



# Is the High Frequency of Machado-Joseph Disease in China Due to New Mutational Origins?

Tianjiao Li<sup>1</sup>, Sandra Martins<sup>2</sup>, Yun Peng<sup>1</sup>, Puzhi Wang<sup>1</sup>, Xiaocan Hou<sup>1</sup>, Zhao Chen<sup>1</sup>, Chunrong Wang<sup>1</sup>, Zhaoli Tang<sup>1</sup>, Rong Qiu<sup>3</sup>, Chao Chen<sup>4</sup>, Zhengmao Hu<sup>4</sup>, Kun Xia<sup>4</sup>, Beisha Tang<sup>1,4,5,6,7,8,9</sup>, Jorge Sequeiros<sup>10</sup> and Hong Jiang<sup>1,4,6\*</sup>

<sup>1</sup> Department of Neurology, Xiangya Hospital, Central South University, Changsha, China, <sup>2</sup> IPATIMUP – Institute of Molecular Pathology and Immunology of the University of Porto, Instituto de Investigação e Inovação em Saúde (i3S), Porto, Portugal, <sup>3</sup> School of Information Science and Engineering, Central South University, Changsha, China, <sup>4</sup> Laboratory of Medical Genetics, Central South University, Changsha, China, <sup>5</sup> National Clinical Research Center for Geriatric Diseases, Xiangya Hospital, Central South University, Changsha, China, <sup>6</sup> Key Laboratory of Hunan Province in Neurodegenerative Disorders, Central South University, Changsha, China, <sup>7</sup> Parkinson's Disease Center of Beijing Institute for Brain Disorders, Beijing, China, <sup>8</sup> Collaborative Innovation Center for Brain Science, Shanghai, China, <sup>9</sup> Collaborative Innovation Center for Genetics and Development, Shanghai, China, <sup>10</sup> IBMC – Institute for Molecular and Cell Biology, i3S – Instituto de Investigação e Inovação em Saúde, ICBAS (Instituto de Ciências Biomédicas Abel Salazar), University of Porto, Porto, Portugal

## OPEN ACCESS

### Edited by:

James J. Cai,  
Texas A&M University, United States

### Reviewed by:

Tina T. Hu,  
Princeton University, United States  
Sajid Ali,  
University of Agriculture, Peshawar,  
Pakistan

### \*Correspondence:

Hong Jiang  
jianghong73868@126.com

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Genetics

Received: 24 June 2018

Accepted: 22 December 2018

Published: 20 February 2019

### Citation:

Li T, Martins S, Peng Y, Wang P,  
Hou X, Chen Z, Wang C, Tang Z,  
Qiu R, Chen C, Hu Z, Xia K, Tang B,  
Sequeiros J and Jiang H (2019) Is  
the High Frequency  
of Machado-Joseph Disease in China  
Due to New Mutational Origins?  
Front. Genet. 9:740.  
doi: 10.3389/fgene.2018.00740

Machado-Joseph disease (MJD, also known as spinocerebellar ataxia 3 or SCA3) is the most common dominant ataxia worldwide, with an overall average prevalence of 1–5/100,000. To this date, two major ancestral lineages have been found throughout the world. In China, the relative frequency of MJD among the SCAs reaches as high as 63%, however, little is known about its mutational origin in this country. We analyzed 50 families with MJD patients in two or more generations to study the hypothesis that new mutational events have occurred in this population. Haplotypes based on 20 SNPs have shown new genetic backgrounds segregating with MJD mutations in our cohort from China. We found the “Joseph-derived” lineage (Joseph lineage with a G variant in rs56268847) to be very common among Chinese MJD patients. Moreover, we estimated the time for the origin of this MJD SNP background based on STR diversity flanking the (CAG)<sub>n</sub> of ATXN3. It was surprising to find that the Chinese MJD population originated from 8,000 to 17,000 years ago, far earlier than the previous literature reports, which will be an important evidence to explain the origin, spread and founder effects of MJD.

**Keywords:** spinocerebellar ataxia type 3, SCA3, Machado-Joseph disease, founder effect, haplotype, mutational origins

## INTRODUCTION

Machado-Joseph disease (MJD, OMIM#109150), also known as spinocerebellar ataxia type 3 (SCA3), is one of the polyQ diseases. It is a rare autosomal dominantly inherited neurodegenerative disease that causes progressive cerebellar ataxia, resulting in lacks of muscle control and coordination of the upper and lower extremities, with symptoms including dysarthria, dysphagia, pyramidal signs, progressive external ophthalmoplegia, and dystonia (Bettencourt and Lima, 2011; Costa Mdo and Paulson, 2012; Wang et al., 2015; Paulson et al., 2017). This autosomal dominant disease is caused by an elongated polyglutamine stretch encoded by more than 51 CAG

repeats on the *ATXN3* allele (Rüb et al., 2013; Ding et al., 2016; Chen et al., 2017; Long et al., 2018). MJD is spread worldwide, but it has a higher relative frequency among the SCAs in Portugal (57.8%) (Vale et al., 2010), Brazil (59.6%) (de Castilhos et al., 2014), Japan (43%) (Takano et al., 1998), and Germany (42%) (Schöls et al., 1997). Our previous study has shown that mainland China has a high prevalence of MJD with a relative frequency of 62.6% (Chen et al., 2018).

