

DOI: 10.3724/SP.J.1005.2011.00353

## 不同群体中 *ATXN2* 基因编码区 CAG 重复的变异研究

陈晓晨, 孙浩, 米冬青, 黄小琴, 林克勤, 易文, 于亮, 史磊, 史荔, 杨昭庆, 褚嘉祐

中国医学科学院北京协和医学院医学生物学研究所遗传室, 昆明 650118

**摘要:** 在中国 6 个生活环境差异较大的少数民族群体中进行 *ATXN2* 基因编码区 CAG 重复的变异研究, 以衡量其是否受到正选择的作用以及寻找推动选择作用的因素。采集 6 个民族群体共 291 个健康无关个体, 对其进行 STR 分型, 直接计数其等位基因及等位基因型频率, 计算其线性 *Fst* 值, 构建针对该基因的系统进化树, 并对各群体进行 MDS 分析。线性 *Fst* 值结果显示: 回族和彝族群体间 *ATXN2* 基因 STR 位点进化的差异具有显著性, 其他 4 个群体相互间无显著性差异。结合已报道的其他群体进一步分析, 回族、哈尼族、云南蒙古族以及内蒙古自治区蒙古族每个个人群都与日本人群有显著性差异; 回族、内蒙古自治区蒙古族与汉族具有显著性差异。6 个群体中 *ATXN2* 基因 STR 的等位基因频率有各自的分布特点, 稀有等位基因频率变化产生的原因可能是选择作用的结果。

**关键词:** *ATXN2* 基因; 民族群体; 自然选择; 基因多态性

## Variation of CAG repeats in coding region of *ATXN2* gene in different ethnic groups

CHEN Xiao-Chen, SUN Hao, MI Dong-Qing, HUANG Xiao-Qin, LIN Ke-Qin, YI Wen, YU Liang, SHI Lei, SHI Li, YANG Zhao-Qing, CHU Jia-You

Department of Medical Genetics, Institute of Medical Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Kunming 650118, China

**Abstract:** To investigate CAG repeats variation of *ATXN2* gene coding region in six ethnic groups that live in comparatively different environments, to evaluate whether these variations are under positive selection, and to find factors driving selection effects, 291 unrelated healthy individuals were collected from six ethnic groups and their STR genotyping was performed. The frequencies of alleles and genotypes were counted and thereby Slatkin's linearized *Fst* values were calculated. The UPGMA tree against this gene was constructed. The MDS analysis among these groups was carried out as well. The results from the linearized *Fst* values indicated that there were significant evolutionary differences of the STR in *ATXN2* gene between Hui and Yi groups, but not among the other 4 groups. Further analysis was performed by combining our data with published data obtained from other groups. These results indicated that there were significant differences between Japanese and other groups including Hui, Hani, Yunnan Mongolian, and Inner Mongolian. Both Hui and Mongolian from Inner Mongolia were significantly different from Han. In conclusion, the six ethnic groups had their own distribution characterizations of allelic frequencies of *ATXN2* STR, and the potential cause of frequency changes in rare alleles could be the consequence of positive selection.

