a recombinant of H'1. In this population, the finding of a third SNP-based haplotype (GTGCCA), explainable by a single (recurrent) mutation in $\overline{C}^{987}GG/\overline{G}^{987}GG$, is also noteworthy.

AGE OF MUTATIONAL EVENTS

Assuming a generation time of 25 years and based on the most accurate recombination estimates (from family data), the age calculated for the worldwide-spread lineage TTACAC was 5774 ± 1116 years (Table 2). Introduction of this mutation in North America seems to have occurred much later (3396 ± 1182 years ago), in concordance with the genetic diversity found; nevertheless, these ignore the possibility of different introductions of mutations identical by descent in a given population. In Europe, TTACAC seems to be even more recent, with ages less than 2000 years in Germany and France and less than 1000 years in Portugal. As for GTGGCA, the estimated age regardless of its origin was 1416 ± 434 years.

COMMENT

This study focuses on the main MJD mutations spread worldwide, explaining the current disease distribution. Analysis of haplotype backgrounds of fast-evolving polymorphisms (STRs) within distinct lineages (identified by genotyping of SNPs, which are rather stable) allowed us to assess their origins and follow their spreading routes. It is expected that, in the absence of strong selective pressures, the ancestral population of a given lineage will show the largest diversity, as mutation and recombination would have more time to generate new STR haplotypes. This should be taken with caution owing to the large variance that stochastic events underlying the evolutionary process carry: a large sample size is needed to infer origin, and inferences based on small samples may not be reliable. Moreover, accumulation of variation is also proportional to the effective population size; in smaller populations, haplotypes will coalesce earlier and thus accumulate less variation.

TTACAC LINEAGE

The high haplotype diversity among Japanese families as compared with all of the other families with the TTACAC lineage supports its Asian origin. Regarding the high frequency of this SNP-based haplotype among Asian controls as compared with the European and African control populations,²⁰ it could be hypothesized that recurrent mutations have been contributing to this high diversity. However, contrary to other trinucleotide repeat disorders, there are no reports to our knowledge of de novo MJD expansions, and intermediate alleles are rarely found. This is in accordance with a model where very few mutational events account for the cases observed nowadays.

The higher levels of diversity among French and German families as compared with Portuguese families do not support the hypothesis of an introduction of the TTACAC

Table 2. Age Estimation of the 2 Ancestral Machado-Joseph Disease Mutational Events and of the Introduction of the Worldwide-Spread TTACAC Lineage in Different Populations

MJD Lineage	Origin of Families	Haplotype ^a	Families, No. ^b	Mutation Steps, No.	Age ± δ , y ^c	
					$\epsilon = 0.0087$	$\epsilon = 0.0104$
TTACAC	Japanese	H1: 8-25-14-19 ^d	7	0	6923 ± 1334	5774 ± 1116
		H2*: 8-25-14-16	2 (0.8)	1		
		H3*: 16-22- 14-19	2 (0.6; 0.9)	1		
		H4*: 17-22- 14-19	1 (0.7)	2		
		H5: 16-23- 14-19	1	2		
		H6*: 15-27- 14-19	1 (0.6)	7		
		H7: 15-23-12- 19	1	5		
		H8*: 10-24-12-16	4 (0.7; 0.9)	4		
		H9: 10-22-12-16	1	6		
		H10*: 10-24-12- 19	1 (0.7)	5		
		H11: 10-21- 14-19	1	6		
	North American	H12: 11-21-14-15 ^d	4	0	4071 ± 1411	3396 ± 1182
		H13: 10-21- 14-15	3	1		
		H14: 11-21-14-13	3	2		
		H15*: 10-24- 14-15	1 (0.6)	4		
		H16: 10-23- 14-19	1	4		
		H13: 10-21-14-15 ^d	9	0		
	German	H12: 11-21-14-15	2	1	1547 ± 661	1291 ± 553
		H17: 16-23- 14-15	1	1		
		H18: 16-21- 14-19	1	4		
	French	H12: 11-21-14-15 ^d	16	0	1494 ± 472	1246 ± 395
		H19: 11-21-13-15	2	1		
		H13: 10-21- 14-15	1	1		
		H20: 11-20-14-15	2	1		
		H21: 11-21-14-19	1	1		
		H22*: 16-25- 14-15	1 (0.7)	1		
		H23: 10-29- 14-16	1	2		
		H24*: 15-21- 14-15	1 (0.6)	4		
	Portuguese	H12: 11-21-14-15 ^d	27	0	784 ± 288	654 ± 241
		H13: 10-21- 14-15	1	1		
		H19: 11-21-13-15	2	1		
		H11: 10-21- 14-19	1	2		
		H25*: 11-21-14-16	1 (0.7)	1		
H26*: 14-21- 14-15		1 (0.7)	3			
H12: 11-21-14-15 ^d		24	0			
Brazilian	H13: 10-21- 14-15	1	1	1189 ± 342	992 ± 286	
	H14: 11-21-14-13	1	2			
	H15*: 10-24- 14-15	1 (0.9)	1			
	H26*: 14-21- 14-15	1 (0.7)	3			
	H27*: 16-21- 14-15	1 (0.7)	5			
	H'1: 10-28-18-19 ^d	43	0			
	H'2: 16-22- 18-19	8	1			
GTGGCA	Worldwide	H'3: 16-23- 18-19	2	2	1698 ± 519	1416 ± 434
		H'4: 17-22- 18-19	5	2		
		H'5: 17-22-19- 19	1	3		
		H'6: 10-27-18-19	1	1		
		H'7: 11-23- 18-19	1	1		
		H'8: 10-28-18-15	1	1		
		H'9: 11-28-14-18	2	2		
		H'10: 11-27-14-18	1	3		
		H'11*: 16-28-14-19	1 (0.9)	4		