This disease was first reported in an extended family of Portuguese-Azorean ancestry, in 1972 (Nakano et al., 1972), and was believed to be very frequent in Portugal due to founder effects. Also, Japan, Brazil and France all had founder effects of MJD (Marie-Françoise Chesselet, 2001). Furthermore, a linkage disequilibrium analysis was carried out with three single nucleotide polymorphisms (SNPs, A<sup>669</sup>TG/G<sup>669</sup>TG, C<sup>987</sup>GG/G<sup>987</sup>GG, and TAA<sup>1118</sup>/TAC<sup>1118</sup>) and five short tandem repeats (STRs, D14S1015, D14S995, D14S973, D14S1016, and D14S977), in 249 families of various ethnic backgrounds. Four different SNP haplotypes were identified segregating with MJD expansions: A-C-A, A-G-A, G-G-A, and G-G-C (Gaspar et al., 2001). Following this discovery, a worldwide study of extended haplotypes was performed in 264 MJD families, and two major ancestral lineages were confirmed: the GTGGCA background or the Machado lineage, probably originated in Portugal; and the TTACAC or Joseph lineage, observed in 19 countries, including Japan (Martins et al., 2007). To determine the occurrence of new mutation events and clarify the spread of ancestral MJD lineages in different populations, genetic distances were determined using a total of 20 SNPs and 4 microsatellites for the purpose of haplotypes identification (Martins et al., 2012), however, little information on MJD haplotypes is available in China, although the frequency of MJD is very high in this Asian population.

In this study, to determine the origin and estimate the age of MJD mutational event(s) in China, we performed a haplotype analysis of 20 SNPs and 7 microsatellites in 50 Chinese MJD families.

## MATERIALS AND METHODS

### Patients and Controls

We analyzed 50 families with MJD, from Southern China, including 109 patients, and 105 healthy individuals. All participants agreed with our request for collection of

peripheral blood samples and signed an informed consent form. The study was approved by Ethics Committee of Xiangya Hospital of Central South University in China. Genomic DNA was isolated using a standard protocol. Each family has at least two generations and has at least four individuals. Healthy spouses served as the control group. In one of the families studied, a patient homozygous for an MJD expansion was analyzed together with both parents and his sister.

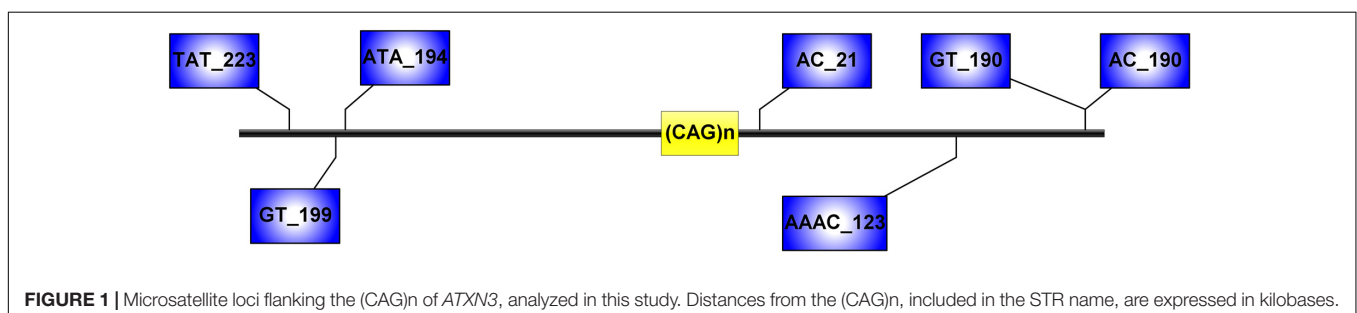
### Genotyping and Haplotype Analysis

We analyzed 20 SNPs located upstream and downstream the (CAG)<sub>n</sub>, including the previously analyzed six core SNPs and variants within a 4 kb area flanking the repeat (**Figure 1**; Martins et al., 2006; Costa et al., 2019). PCR amplification reactions were done with TSE 101 Golden Star T6 Super PCR Mix (1.1×). The location of all loci and Primers for amplification are listed in the **Supplementary Table S1**. Genotyping of SNPs was performed through Sanger sequencing. We inferred allelic phases by segregation for most of analyzed families; some haplotypes were reconstructed by PHASE software version 2.1.1<sup>1</sup> (only haplotypes with probabilities greater than 0.6 were used for subsequent analyses). We compared the distribution of SNPs in the Chinese MJD population with the two ancestral lineages described previously, and selected four different SNPs among these to test for Hardy-Weinberg equilibrium and Chi-square test ( $X^2$ ). In this test, 49 healthy spouses of probands served as controls. We first distinguished the haplotypes of the six core SNPs and conducted a Fisher's exact test. Basically, the formula  $\delta = (Fd-Fc) / (1-Fc)$  is used to calculate the approximate risk of population attribution and provide evidence of LD. Then, we extended the haplotype to 20 SNPs for further analysis (Devlin and Risch, 1995; Gaspar et al., 2001).

### STR-Based Haplotype Analysis and Age Estimation

Capillary electrophoresis was used to detect the seven STRs (TAT<sub>223</sub>)<sub>n</sub>, (GT<sub>199</sub>)<sub>n</sub>, (ATA<sub>194</sub>)<sub>n</sub>, (AC<sub>21</sub>)<sub>n</sub>, (AAAC<sub>123</sub>)<sub>n</sub>, (GT<sub>190</sub>)<sub>n</sub>, and (AC<sub>190</sub>)<sub>n</sub> (**Figure 1**), and haplotypes reconstructed by segregation and with the PHASE software version 2.1.1 (only haplotypes with probabilities greater than 0.6 were used for subsequent analyses). We determine