**Keywords:** *ATXN2* gene; ethnic groups; natural selection; gene polymorphisms

收稿日期: 2010-08-10; 修回日期: 2010-10-19

基金项目: 国家自然科学基金项目(编号: 30971578)和云南省自然科学基金项目(编号: 2008ZC068M)资助

作者简介: 陈晓晨, 在读博士研究生, 研究方向: 医学遗传学。E-mail: fireflyxchen@imbcams.com.cn

通讯作者: 褚嘉祐, 研究员, 博士生导师, 研究方向: 医学遗传学。E-mail: chujy@imbcams.com.cn

*ATXN2* 基因定位于 12q24, 其编码区 CAG 三核苷酸重复序列在正常人中的重复次数一般为 14~31 次<sup>[1-6]</sup>, 而 CAG 的异常扩增引起编码蛋白中多聚谷氨酰胺堆积, 将导致脊髓小脑共济失调 2 型<sup>[7]</sup>的发生, 患者的 CAG 重复一般在 35~56 次左右。脊髓小脑共济失调 2 型(Spinocerebellar ataxia type 2, SCA2)是一种进行性神经系统退行性疾病, 为常染色体显性遗传, 临床上以进行性小脑性共济失调为主要症状, 此外还常常伴有帕金森综合征、肌张力障碍等, 严重影响正常的感觉、运动控制功能。有研究表明, *ATXN2* 基因编码区 CAG 重复在现代欧洲人群体共 55 人中受到近期的正选择作用, 而在北京汉族和东京日本人群体中(分别为 43 人和 45 人)没有显示正选择的信号<sup>[1]</sup>。这些会产生不利的遗传性疾病的危险突变仍然受到正选择作用被保留下来, 而不是受到负选择作用被淘汰掉<sup>[8]</sup>。这一现象可能与群体的地理位置、生活环境等驱动因素有关而使其保留下来。本研究选取中国 6 个地理位置、生活环境等不同的少数民族群体进行 *ATXN2* 基因编码区 CAG 重复的变异研究, 对其进行 *ATXN2* 基因 STR 的分型, 同时运用本研究和相关文献数据计算这 6 个民族群体和其他人群的线性 *F<sub>st</sub>* 值及其显著性差异, 对这些群体是否受到自然选择作用以及该基因的进化等进行分析和探讨。

## 1 对象和方法

### 1.1 研究对象

依据“知情同意”原则, 在民族聚居地分别采集 6 个民族群体共 291 个健康无关个体, 其中宁夏回族 50 人, 云南哈尼族 44 人, 云南通海蒙古族 48 人, 内蒙古自治区蒙古族 52 人, 云南丽江彝族 47 人以及云南新平傣族 50 人。采集静脉血样本每人约 5 mL, EDTA 或肝素钠抗凝, AXYGEN 公司血液基因组试剂盒提取 DNA, 测定样本基因组 DNA 的浓度, 并稀释至大约 10 ng/ $\mu$ L, -20 $^{\circ}$ C 保存待用。

### 1.2 方法

#### 1.2.1 *ATXN2* 基因 STR 的遗传检测

本研究选取 *ATXN2* 基因第一外显子中的 STR 进行研究。采用 PCR 方法对该 STR 进行扩增。引物<sup>[2]</sup>序列如下:

SCA-A: 5'-GGGCCCTCACCATGTCG-3'(5'端

用 FAM 标记),

SCA-B: 5'-CGGGCTTGCGGACATTGG-3'。

PCR 反应条件: 94 $^{\circ}$ C 7 min; 94 $^{\circ}$ C 30 s, 62 $^{\circ}$ C 30 s, 72 $^{\circ}$ C 30 s, 共 30 个循环; 72 $^{\circ}$ C 10 min。PCR 扩增产物在 2% 琼脂糖凝胶电泳检测, 溴化乙啶染色, 在紫外凝胶电泳成像分析系统上观察结果。随后将 PCR 产物稀释 200 倍, 并与基因分型分子内标 GS500LIZ 充分混匀, 在 ABI3130 测序仪上进行基因分型, 数据分析使用 Genemapper V4.0 软件。

#### 1.2.2 统计学方法

群体中等位基因及等位基因型频率采用直接计数法; 应用 Arlequin 软件<sup>[9]</sup>进行统计学分析, 对各群体进行 Hardy-Weinberg 平衡检验 (Hardy-Weinberg Equilibrium, HWE) 和 Ewens-Watterson 中性检验 (Ewens-Watterson Neutrality Test, EW); 群体间的比较用线性 *F<sub>st</sub>* 值<sup>[10]</sup>表示, 并通过置换的方法计算其显著性差异 *P* 值; 针对 *ATXN2* 基因 STR 的系统进化树通过 MEGA 软件用 UPGMA 法构建; 应用 SPSS17.0 软件对各群体进行 MDS 多维尺度分析。