Abbreviation: MJD, Machado-Joseph Disease.

^a Alleles conserved in each population appear in bold. Haplotypes with an asterisk were reconstructed by the Bayesian method implemented in PHASE software version 2.0.

^b The probabilities for the haplotypes reconstructed in PHASE software version 2.0 are shown in parentheses.

^c The recombination rate (c) was based on the physical distance between the 2 most distant markers (413.2 kilobases; $c = 0.0058$ cM) using both a conversion factor of 1 megabase = 1.41 cM and the observed meiotic events (2 recombinations per 268 meioses; $c = 0.0075$ cM). The estimated probability of mutation per generation and per short tandem repeat was 2.95×10^{-3} (as 1 trinucleotide and 3 dinucleotide short tandem repeats were studied). The mutation rate for trinucleotides (6.13×10^{-4}) was calculated as the median value between the rate for dinucleotide repeats (7.8×10^{-4}) and tetranucleotide repeats (1.5-2.0 times lower than dinucleotides).

^d Ancestral or founder haplotypes.

lineage in Europe exclusively from Portugal. In Europe, the presence of TTACAC expanded chromosomes is recent and indicates that this has not spread along with the Neolithic culture. Moreover, the low diversity and close phyloge-

netic relationships among haplotypes of Portuguese and Brazilian families suggest that historical links between these 2 populations are responsible for the presence of MJD in Brazil. In North America, separate introductions of the

TTACAC lineage may have occurred directly from the ancestral Asian population, Europe, or both.

GTGGCA LINEAGE

As for GTGGCA, results are more controversial. Although it could be argued that haplotype variability was underestimated owing to sampling effects, this seems unlikely if we consider the high number of families studied and their rather distinct origins; thus, our findings might still be compatible with a Portuguese origin. Moreover, GTGGCA reaches its higher frequency in controls in Europe.²⁰ Taken at face value, the lower diversity among Portuguese haplotypes would imply a non-Portuguese origin for the ancestral line of GTGGCA; nevertheless, most of these chromosomes came from small isolated populations such as the Azores, which could lead to an underestimation of its age. One could contemplate the hypothesis of a non-Portuguese origin with a subsequent expansion within Portugal and loss of the ancestral haplotype; this is, however, a less parsimonious explanation.

AGE AND SPREAD OF ANCESTRAL MUTATIONAL EVENTS

The ancient TTACAC lineage may have occurred during the Bronze Age; the GTGGCA lineage, regardless of its population of origin, may have occurred less than 2000 years BP. Thus, both MJD mutations seem to be fairly recent as compared with other expanded repeat disorders. In Friedreich ataxia, the origin of expanded alleles was dated at about 25 000 years ago, following the population expansion of the Upper Paleolithic²³; for type 2 myotonic dystrophy, the founding haplotype was calculated to be 5000 to 12 500 years old²⁴; in Huntington disease, the presence of the ancestral haplotype only in the Land of Valencia (Spain) was estimated to range from 4700 to 10 000 years old.²⁵

The age estimated for the expansion of TTACAC could explain its diffusion westward from East Asia through the trade routes established in central Asia known as the Silk Road; however, these results do not support admixture between Indian and Portuguese MJD populations and are against the hypothesis of MJD in India having been introduced from Portugal¹⁹ as its reverse.¹⁵ Historical links with Portugal, on the other hand, would be responsible for MJD among Brazilian families.