<sup>1</sup><http://stephenslab.uchicago.edu/phase/download.html>



**TABLE 1 |** SNP-based haplotypes of MJD Chinese families in comparison to Joseph MJD lineage.

| SNP                                      | refSNP ID   | Joseph lineage | Haplotypes of MJD Chinese families <sup>abc</sup> |            |           |           |           |           |            |           |           |           |           |           |           |   |
|--|-------------|----------------|---|------------|-----------|-----------|-----------|-----------|------------|-----------|-----------|-----------|-----------|-----------|-----------|---|
|  |             |                | A (n = 10)  | B (n = 12) | C (n = 1) | D (n = 8) | E (n = 1) | F (n = 1) | G (n = 11) | H (n = 1) | I (n = 1) | J (n = 1) | K (n = 2) | L (n = 1) | M (n = 1) |   |
| IVS6-30G > T                             | rs12590497  | T              | T   | T          | T         | T         | T         | T         | T          | T         | T         | T         | T         | T         | T         | T |
| GTT <sup>527</sup> /GTC <sup>527</sup>   | rs16999141  | T              | T   | T          | T         | T         | T         | T         | T          | T         | T         | T         | T         | T         | T         | T |
| IVS8-86T > G                             | rs10146519  | G              | G   | G          | G         | G         | G         | G         | G          | G         | G         | G         | G         | G         | G         | G |
| A <sup>669</sup> TG/G <sup>669</sup> TG  | rs1048755   | A              | A   | A          | A         | A         | A         | A         | A          | A         | A         | A         | A         | A         | A         | A |
| C/T.124                                  | rs12586535  | T              | T   | T          | T         | T         | T         | T         | T          | T         | T         | T         | T         | T         | T         | T |
| T/C.248                                  | rs12586471  | C              | C   | C          | C         | C         | C         | C         | C          | C         | C         | C         | C         | C         | C         | C |
| A/G.485                                  | rs56268847  | A              | A   | A          | A         | A         | A         | A         | A          | A         | A         | A         | A         | A         | A         | A |
| G/A.868                                  | rs10467858  | A              | A   | A          | A         | A         | A         | A         | A          | A         | A         | A         | A         | A         | A         | A |
| C/G.910                                  | rs10467857  | G              | G   | G          | G         | G         | G         | G         | G          | G         | G         | G         | G         | G         | G         | G |
| T/C.921                                  | rs10467856  | C              | C   | C          | C         | C         | C         | C         | C          | C         | C         | C         | C         | C         | C         | C |
| <b>(CAG)<sub>n</sub></b>                 |             |                |   |            |           |           |           |           |            |           |           |           |           |           |           |   |
| C <sup>987</sup> GG/G <sup>987</sup> GG  | rs12895357  | C              | C   | C          | C         | C         | C         | C         | C          | C         | C         | C         | C         | C         | C         | C |
| TAA <sup>1118</sup> /TAC <sup>1118</sup> | rs7158733   | A              | A   | A          | A         | A         | A         | A         | A          | A         | A         | A         | A         | A         | A         | A |
| C <sup>1178</sup> /A <sup>1178</sup>     | rs3092822   | C              | C   | C          | C         | C         | C         | C         | C          | C         | C         | C         | C         | C         | C         | C |
| C/T                                      | rs77086230  | C              | C   | C          | C         | C         | C         | C         | C          | C         | C         | C         | C         | C         | C         | C |
| C/T                                      | rs79316375  | C              | C   | C          | C         | C         | C         | C         | C          | C         | C         | C         | C         | C         | C         | C |
| G/A.3738                                 | rs8004149   | A              | A   | A          | A         | A         | A         | A         | A          | A         | A         | A         | A         | A         | A         | A |
| G/A.3770                                 | rs111735934 | G              | G   | G          | G         | G         | G         | G         | G          | G         | G         | G         | G         | G         | G         | G |
| C/T.3874                                 | rs181752420 | C              | C   | C          | C         | C         | C         | C         | C          | C         | C         | C         | C         | C         | C         | C |
| A/G.3912                                 | rs7142326   | G              | G   | G          | G         | G         | G         | G         | G          | G         | G         | G         | G         | G         | G         | G |
| C/T.3980                                 | rs74071847  | C              | C   | C          | C         | C         | C         | C         | C          | C         | C         | C         | C         | C         | C         | C |

SNP, single-nucleotide polymorphism.

<sup>a</sup>A to M indicate different haplotypes observed in Chinese MJD patients. n is the number of families. Haplotype A: families 1, 5, 8, 10, 11, 24, 29, 34, 44, 47; Haplotype B: families 4, 16, 23, 25, 30, 35, 37, 40, 41, 45, 48, 49; Haplotype C: family 15; Haplotype D: families 1, 3, 7, 9#, 19, 26, 38#, 50; Haplotype E: family 31; Haplotype F: family 14; Haplotype G: families 6, 13, 17, 22, 27, 32, 33#, 36, 42, 43, 46; Haplotype H: family 18; Haplotype I: family 12; Haplotype J: family 2; Haplotype K: families 20, 39#; Haplotype L: family 21; Haplotype M: family 28# (# indicate families with some alleles inferred by PHASE).

<sup>b</sup>In family 1, both parents are affected and the patient harbours two different MJD lineages.

<sup>c</sup>Bold indicates different SNP alleles when compared to the Joseph lineage.

ancestral haplotypes and steps of mutation by STR-based haplotype from the most common haplotype based on 20 SNPs and assess the age of mutation events. This formula  $\epsilon = 1 - [(1-c)(1-\mu)]$  is used to calculate the probability of change for each generation, where  $c$  represents the recombination rate and  $\mu$  represents the mutation rate. The multiplication of  $\epsilon$  and  $t$  yields the average of mutations and reorganizations ( $\lambda$ ), where  $t$  is the number of generations (Martins et al., 2007).

## RESULTS

### SNP Haplotyping of Chinese MJD Families

We identified 13 disease-associated haplotypes in our cohort of Chinese MJD families. We compared these haplotypes with the Joseph lineage and indicated their differences (Table 1). It is worthy of mentioning that, for four SNPs, we found new alleles segregating with expanded alleles, not previously associated to Machado or Joseph ancestral MJD lineages (Table 2). The G allele occurs in the rs56268847 of the pathogenic chromosomes in the 28 families, exceeding 50% of the total fifty families involved in the study. Regarding SNPs rs16999141, rs10467857 and rs77086230, alleles never found in phase with expanded MJD alleles were here observed in families 21, 31 and 15, respectively.