## 2 结果与分析

### 2.1 群体中等位基因频率分布

对本研究中的 6 个民族群体样本进行 STR 基因分型, 分型结果采用直接计数法进行统计。在所有群体中共检出等位基因 14 个, 等位基因型 16 个, 其中等位基因频率最高的是 22 次重复 (91%~97%), 等位基因多态性分布范围较广 (14~31 次重复), 结果见表 1。根据这 6 个群体基因型分布做 Hardy-Weinberg 平衡检验, 结果显示: 这 6 个群体均符合 Hardy-Weinberg 平衡 ( $P>0.05$ ), 表明其均处于平衡状态。

### 2.2 不同群体间 STR 多态性的比较

应用 Arlequin 软件<sup>[9]</sup>对群体间该 STR 位点多态性分布进行比较分析, 其线性 *F<sub>st</sub>* 值及其显著性差异 (表 2) 结果显示: 回族和彝族群体间该 STR 位点进化的差异具有显著性, 其他 4 个群体相互间无显著性差异。结合已报道的其他群体<sup>[1]</sup>进一步分析, 回族与汉族、日本人有显著性差异; 哈尼族、云南蒙古族 2 个群体均与日本人有显著性差异; 内蒙古自治区蒙古族与汉族、日本人具有显著性差异。

表 1 不同群体中 *ATXN2* 基因 STR 等位基因频率

CAG 重复 次数	不同群体中的等位基因频率							
	回族 2n=100	哈尼族 2n=88	蒙古族(云南) 2n=96	蒙古族(内蒙) 2n=104	彝族 2n=94	傣族 2n=100	汉族 <sup>[1]</sup> 2n=86	日本人 <sup>[1]</sup> 2n=90
14		3(0.0341)						
18				1(0.0096)				
19	6(0.0600)	1(0.0114)				2(0.0200)		
20			2(0.0208)					
21	1(0.0100)			1(0.0096)				
22	91(0.9100)	82(0.9318)	90(0.9375)	97(0.9327)	91(0.9681)	97(0.9700)	86(1.0000)	90(1.0000)
23				2(0.0192)	1(0.0106)			
24			1(0.0104)					
25				3(0.0288)	1(0.0106)			
26		2(0.0227)						
27			2(0.0208)		1(0.0106)	1(0.0100)		
29			1(0.0104)					
30	1(0.0100)							
31	1(0.0100)							

注: 表格中的数字表示 *ATXN2* 基因中不同的 CAG 重复次数在 6 个民族群体中检出的数目, 括号内的数字表示其相应频率。

表 2 不同群体的线性 *Fst* 值及其显著性差异

	回族	哈尼族	蒙古族(云南)	蒙古族(内蒙)	彝族	傣族	汉族 <sup>[1]</sup>	日本人 <sup>[1]</sup>
回族								
哈尼族	0.00520							
蒙古族(云南)	0.00946	0.00071						
蒙古族(内蒙)	0.00896	0.00174	0.00000					
彝族	0.02192*	0.00701	0.00000	0.00000				
傣族	0.01447	0.00713	0.00221	0.00740	0.00000			
汉族	0.05490*	0.03708	0.02797	0.03093*	0.00945	0.01117		
日本人	0.05659*	0.03846*	0.02909*	0.03208*	0.01017	0.01190	0.00000	

注: \*:  $P<0.05$ , 具有显著性差异。

表 3 不同群体中 Ewens-Watterson 中性检验结果

	回族	哈尼族	蒙古族(云南)	蒙古族(内蒙)	彝族	傣族	汉族 <sup>[1]</sup>	日本人 <sup>[1]</sup>
样本数	100	88	96	104	94	100	86	90
等位基因数	5	4	5	5	4	3	1	1
F(观察)	0.83200	0.87009	0.87999	0.87130	0.93753	0.94140	N.A.	N.A.
F(期望)	0.49265	0.56127	0.48439	0.48231	0.56984	0.67879	N.A.	N.A.
P 值	0.97200	0.94400	0.99600	0.98700	1.00000	0.96700	N.A.	N.A.