If the GTGGCA lineage indeed has a Portuguese origin, we could speculate about the Azores vs mainland Portugal as its starting point. Taking into account only the haplotype diversity in the Azores together with a recurrent mutation in C⁹⁸⁷GG/G⁹⁸⁷GG (a rare event) in that population, the H'1 haplotype could be a recombinant from the ancestral H'2, whereas the high number of mainland Portuguese families with the H'1 haplotype would be explained by a founder effect. Nevertheless, phylogenetic relationships among haplotypes from different populations suggest that more diversity has been generated from H'1, supporting its mutational origin in mainland Portugal. This fits well with the fact that the Azores were inhabited when they were first discovered in 1439, subsequently being colonized mostly from mainland Portugal.

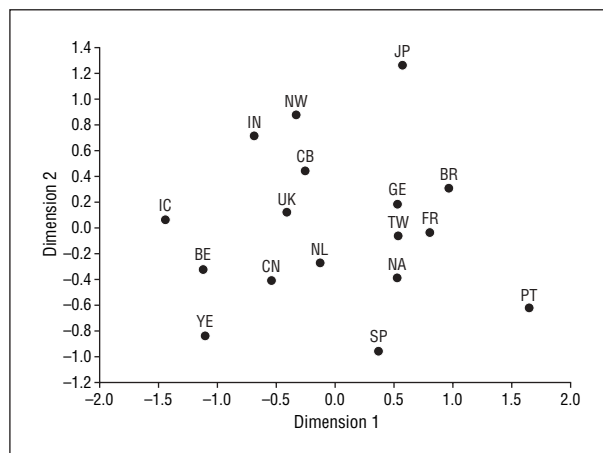


Figure 5. Multidimensional scaling analysis of interpopulational genetic distance (R_{ST} values) from short tandem repeat-based haplotypes of all of the families with the TTACAC lineage. PT indicates Portugal; SP, Spain; CB, Cambodia; NA, North America; JP, Japan; CN, China; TW, Taiwan; UK, United Kingdom; FR, France; GE, Germany; BE, Belgium; BR, Brazil; IN, India; NL, the Netherlands; NW, Norway; YE, Yemen; and IC, the Ivory Coast.

CONCLUSIONS

The scarcity of de novo expansions in MJD led us to question worldwide disease frequencies. Considering the clinical phenomenon of anticipation together with the rarity of intermediate alleles that could act as a reservoir for recurrent full mutations, the extinction of expanded MJD alleles would be expected; however, we must take into account that a theoretical progressive elimination of alleles from the human gene pool must be faced over an evolutionary timescale where other demographic events interfere as well. Moreover, distinct mutational mechanisms may apply for different repeat disorders; it would be interesting to explore *cis*- or *trans*-acting factors involved in repeat stability in MJD.

Accepted for Publication: November 28, 2006.

Author Affiliations: Instituto de Patologia e Imunologia Molecular da Universidade do Porto and Faculdade de Ciências (Drs Martins and Amorim), Instituto de Biologia Molecular e Celular (Drs Silveira, Loureiro, and Sequeiros), and Instituto de Ciências Biomédicas Abel Salazar (Dr Sequeiros), University of Porto, Porto, Portugal; Unitat Biologia Evolutiva, Departament de Ciències Experimentals i de la Salut, University of Pompeu Fabra, Barcelona, Spain (Dr Calafell); Centre Hospitalier de l'Université de Montréal, Hôpital Notre Dame, Montreal, Quebec, Canada (Drs Gaspar and Rouleau); Division of Child Neurology, Developmental Paediatrics, and Neurohabilitation, Department of Paediatrics and Adolescent Medicine, The University of Hong Kong, Hong Kong, China (Dr Wong); Department of Medicine, Concord Hospital, University of Sydney, Sydney, Australia (Dr Nicholson); University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands (Dr Brunt); Department of Audiology, Bispebjerg Hospital, Copenhagen, Denmark (Dr Tranebjaerg); Department of Medical Genetics, University Hospital of Northern Norway, Tromsø, Norway (Dr Tranebjaerg); Institut National de la Santé et de la Recherche Médicale Unit 679,