### Hardy-Weinberg Equilibrium and Chi-Square Test for the 4 of SNPs

Hardy-Weinberg equilibrium and chi-square test were carried out to examine the SNP variations. The control group consisting of 49 spouses confirmed by the Hardy-Weinberg equilibrium

that satisfy the genetic equilibrium in the four heterogeneous SNPs. In rs16999141, no genotype CC was found in MJD patients, 57 of 109 patients shared TT, and the rest were TC. For SNP rs56268847, significant variation from AA, AG to GG was detected with the  $p$ -value of the Hardy-Weinberg equilibrium at this position is 0.007. More specifically, 46 out of 109 SCA3/MJD patients had a genotype of AA, while 60 patients had AG, and by contrast, only three patients had GG. Different distribution for SNP rs10467857 was detected. Of the 109 identical patients, 56 are genotyped as GG and 53 as GC ( $p$ -value of the Hardy-Weinberg equilibrium 0.004), respectively. Similar situation was found in the SNP rs77086230 in a way that Only CC and CT were discovered with a predominance of CC over CT. The  $p$ -value of Hardy-Weinberg equilibrium of this SNP in SCA3/MJD group and three SNPs in the control group are more significant than 0.05. Moreover, all  $p$ -values used chi-square tests for three SNPs are less than 0.001 (Table 2).

### Analysis of Haplotypes Based on Six SNPs

Through linkage analysis of six SNPs such as IVS6-30G > T, GTT<sup>527</sup>/GTC<sup>527</sup>, A<sup>669</sup>TG/G<sup>669</sup>TG, C<sup>987</sup>GG/G<sup>987</sup>GG, TAA<sup>1118</sup>/TAC<sup>1118</sup>, and C<sup>1178</sup>/A<sup>1178</sup>, we found that Chinese MJD patients only shared 7 haplotypes, while healthy people shared 21 haplotypes (Table 3). Among them, T-T-A-G-C-C only appears in the MJD patients. The remaining 6 haplotypes were confirmed statistically significant by chi-square test, and all the  $p$ -values were less than 0.001.

### Age Estimation

We performed STR analysis and age estimation on four major haplotypes A, B, D, and G based on 20 SNPs, and inferred the Phylogenetic networks of the four haplotypes (Figure 2).

**TABLE 2** | Allele frequency of 4 different SNPs analyzed by Hardy-Weinberg equilibrium and chi-square test.

| refSNP ID  | Group                | N <sup>a</sup> | Genotype <sup>b</sup> |    |    | P-value <sup>c</sup> | Allele <sup>d</sup> |           | P-value <sup>e</sup> |
|------------|----------------------|----------------|-----------------------|----|----|----------------------|---------------------|-----------|----------------------|
| rs16999141 | MJD                  | 109            | TT                    | TC | CC | 0.005                | T                   | C         | <0.001               |
|            | Control <sup>f</sup> | 49             | 12                    | 29 | 8  | 0.407                | 53 (0.54)           | 45 (0.46) |                      |
| rs56268847 | MJD                  | 109            | AA                    | AG | GG | 0.007                | A                   | G         | <0.001               |
|            | Control <sup>f</sup> | 49             | 43                    | 5  | 1  | 0.271                | 91 (0.93)           | 7 (0.07)  |                      |
| rs10467857 | MJD                  | 109            | GG                    | GC | CC | 0.004                | G                   | C         | <0.001               |
|            | Control <sup>f</sup> | 49             | 13                    | 26 | 10 | 0.901                | 52 (0.53)           | 46 (0.47) |                      |
| rs77086230 | MJD                  | 109            | CC                    | CT | TT | 0.957                | C                   | T         | <0.001               |
|            | Control <sup>f</sup> | 49             | 44                    | 5  | 0  | 0.932                | 93 (0.95)           | 5 (0.05)  |                      |

MJD, Machado-Joseph disease; SNP, single-nucleotide polymorphism.

<sup>a</sup>The number of each group.

<sup>b</sup>Value of the Hardy-Weinberg equilibrium for each comparison (MJD vs. controls).

<sup>c</sup>The  $P$ -value of Hardy-Weinberg equilibrium.

<sup>d</sup>Value of the Hardy-Weinberg equilibrium for each comparison (MJD vs. controls).

<sup>e</sup>The  $P$ -value of chi-square test.