注: N.A.: 未分析。

Ewens-Watterson 中性检验结果显示: 本研究中的 6 个群体均未受到明显的自然选择作用( $P>0.05$ )(表 3)。

### 2.3 *ATXN2* 基因 STR 在群体中的聚类分析结果

应用群体间的线性 *Fst* 值, 通过 UPGMA 法构建针对 *ATXN2* 基因 STR 的系统进化树, 以表明在不同群体中该基因 STR 的起源关系。在本研究所分析的 8 个群体中, 2 个地区的蒙古族群体 *ATXN2* 基因 STR 的起源相近聚为一类, 这与其民族起源、遗传背景

等相一致, 后再与哈尼族群体聚类; 同时, 彝族和傣族群体中该位点起源较为接近; 而回族则为单独一支, 在该基因的进化上与其他群体分离较早, 不与之聚类; 在汉族和日本人 群体中均只检出同一种等位基因, 而单独聚类(图 1)。

### 2.4 MDS 分析结果

对上述分析的 8 个群体进行 MDS 多维尺度分析, 如图 2 所示: 从各群体的空间分布来看, 回族与其

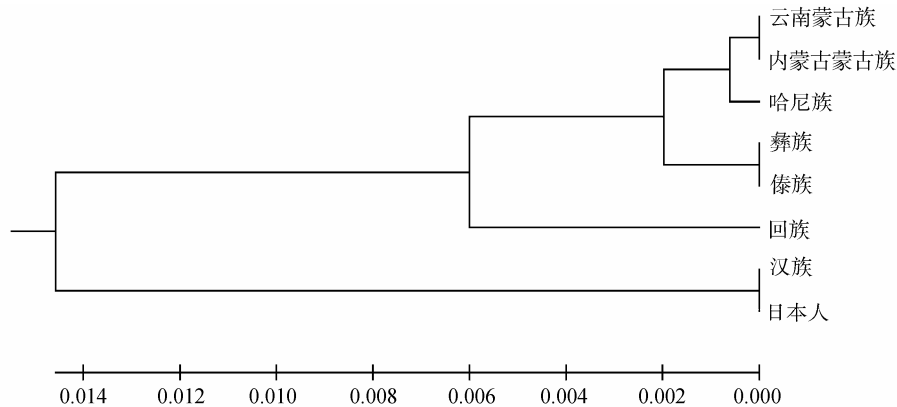


图 1 不同群体中 *ATXN2* 基因 STR 的系统进化树

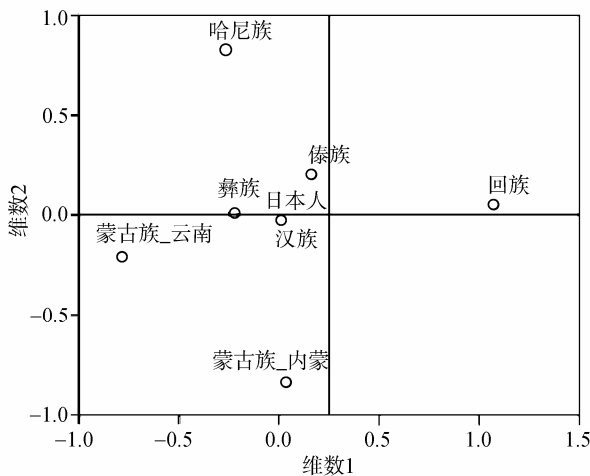


图 2 不同群体的多维尺度分析结果

他群体相距较远；两个地区的蒙古族群体聚类；哈尼族和傣族相近。

### 3 讨论

中国是一个多民族国家，拥有极丰富的群体遗传资源。本研究选择回族、蒙古族等 6 个民族群体进行 *ATXN2* 基因 CAG 重复与自然选择关系的研究。这些群体在遗传背景、生活习惯等方面不尽相同。对其进行研究，可能有助于分析此基因在不同民族或同一民族不同群体间的遗传、进化、环境与自然选择的相互关系等。

根据群体间 STR 位点多态性分布计算其线性 *F<sub>st</sub>* 值<sup>[10]</sup>，可以对多个群体同一位点上多个等位基因进行比较分析，并通过置换的方法计算 *P* 值，从而显示该基因进化的差异。本研究结果显示：在 6 个民族群体中 *ATXN2* 基因 STR 等位基因频率最高的是 22 次重复 (91%~97%)，回族和彝族群体间 *ATXN2* 基因 STR 位点进化的差异具有显著性，其他 4 个群体之