Federative Institute for Neuroscience Research, Pitié-Salpêtrière Group, Paris, France (Drs Stevanin, Dürr, and Brice); Department of Life Science, Tunghai University, Taichung, Taiwan (Dr Hsieh); Neurological Institute, Veterans General Hospital, Taipei, Taiwan (Dr Soong); Department of Neurology, University of Tokyo, Tokyo, Japan (Dr Tsuji); Department of Neurology, Gunma University School of Medicine, Gunma, Japan (Dr Watanabe); Serviço de Genética Médica, Hospital Clínicas, Porto Alegre, Brazil (Dr Jardim); Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London, England (Dr Giunti); Department of Medical Genetics, University of Tübingen, Tübingen, Germany (Dr Riess); Department of Genetics, Cell Biology, and Development and Institute of Human Genetics, University of Minnesota, Minneapolis (Dr Ranum); and Serviço de Neurologia, Hospital São Sebastião, Santa Maria da Feira, Portugal (Dr Coutinho). Dr Watanabe is now with the Department of Neurology, Hirosaki University Graduate School of Medicine, Hirosaki, Japan.

Correspondence: Jorge Sequeiros, MD, PhD, Unidade de Investigação Genética e Epidemiológica em Doenças Neurológicas, Instituto de Biologia Molecular e Celular, Rua Campo Alegre 823, 4150-180 Porto, Portugal (jsqueir@ibmc.up.pt).

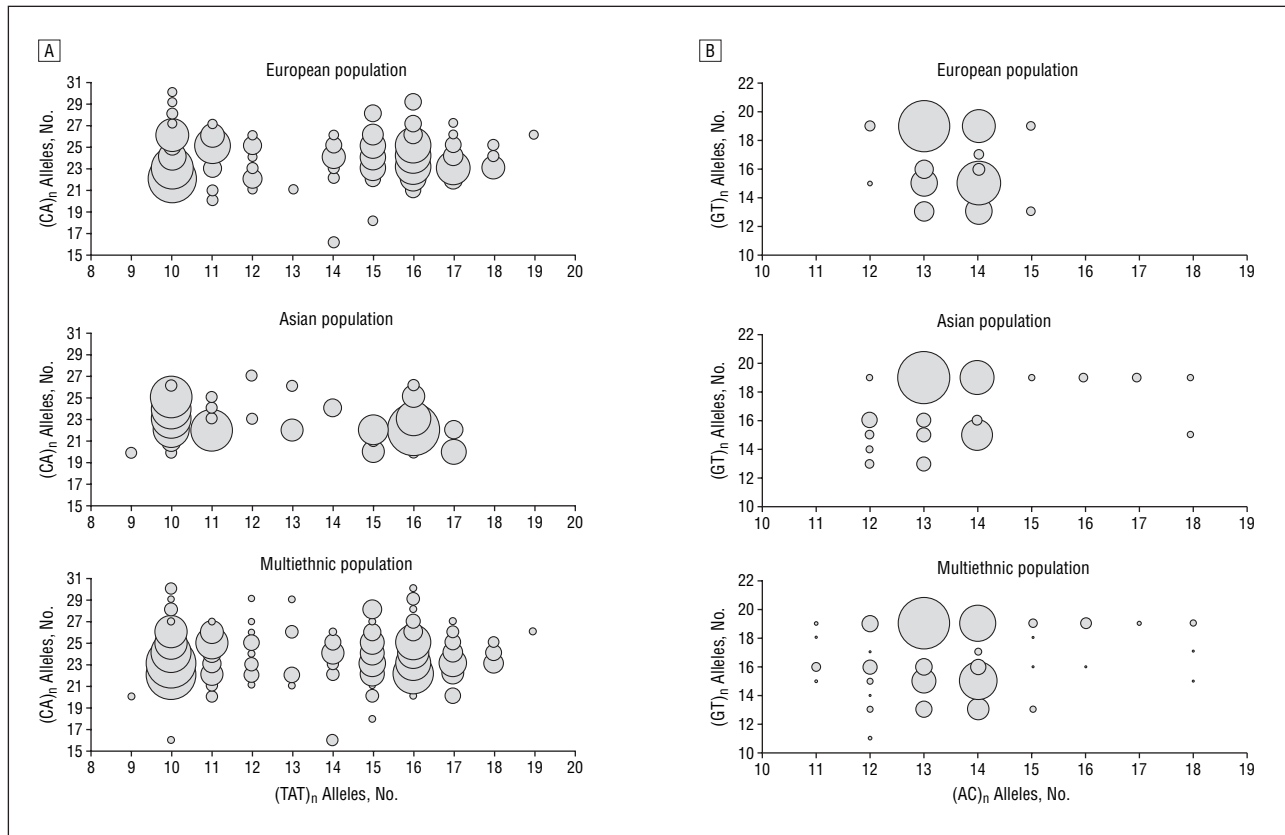
Author Contributions: *Study concept and design:* Martins, Coutinho, Amorim, and Sequeiros. *Acquisition of data:* Martins, Gaspar, Wong, Silveira, Nicholson, Brunt, Tranebjaerg, Stevanin, Hsieh, Soong, Loureiro, Dürr, Tsuji, Watanabe, Jardim, Giunti, Riess, Ranum, Brice, Rouleau, and Coutinho. *Analysis and interpretation of data:* Martins, Calafell, Silveira, Amorim, and Sequeiros. *Drafting of the manuscript:* Martins, Amorim, and Sequeiros. *Critical revision of the manuscript for important intellectual content:* Calafell, Gaspar, Wong, Silveira, Nicholson, Brunt, Tranebjaerg, Stevanin, Hsieh, Soong, Loureiro, Dürr, Tsuji, Watanabe, Jardim, Giunti, Riess, Ranum, Brice, Rouleau, and Coutinho. *Statistical analysis:* Martins, Calafell, and Amorim. *Obtained funding:* Amorim and Sequeiros. *Study supervision:* Calafell, Coutinho, Amorim, and Sequeiros. **Financial Disclosure:** None reported.