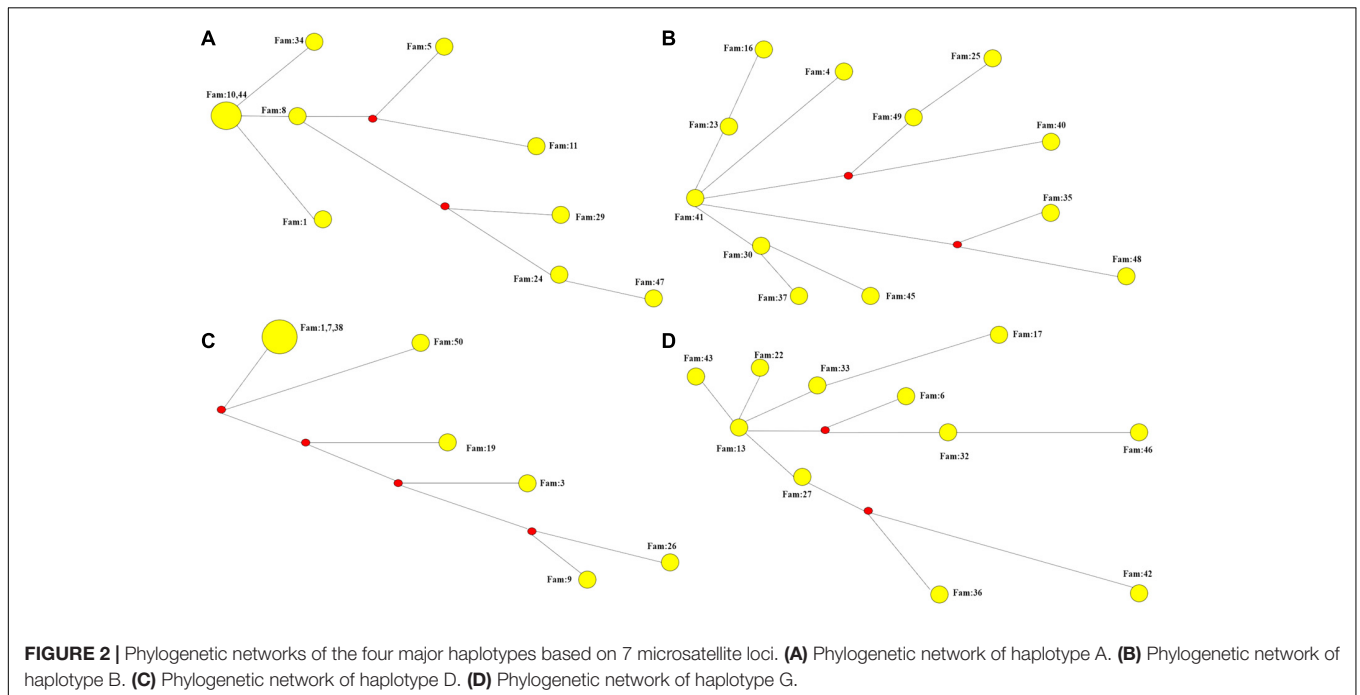
<sup>f</sup>49 healthy spouses of probands of 50 Chinese families with MJD served as controls.

**TABLE 3** | Overall linkage disequilibrium analysis of Chinese families for intragenic haplotypes based on 6 SNPs.

| Haplotype <sup>a</sup> | Frequency in     |                   | $\delta^b$ | P-value |
|------------------------|------------------|-------------------|------------|---------|
|                        | Control subjects | Subjects with MJD |            |         |
| T-T-A-C-A-C            | 0.283            | 0.482             | 0.278      | <0.001  |
| T-T-A-C-C-A            | 0.013            | 0                 | ...        | ...     |
| T-T-A-G-A-C            | 0.119            | 0.382             | 0.298      | <0.001  |
| T-T-A-G-C-C            | 0.000            | 0.045             | 0.045      | ...     |
| T-T-A-G-C-A            | 0.003            | 0                 | ...        | ...     |
| T-T-G-G-A-C            | 0.013            | 0.009             | 0.003      | <0.001  |
| T-T-G-G-C-A            | 0.013            | 0                 | ...        | ...     |
| T-T-G-C-A-C            | 0.006            | 0.018             | 0.012      | <0.001  |
| T-C-G-G-C-A            | 0.019            | 0                 | ...        | ...     |
| G-C-G-G-C-A            | 0.321            | 0                 | ...        | ...     |
| G-C-G-G-A-C            | 0.006            | 0                 | ...        | ...     |
| G-C-G-C-C-A            | 0.035            | 0                 | ...        | ...     |
| G-C-G-C-A-C            | 0.013            | 0                 | ...        | ...     |
| G-C-A-G-C-A            | 0.028            | 0                 | ...        | ...     |
| G-C-A-G-A-C            | 0.009            | 0.009             | 0          | <0.001  |
| G-T-A-G-A-C            | 0.016            | 0.055             | 0.040      | <0.001  |
| G-T-A-G-C-A            | 0.003            | 0                 | ...        | ...     |
| G-T-A-C-A-C            | 0.009            | 0                 | ...        | ...     |
| G-T-A-C-C-A            | 0.009            | 0                 | ...        | ...     |
| G-T-G-G-A-C            | 0.003            | 0                 | ...        | ...     |
| G-T-G-G-C-A            | 0.066            | 0                 | ...        | ...     |
| G-T-G-C-C-A            | 0.013            | 0                 | ...        | ...     |

<sup>a</sup>Haplotype based on IVS6-30G > T, GTT<sup>527</sup>/GTC<sup>527</sup>, A<sup>669</sup>TG/G<sup>669</sup>TG, C<sup>987</sup>GG/G<sup>987</sup>GG, TAA<sup>1118</sup>/TAC<sup>1118</sup>, C<sup>1178</sup>/A<sup>1178</sup>.

<sup>b</sup>The approximate risk of population attribution.



**FIGURE 2** | Phylogenetic networks of the four major haplotypes based on 7 microsatellite loci. **(A)** Phylogenetic network of haplotype A. **(B)** Phylogenetic network of haplotype B. **(C)** Phylogenetic network of haplotype D. **(D)** Phylogenetic network of haplotype G.

Surprisingly, MJD haplotypes A and D seemed to be present in the Chinese population as remote as  $16,335 \pm 1,966$  and  $11,837 \pm 1,871$  years, respectively; whereas introduction of

haplotypes B and G probably occurred simultaneously, a few million years later,  $9,272 \pm 1,352$  and  $9,254 \pm 1,411$  years ago, respectively (Table 4).