间无显著性差异。回族群体的 19 次重复在稀有等位基因中具有最大频率 (6%)，并且其突变等位基因多态性最高、分布最广 (19~31 次重复)，而彝族稀有等位基因频率较其他群体为低 (1%)，加之多态性分布也较为局限 (22~27 次重复)，可能是其产生显著性差异的原因。与文献报道的群体比较后发现，在汉族和日本人<sup>[11]</sup>群体中，STR 重复全部为 22 次 (100%)，等位基因多态性较为单一；而回族、哈尼族、内蒙古自治区蒙古族以及云南蒙古族多态性较其他群体为高。总的来说，与汉族及日本人相比，本研究的 6 个民族群体 STR 多态性程度较高，有明显的差异。同时，针对 *ATXN2* 基因 STR 的系统进化树也显示：回族为单独一支，不与其他群体聚类，与我们的实验数据显示的该群体具有高频稀有等位基因 (19 次重复)、多态性最高、分布最广 (19~31 次重复) 等结果一致；而在汉族和日本人<sup>[11]</sup>群体中只检出了一种等位基因，因而单独聚为一类。MDS 多维尺度分析的结果与系统进化树所示的结果一致，提示我们 *ATXN2* 基因的进化可能受到各民族群体的遗传背景影响。

STR 重复次数的变化在群体中是一个逐渐突变的过程，在一代一代的繁殖传代中优势等位基因将被保留，而稀有等位基因则可能被淘汰。本研究结果发现，在所有群体中，该 STR 等位基因分布高度倾向于 22 次重复，但是仍有 13 个稀有等位基因被检出。相对于其他疾病相关的编码三联体重复来说，*ATXN2* 基因 STR 重复次数的变化非常低<sup>[11]</sup>。在这个位点上可能由于 CAA 的插入使其保持稳定，从而导致突变率相对较低<sup>[12]</sup>，然而我们在 *ATXN2* 基因中检出的这些稀有等位基因在群体中占有一定可比较的数量，所以这一说法并不能完全解释稀有突变频率

的变化。本研究在回族群体中检出了 30 次、31 次重复的等位基因, 该 STR 的过度扩增可能导致脊髓小脑共济失调的患病风险, 如果该基因 STR 在传代过程中发生扩展突变, 这对于后代来说可能存在患病危险。但这些不利的等位基因在群体中保留下来, 并且能够轻易地从常见等位基因进化而来, 也暗示选择压力可能作用于维持少数等位基因在群体中的优势。为什么这些会产生不利的遗传性疾病的危险的突变仍然受到正选择作用? 这一现象可能与基因多效性有关, 突变在一种功能上是有利的, 但在另一功能上是有害的<sup>[13,14]</sup>。例如, 携带葡萄糖-6-磷酸脱氢酶基因(*G6PD*)的一个变异体 202A 的个体具有抗疟疾感染的能力<sup>[15]</sup>, 而在另一方面这个突变又会造成镰刀型贫血症。由此猜测, 是否 *ATXN2* 基因的突变也在某个疾病的抵抗性方面起到重要作用, 或者存在一个“搭载效应”使其产生优势从而被保留下来, 这有待于进一步的考证。

Ewens-Watterson 中性检验并没有显示本研究的群体受到明显的自然选择作用。然而, 由于在本研究中有 13 个稀有等位基因被检出并占有一定的比例, 与汉族和日本人群体相比, 我们所研究的 6 个少数民族群体有各自特有的等位基因, 频率分布有其各自的特点, 且在回族群体中检出的等位基因有扩增的趋势, 如果在传代过程中发生扩展突变达到一定阈值, 可能会在后代中产生患病风险, 到底是否存在选择的作用, 还有待于进一步的研究。这些群体生活在相对隔离的环境, 加之 Ewens-Watterson 中性检验本身存在一定的保守性<sup>[16]</sup>, 不足以解释本研究的数据, 从而我们推测其稀有等位基因多态性程度升高、频率增加的原因可能是选择作用的结果, 因此本研究结果并不能排除这一现象受到自然选择的作用。我们将进一步对该基因邻近区域相关位点进行深入研究, 以期阐明该 STR 位点在少数民族群体中是否确实受到自然选择的作用, 以及环境等因素是否影响了选择的方向, 对于分析基因的进化等方面具有重要意义。