Funding/Support: This work was supported in part by Fundação para a Ciência e Tecnologia, through a research grant from Programa Operacional Ciência e Inovação (Instituto de Patologia e Imunologia Molecular da Universidade do Porto), and through Financiamento Plurianual de Unidades (Unidade de Investigação Genética e Epidemiológica em Doenças Neurológicas, Instituto de Biologia Molecular e Celular). Dr Martins was a recipient of scholarship SFRH/BD/8880/2002 from Fundação para a Ciência e Tecnologia.

Additional Information: The eFigure is available at <http://www.archneuro.com>.

Additional Contributions: Albertino Damasceno, MD, and Benilde Soares, MD, of the Eduardo Mondlane University, Maputo, provided the Mozambican control samples. Nicholas Wood, MD, PhD, provided collaboration. Paula Magalhães, MSc, Susana Carrilho, MSc, and Joana Cerqueira, MSc, provided efficient technical assistance at Centro de Genética Preditiva e Preventiva, Instituto de Biologia Molecular e Celular. We are grateful to the patients and relatives who agreed to collaborate in this project.

- Nakano KK, Dawson DM, Spence A. Machado disease: a hereditary ataxia in Portuguese emigrants to Massachusetts. *Neurology*. 1972;22(1):49-55.
- Woods BT, Schaumburg HH. Nigro-spino-dentatal degeneration with nuclear ophthalmoplegia: a unique and partially treatable clinico-pathological entity. *J Neurol Sci*. 1972;17(2):149-166.
- Rosenberg RN, Nyhan WL, Bay C, Shore P. Autosomal dominant striato-nigral degeneration: a clinical, pathologic and biochemical study of a new genetic disorder. *Neurology*. 1976;26(8):703-714.
- Silveira I, Miranda C, Guimarães L, et al. Trinucleotide repeats in 202 families with ataxia: a small expanded (CAG)_n allele at the SCA17 locus. *Arch Neurol*. 2002;59(4):623-629.
- Jiang H, Tang B, Xu B, et al. Frequency analysis of autosomal dominant spinocerebellar ataxias in Han population in the Chinese mainland and clinical and molecular characterization of spinocerebellar ataxia type 6 [in Chinese]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi*. 2005;22(1):1-4.
- Lopes-Cendes I, Teive HG, Calcagnotto ME, et al. Frequency of the different mutations causing spinocerebellar ataxia (SCA1, SCA2, MJD/SCA3 and DRPLA) in a large group of Brazilian patients. *Arq Neuropsiquiatr*. 1997;55(3B):519-529.
- Jardim LB, Silveira I, Pereira ML, et al. A survey of spinocerebellar ataxia in South Brazil: 66 new cases with Machado-Joseph disease, SCA7, SCA8, or unidentified disease-causing mutations. *J Neurol*. 2001;248(10):870-876.
- van de Warrenburg BP, Sinke RJ, Verschuuren-Bemelmans CC, et al. Spinocerebellar ataxias in the Netherlands: prevalence and age at onset variance analysis. *Neurology*. 2002;58(5):702-708.
- Takano H, Cancel G, Ikeuchi T, et al. 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(4):1060-1066.
- Schöls L, Amoiridis G, Büttner T, Przuntek H, Epplen JT, Riess O. Autosomal dominant cerebellar ataxia: phenotypic differences in genetically defined subtypes? *Ann Neurol*. 1997;42(6):924-932.
- Stevanin G, Dürr A, David G, et al. Clinical and molecular features of spinocerebellar ataxia type 6. *Neurology*. 1997;49(5):1243-1246.