**TABLE 4 |** Haplotypes and age estimation with 7 STR flanking the (CAG)<sub>n</sub> at the *ATXN3*, from families sharing the four most common SNP-based haplotypes.

| SNP-based haplotype  | Family number  | (TAT_223) <sub>n</sub> | (GT_199) <sub>n</sub> | (ATA_194) <sub>n</sub> | (AC_21) <sub>n</sub> | (AAAC) <sub>n</sub> | (GT) <sub>n</sub> | (AC_190) <sub>n</sub> | Age <sup>a</sup> |
|--|--|------------------------|-----------------------|------------------------|----------------------|---------------------|-------------------|-----------------------|------------------|
| Haplotype<br>A:TTGATCAAGC-(CAG) <sub>exp</sub> -<br>CACCCAGCGC | Fam 10   | 10                     | 20                    | 10                     | 13                   | 7                   | 16                | 15                    | 16,335 ± 1,966   |
|  | Fam 44   | 10                     | 20                    | 10                     | 13                   | 7                   | 16                | 15                    |                  |
|  | Fam 34   | 11                     | 20                    | 10                     | 13                   | 7                   | 15                | 15                    |                  |
|  | Fam 1  | 10                     | 20                    | 10                     | 13                   | 7                   | 20                | 18                    |                  |
|  | Fam 8  | 10                     | 22                    | 10                     | 13                   | 7                   | 16                | 15                    |                  |
|  | Fam 5  | 10                     | 26                    | 10                     | 13                   | 7                   | 16                | 15                    |                  |
|  | Fam 11   | 10                     | 24                    | 10                     | 12                   | 7                   | 14                | 18                    |                  |
|  | Fam 24   | 14                     | 22                    | 10                     | 12                   | 8                   | 20                | 15                    |                  |
|  | Fam 29   | 16                     | 22                    | 10                     | 13                   | 7                   | 20                | 15                    |                  |
|  | Fam 47   | 16                     | 24                    | 10                     | 14                   | 5                   | 20                | 16                    |                  |
| Haplotype<br>B:TTGATCGAGC-<br>(CAG) <sub>exp</sub> -CACCCAGCGC | Fam 41   | 10                     | 22                    | 10                     | 12                   | 7                   | 20                | 16                    | 9,272 ± 1,352    |
|  | Fam 23   | 10                     | 22                    | 10                     | 13                   | 7                   | 20                | 16                    |                  |
|  | Fam 16   | 10                     | 21                    | 11                     | 13                   | 7                   | 20                | 16                    |                  |
|  | Fam 4  | 11                     | 24                    | 10                     | 12                   | 7                   | 20                | 15                    |                  |
|  | Fam 49   | 10                     | 22                    | 10                     | 12                   | 7                   | 16                | 16                    |                  |
|  | Fam 25   | 10                     | 22                    | 10                     | 14                   | 7                   | 16                | 16                    |                  |
|  | Fam 40   | 10                     | 22                    | 9                      | 12                   | 7                   | 17                | 19                    |                  |
|  | Fam 30   | 10                     | 22                    | 9                      | 12                   | 7                   | 20                | 16                    |                  |
|  | Fam 37   | 10                     | 22                    | 9                      | 12                   | 8                   | 20                | 16                    |                  |
|  | Fam 45   | 10                     | 20                    | 9                      | 12                   | 7                   | 20                | 16                    |                  |
|  | Fam 35   | 15                     | 22                    | 8                      | 12                   | 7                   | 20                | 16                    |                  |
|  | Fam 48   | 16                     | 22                    | 9                      | 12                   | 6                   | 20                | 15                    |                  |
|  | Haplotype<br>D:TTGATCAAGC-<br>(CAG) <sub>exp</sub> -GACCCAGCGC | Fam 1                  | 10                    | 20                     | 10                   | 13                  | 7                 | 16                    |                  |
| Fam 7  |  | 10                     | 20                    | 10                     | 13                   | 7                   | 16                | 15                    |                  |
| Fam 38   |  | 10                     | 20                    | 10                     | 13                   | 7                   | 16                | 15                    |                  |
| Fam 50   |  | 11                     | 19                    | 12                     | 13                   | 7                   | 15                | 15                    |                  |
| Fam 19   |  | 13                     | 22                    | 9                      | 13                   | 7                   | 16                | 15                    |                  |
| Fam 3  |  | 16                     | 23                    | 10                     | 13                   | 7                   | 16                | 15                    |                  |
| Fam 9  |  | 17                     | 21                    | 10                     | 13                   | 7                   | 17                | 15                    |                  |
| Fam 26   |  | 16                     | 21                    | 10                     | 13                   | 7                   | 20                | 15                    |                  |
| Haplotype<br>G:TTGATCGAGC-<br>(CAG) <sub>exp</sub> -GACCCAGCGC |  | Fam 13                 | 10                    | 22                     | 9                    | 12                  | 7                 | 20                    | 16               |
|  | Fam 43   | 10                     | 22                    | 8                      | 12                   | 7                   | 20                | 16                    |                  |
|  | Fam 22   | 10                     | 22                    | 9                      | 12                   | 7                   | 20                | 15                    |                  |
|  | Fam 33   | 10                     | 22                    | 9                      | 15                   | 7 (5)               | 19                | 16                    |                  |
|  | Fam 17   | 10                     | 22                    | 9                      | 14                   | 8                   | 16                | 16                    |                  |
|  | Fam 6  | 12                     | 22                    | 10                     | 12                   | 7                   | 20                | 16                    |                  |
|  | Fam 32   | 16                     | 22                    | 9                      | 12                   | 7                   | 20                | 16                    |                  |
|  | Fam 46   | 16                     | 21                    | 10                     | 12                   | 7                   | 20                | 15                    |                  |
|  | Fam 27   | 10                     | 22                    | 9                      | 12                   | 5                   | 20                | 16                    |                  |
|  | Fam 36   | 10                     | 22                    | 10                     | 12                   | 5                   | 20                | 15                    |                  |
|  | Fam 42   | 11                     | 22                    | 11                     | 12                   | 5                   | 20                | 20                    |                  |

<sup>a</sup>Age is the estimated age of occurrence for each haplotype based on 20 SNP.