#### 参考文献(References):

- [1] Yu FL, Sabeti PC, Hardenbol P, Fu Q, Fry B, Lu XH, Ghose S, Vega R, Perez A, Pasternak S, Leal SM, Willis TD, Nelson DL, Belmont J, Gibbs RA. Positive selection of a pre-expansion CAG repeat of the human *SCA2* gene. *PLoS Genet*, 2005, 1(3): e41. [\[DOI\]](#)
- [2] Mutesa L, Pierquin G, Segers K, Vanbellinghen JF, Gahimbare L, Bours V. Spinocerebellar ataxia type 2 (SCA2): Clinical features and genetic analysis. *Trop Pediatr*, 2008, 54(5): 350–352. [\[DOI\]](#)
- [3] 陈朴, 马明义, 商慧芳, 苏丹, 张思仲, 杨元. 常染色体显性小脑性共济失调致病基因动态突变位点三核苷酸重复变异的研究. *中华医学遗传学杂志*, 2009, 26(6): 626–633. [\[DOI\]](#)
- [4] 谭建强, 汪洋, 胡启平, 李松峰, 舒伟, 马军, 方玲, 华荣, 丁晔, 袁志刚. 广西地区脊髓小脑性共济失调病人的基因诊断和 CAG 重复扩增. *遗传*, 2009, 31(6): 605–610. [\[DOI\]](#)
- [5] 王进, 李桂冰, 罗曼, 王栋慧. 脊髓小脑性共济失调 2 型的分子遗传学诊断及临床分析. *中国临床神经科学*, 2009, 17(1): 12–15. [\[DOI\]](#)
- [6] 沈璐, 唐北沙, 汤熙翔, 谭斯品, 戴和平, 潘乾, 严新翔, 欧阳珊. 脊髓小脑型共济失调 II 型(CAG)<sub>n</sub> 突变检测研究. *中华内科杂志*, 2000, 39(4): 259–261. [\[DOI\]](#)
- [7] 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(3): 269–276. [\[DOI\]](#)
- [8] 吴东东, 张亚平. 人类群体中的达尔文正选择. *科学通报*, 2008, 53(7): 737–746. [\[DOI\]](#)
- [9] Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online*, 2005, 1: 47–50. [\[DOI\]](#)
- [10] Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 1995, 139(1): 457–462. [\[DOI\]](#)
- [11] Andrés AM, Lao O, Soldevila M, Calafell F, Bertranpetit J. Dynamics of CAG repeat loci revealed by the analysis of their variability. *Hum Mutat*, 2002, 21(1): 61–70. [\[DOI\]](#)
- [12] Sobczak K, Krzyzosiak WJ. Patterns of CAG repeat interruptions in *SCA1* and *SCA2* genes in relation to repeat instability. *Hum Mutat*, 2004, 24(3): 236–247. [\[DOI\]](#)
- [13] Clark NL, Swanson WJ. Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet*, 2005, 1(3): e35. [\[DOI\]](#)
- [14] Harpending H, Cochran G. In our genes. *Proc Natl Acad Sci USA*, 2002, 99(1): 10–12. [\[DOI\]](#)
- [15] Ruwende C, Hill A. Glucose-6-phosphate dehydrogenase deficiency and malaria. *J Mol Med*, 1998, 76(8): 581–588. [\[DOI\]](#)
- [16] Watterson GA. The homozygosity test of neutrality. *Genetics*, 1978, 88(2): 405–417. [\[DOI\]](#)

[1] Yu FL, Sabeti PC, Hardenbol P, Fu Q, Fry B, Lu XH, Ghose S, Vega R, Perez A, Pasternak S, Leal SM, Willis TD, Nelson DL, Belmont J, Gibbs RA. Positive selec-