- Moseley ML, Benzow KA, Schut LJ, et al. Incidence of dominant spinocerebellar and Friedreich triplet repeats among 361 ataxia families. *Neurology*. 1998;51(6):1666-1671.
- Storey E, du Sart D, Shaw JH, et al. Frequency of spinocerebellar ataxia types 1, 2, 3, 6, and 7 in Australian patients with spinocerebellar ataxia. *Am J Med Genet*. 2000;95(4):351-357.
- Giunti P, Sabbadini G, Sweeney MG, et al. The role of the SCA2 trinucleotide repeat expansion in 89 autosomal dominant cerebellar ataxia families: frequency, clinical and genetic correlates. *Brain*. 1998;121:459-467.
- Mittal U, Srivastava AK, Jain S, Jain S, Mukerji M. Founder haplotype for Machado-Joseph disease in the Indian population: novel insights from history and polymorphism studies. *Arch Neurol*. 2005;62(4):637-640.
- Brusco A, Gellera C, Cagnoli C, et al. Molecular genetics of hereditary spinocerebellar ataxia: mutation analysis of spinocerebellar ataxia genes and CAG/CTG repeat expansion detection in 225 Italian families. *Arch Neurol*. 2004;61(5):727-733.
- Coutinho P, Andrade C. Autosomal dominant systemic degeneration in Portuguese families of the Azores Islands: a new genetic disorder involving cerebellar, pyramidal, extrapyramidal and spinal cord motor functions. *Neurology*. 1978;28(7):703-709.
- Maciel P, Costa MC, Ferro A, et al. Improvement in the molecular diagnosis of Machado-Joseph disease. *Arch Neurol*. 2001;58(11):1821-1827.
- Gaspar C, Lopes-Cendes I, Hayes S, et al. Ancestral origins of the Machado-Joseph disease mutation: a worldwide haplotype study. *Am J Hum Genet*. 2001;68(2):523-528.
- Martins S, Calafell F, Wong VCN, Sequeiros J, Amorim A. A multi-step mutation mechanism drives the evolution of the CAG repeat at MJD/SCA3 locus. *Eur J Hum Genet*. 2006;14(8):932-940.
- Morral N, Bertranpetit J, Estivill X, et al. The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nat Genet*. 1994;7(2):169-175.
- Cobo AM, Saenz A, Poza JJ, et al. A common haplotype associated with the Basque 2362AG→TCATCT mutation in the muscular calpain-3 gene. *Hum Biol*. 2004;76(5):731-741.
- Labuda M, Labuda D, Miranda C, et al. Unique origin and specific ethnic distribution of the Friedreich ataxia GAA expansion. *Neurology*. 2000;54(12):2322-2324.
- Bachinski LL, Udd B, Meola G, et al. Confirmation of the type 2 myotonic dystrophy (CCTG)_n expansion mutation in patients with proximal myotonic myopathy/proximal myotonic dystrophy of different European origins: a single shared haplotype indicates an ancestral founder effect. *Am J Hum Genet*. 2003;73(4):835-848.
- García-Planells J, Burguera JA, Solís P, et al. Ancient origin of the CAG expansion causing Huntington disease in a Spanish population. *Hum Mutat*. 2005;25(5):453-459.



eFigure. Frequency of Machado-Joseph disease repeat-flanking haplotypes in control populations of European (n=277), Asian (n=100), and multiethnic (including 54 African normal chromosomes in addition to those analyzed from European and Asian populations) origin. A, Upstream haplotypes with (TAT)_n and (CA)_n markers. B, Downstream haplotypes with (AC)_n and (GT)_n markers.