## DISCUSSION

For the selection of participants, we require that the proband must have more than one children

or proband's family must bear more than two generations. The stringency of the selection criteria for the participants made only 5 of the families inferred as haplotypes by software, suggesting the vast



majority of families inferred haplotypes through the pedigree structure to make our results more accurate and unquestionable.

The rs56268847 found in Asian SCA3/MJD patients was not statistically significant in the previous report (Martins et al., 2012), most probably due to the small sample size and different national backgrounds. Pathogenic chromosomes of 28 probands in 50 families carry G allele with a frequency of 0.56. The *p*-value of the Hardy-Weinberg equilibrium is  $<0.05$  in the SCA3/MJD group, indicating that this allele does not reach a genetic equilibrium. However, the difference between the control group confirmed to achieve the genetic balance after the same test and Chi-square test, and the SCA3/MJD group is statistically significant. Thus, it is reasonable to speculate that the difference might be primarily caused by the disease. Furthermore, since the G allele at rs56268847 has been observed to segregate with the expanded allele in more than half of our Chinese MJD families, this suggests that a point mutation at this SNP must have occurred early in the introduction of MJD in China or, even, in other Asian countries.

Additionally, for the first time we detected base variations (new base addition) in three SNPs rs16999141, rs10467857, and rs77086230 in the Chinese SCA3/MJD patients compared to the previously reported two ancestral lineages. Both ancestral lineages are reported as T on rs16999141, and we detected T/C. While G is reported in the ancestral lineage SNP rs10467857, we detected G/C instead of G only. Similarly, C/T is found in the Chinese SCA3/MJD patients other than C only as reported in the ancestral lineage SNP rs77086230. As for the significance in our new discovery of the new base variation in 3 sites, although so far only one family were examined for each SNP, two test methods have confirmed the significant difference. To be more conclusive to say that the new base variations occurred relatively recently, tests of more families are required.

Previous study showed that the A-G-A haplotype for the SNPs rs1048755, rs12895357, and rs7158733 in the 249 of families, exist in MJD patient (Gaspar et al., 2001), while they were not statistically significant most probably due to the different backgrounds. More convincingly, the *p*-value obtained by Fisher's exact test is less than 0.001. Thus, we believe the significant difference in the frequency of A-G-A haplotype between the SCA3/MJD and the control group.

The first studies on the epidemiology of MJD were based on families described before the gene ATXN3 was known (Sequeiros and Coutinho, 1993). Later, two major MJD haplotypes have been identified (Gaspar et al., 2001; Martins et al., 2007); here, in addition to the worldwide spread Joseph lineage, we found three other major haplotypes that differ from Joseph at (1) rs56268847 (lineage B, 12 families), (2) rs12895357 (lineage D, 8 families), and (3) both rs56268847 and rs12895357 (lineage G, 11 families). Interestingly, lineage B has been previously described among

Australian aborigines, probably introduced in this population via Asia (Martins et al., 2012). Independent mutational origins do not necessarily underlie these four MJD SNP-haplotypes since recurrent mutations on the 2 SNPs may have occurred: backgrounds B and D may have evolved from the Joseph lineage by recurrence at rs56268847 and rs12895357, respectively; recombination could be a possible explanation for the origin of SNP backgrounds G, although the short distance between the two SNPs and the deleterious (CAG)<sub>n</sub> makes it unlikely. Other MJD backgrounds, phylogenetically more distant from the Joseph lineage, were found in single MJD families; de novo expansions may be on the origin of some of them, but a larger cohort should confirm that their low frequency is explained by a recent event or genetic drift.

This discovery is of importance to clarify the prevalence of SCA3/MJD and could become indispensable evidence that supports the founder effect in this disease. More importantly, this finding could help decipher the genetic basis of the SCA3/MJD by further study on the haplotypes.

## INFORMED CONSENT

Informed consent was obtained from all individual participants included in the study.

## AUTHOR CONTRIBUTIONS

TL completed the collection of samples, analysis of data, and writing of the manuscript. SM, JS, ZH, KX, BT, and HJ completed the design of the experiments. YP, PW, XH, ZC, CW, and ZT assisted in the collection of samples. RQ and CC contributed to the analysis of the data.

## FUNDING

This work was supported by the National Key Research and Development Program of China (Nos. 2016YFC0905100 and 2016YFC0901504 to HJ), the National Natural Science Foundation of China (Nos. 81771231 and 81471156 to HJ), the Clinical Research Funds of Xiangya Hospital (2014L03 to HJ), the Clinical and rehabilitation fund of Peking University Weiming Biotech Group (No. xywm2015110 to HJ), Youth Foundation of Xiangya Hospital (No. 2017Q03 to ZC), and Independent Exploration and Innovation Project of Graduate Students of Central South University (No. 1053320170177 to TL). Moreover, SM is funded by the project IF/00930/2013 from FCT. This article is a result of the project NORTE-01-0145-FEDER-000008, supported by Norte Portugal Regional Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

## ACKNOWLEDGMENTS

We sincerely appreciate all patients and their families that participated in the study voluntarily and supported this study. Moreover, we hope to thank all those who gave this study help.

## REFERENCES

- Bettencourt, C., and Lima, M. (2011). Machado-joseph disease: from first descriptions to new perspectives. *Orphanet J. Rare Dis.* 6:35. doi: 10.1186/1750-1172-6-35
- Chen, Z., Wang, C., Zheng, C., Long, Z., Cao, L., Li, X., et al. (2017). Ubiquitin-related network underlain by (CAG)<sub>n</sub> loci modulate age at onset in Machado-Joseph disease. *Brain* 140:e25. doi: 10.1093/brain/awx028
- Chen, Z., Wang, P., Wang, C., Peng, Y., Hou, X., Zhou, X., et al. (2018). Updated frequency analysis of spinocerebellar ataxia in China. *Brain* 141:e22. doi: 10.1093/brain/awy016
- Costa, I. P. D., Almeida, B. C., Sequeiros, J., Amorim, A., and Martins, S. (2019). A pipeline to assess disease-associated haplotypes in repeat expansion disorders: the example of MJD/SCA3 Locus. *Front. Genet.* 10:38. doi: 10.3389/fgene.2019.00038
- Costa Mdo, C., and Paulson, H. L. (2012). Toward understanding machado-joseph disease. *Prog. Neurobiol.* 97, 239–257. doi: 10.1016/j.pneurobio.2011.11.006
- de Castilhos, R. M., Furtado, G. V., Gheno, T. C., Schaeffer, P., Russo, A., Barsottini, O., et al. (2014). Spinocerebellar ataxias in Brazil—frequencies and modulating effects of related genes. *Cerebellum* 13, 17–28. doi: 10.1007/s12311-013-0510-y
- Devlin, B., and Risch, N. (1995). A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29, 311–322. doi: 10.1006/geno.1995.9003
- Ding, D., Li, K., Wang, C., Chen, Z., Long, Z., Peng, Y., et al. (2016). ATXN2 polymorphism modulates age at onset in Machado-Joseph disease. *Brain* 139:e59. doi: 10.1093/brain/aww176
- Gaspar, C., Lopes-Cendes, I., Hayes, S., Goto, J., Arvidsson, K., Dias, A., et al. (2001). Ancestral origins of the Machado-Joseph disease mutation: a worldwide haplotype study. *Am. J. Hum. Genet.* 68, 523–528. doi: 10.1086/318184
- Long, Z., Li, T., Chen, Z., Peng, Y., Wang, C., Hou, X., et al. (2018). Cerebellar lncRNA expression profile analysis of SCA3/MJD Mice. *Int. J. Genom.* 2018, 1–6. doi: 10.1155/2018/5383517
- Marie-Françoise Chesselet, M. D. (2001). *Molecular Mechanisms of Neurodegenerative Diseases*. New York, NY: Humana Press.
- Martins, S., Calafell, F., Gaspar, C., Wong, V. C., Silveira, I., Nicholson, G. A., et al. (2007). Asian origin for the worldwide-spread mutational event in Machado-Joseph disease. *Arch. Neurol.* 64, 1502–1508. doi: 10.1001/archneur.64.10.1502
- Martins, S., Calafell, F., Wong, V. C., Sequeiros, J., and Amorim, A. (2006). A multistep mutation mechanism drives the evolution of the CAG repeat at MJD/SCA3 locus. *Eur. J. Hum. Genet.* 14, 932–940. doi: 10.1038/sj.ejhg.5201643
- Martins, S., Soong, B. W., Wong, V. C., Giunti, P., Stevanin, G., Ranum, L. P., et al. (2012). Mutational origin of machado-joseph disease in the australian aboriginal communities of groote eylandt and yirrkala. *Arch. Neurol.* 69, 746–751. doi: 10.1001/archneurol.2011.2504
- Nakano, K. K., Dawson, D. M., and Spence, A. (1972). Machado disease: a hereditary ataxia in portuguese emigrants to massachusetts. *Neurology* 22, 49–49. doi: 10.1212/wnl.22.1.49
- Paulson, H. L., Shakkottai, V. G., Clark, H. B., and Orr, H. T. (2017). Polyglutamine spinocerebellar ataxias - from genes to potential treatments. *Nat. Rev. Neurosci.* 18, 613–626. doi: 10.1038/nrn.2017.92
- Rüb, U., Schöls, L., Paulson, H., Auburger, G., Kermer, P., Jen, J. C., et al. (2013). Clinical features, neurogenetics and neuropathology of the polyglutamine spinocerebellar ataxias type 1, 2, 3, 6 and 7. *Prog. Neurobiol.* 104, 38–66. doi: 10.1016/j.pneurobio.2013.01.001
- Schöls, L., Amoiridis, G., Büttner, T., Przuntek, H., Epplen, J. T., and Riess, O. (1997). Autosomal dominant cerebellar ataxia: phenotypic differences in genetically defined subtypes? *Ann. Neurol.* 42, 924–932. doi: 10.1002/ana.410420615
- Sequeiros, J., and Coutinho, P. (1993). Epidemiology and clinical aspects of Machado-Joseph disease. *Adv. Neurol.* 61, 139–153.
- Takano, H., Cancel, G., Ikeuchi, T., Lorenzetti, D., Mawad, R., and Stevanin, G. (1998). 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.* 63, 1060–1066. doi: 10.1086/302067
- Vale, J., Bugalho, P., Silveira, I., Sequeiros, J., Guimaraes, J., and Coutinho, P. (2010). Autosomal dominant cerebellar ataxia: frequency analysis and clinical characterization of 45 families from portugal. *Eur. J. Neurol.* 17, 124–128. doi: 10.1111/j.1468-1331.2009.02757.x
- Wang, C., Chen, Z., Yang, F., Jiao, B., Peng, H., Shi, Y., et al. (2015). Analysis of the GGGGCC repeat expansions of the C9orf72 Gene in SCA3/MJD Patients from China. *PLoS One* 10:e0130336. doi: 10.1371/journal.pone.0130336
- Younger, D. S. (1999). *Motor Disorders*. Philadelphia: Lippincott Williams & Wilkins.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00740/full#supplementary-material>

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Li, Martins, Peng, Wang, Hou, Chen, Wang, Tang, Qiu, Chen, Hu, Xia, Tang, Sequeiros and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.