

Founder Haplotype for Machado-Joseph Disease in the Indian Population

Novel Insights From History and Polymorphism Studies

Uma Mittal, MSc; Achal K. Srivastava, DM; Satish Jain, DM; Sanjeev Jain, MD; Mitali Mukerji, PhD

Background: The ACA haplotype is associated with 72% of the expanded repeats in Machado-Joseph disease (MJD) worldwide and has been traced to a Portuguese ancestry. It is present in only 5% of the normal chromosomes in the Portuguese population.

Objective: To trace the origin of expanded alleles of MJD in the Indian population.

Methods: We performed CAG repeat size determination and haplotype analysis for 9 families with MJD and 263 unrelated chromosomes with unexpanded CAG sequences from the Indian population.

Results: All the expanded alleles were exclusively associated with the ACA haplotype in the Indian population. Interestingly, this haplotype was very common in normal alleles (40%) as compared with the Portuguese

population (5%) and was significantly associated with large normal alleles (Pearson $\chi^2=87.1$, $P<.001$) in the Indian population. We also observed a rare intermediate allele of MJD with the ACA haplotype but with a CAG variant instead of CAA at the sixth position in the repeat tract.

Conclusions: Overrepresentation of the ACA haplotype in large normal alleles in India as compared with the Portuguese population suggests that the expansion-prone large normal alleles with the ACA haplotype may have been introduced in the Portuguese population through admixture with South Asian populations. Detailed haplotype analysis of a CAG variant within the repeat tract in an intermediate allele of MJD suggests a mechanism of gene conversion in the expansion.

Arch Neurol. 2005;62:637-640

MACHADO-JOSEPH DISEASE (MJD)/spinocerebellar ataxia type 3 (SCA3) (MIM 109150) is the most common autosomal dominant cerebellar ataxia caused by CAG repeat expansion.¹ Normal individuals have 12 to 40 CAG repeats, which expand to 61 to 86 repeats in affected individuals.¹⁻³ There is a wide gap between the repeat lengths of the normal and expanded alleles. Intermediate alleles of MJD in the range of 46 to 56 CAG repeats have been reported in only 3 rare cases.⁴⁻⁶

Worldwide, 72% of families with MJD have an ACA haplotype that has been traced to Portuguese-Azorean ancestry.⁷ This haplotype, however, is not significantly associated with large normal alleles at the MJD locus in the Portuguese population ($P=.30$).⁸ The role of various genetic factors, such as repeat length and haplotype background, in predisposing the repeats toward instability as well as the mechanism involved in large jumps from normal to expanded alleles at the MJD locus is still not well established.

In the present study, we analyzed intragenic single nucleotide polymorphisms (SNPs) in the Indian population to identify an association between haplotypes and repeat length at the MJD locus. All the affected chromosomes had the ACA haplotype, which was significantly associated with large normal alleles (Pearson $\chi^2=87.1$, $P<.001$), suggesting that the expanded chromosomes arose from a pool of large normal alleles. We propose the involvement of gene conversion as a possible mechanism in the origin of a rare intermediate allele at the MJD locus. We also provide historical evidence for the possible introduction of this haplotype in the Portuguese population through admixture with South Asian populations.

METHODS

SUBJECTS

This study was undertaken in 9 families with MJD, 83 unrelated normal individuals, and 24 normal families of Indian origin. Informed consent was obtained from all subjects before blood samples were collected. Patients were di-

Author Affiliations: Functional Genomics Unit, Institute of Genomics and Integrative Biology, New Delhi (Ms Mittal and Dr Mukerji), Neuroscience Centre, All India Institute of Medical Sciences, New Delhi (Drs Srivastava and Satish Jain), and Molecular Genetics Laboratory, Department of Psychiatry, National Institute of Mental Health and Neurosciences, Bangalore (Dr Sanjeev Jain), India. Dr Satish Jain is now with the Indian Epilepsy Centre, New Delhi.

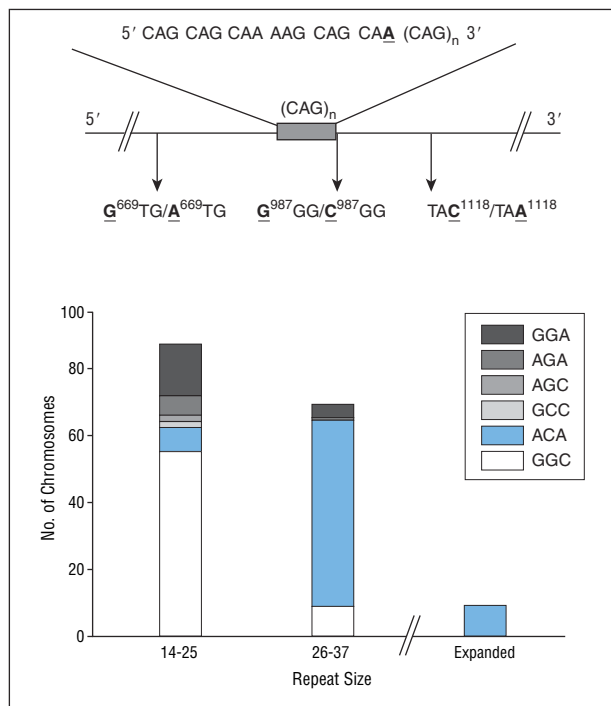


Figure 1. Association of intragenic haplotypes with CAG repeat lengths at the *MJD* locus. Top, Schematic representation of relative locations of intragenic polymorphisms and CAG repeats in Machado-Joseph disease complementary DNA. The CAA/CAG variant at the sixth position in the repeat tract is shown. Bottom, Distribution of intragenic haplotypes according to CAG repeat number in a normal population ($n=161$) and in families ($n=9$) of Indian origin with Machado-Joseph disease.

agnosed for ataxia at the Neuroscience Centre, All India Institute of Medical Sciences, New Delhi. Genomic DNA was isolated from peripheral blood leukocytes using the salting-out procedure.⁹

ESTIMATION OF THE CAG REPEAT LENGTH

Repeat sizes were estimated at the *MJD/SCA3* locus by polymerase chain reaction with primers MJD52 and MJD70¹ using GeneScan software (Applied Biosystems, Foster City, Calif) on an ABI Prism 377 automated DNA sequencer (Applied Biosystems).

SNP GENOTYPING AND HAPLOTYPE ANALYSIS

Haplotype analysis at the *MJD* locus was carried out using 3 closely linked intragenic SNPs, A⁶⁶⁹TG/G⁶⁶⁹TG, C⁹⁸⁷GG/G⁹⁸⁷GG, and TAA¹¹¹⁸/TAC¹¹¹⁸ (Figure 1).^{7,8} The first SNP was typed by sequencing, and the other two were typed using amplification refractory mutation system polymerase chain reaction (ARMS-PCR)⁸ followed by sizing the allele-specific fragments on the ABI Prism 377 automated DNA sequencer. In the case of A⁶⁶⁹TG/G⁶⁶⁹TG heterozygotes, for which phase could not be determined, the polymorphism was typed through pedigree studies. The CAA/CAG polymorphism at the sixth triplet of the repeat tract (Figure 1) was typed in 88 individuals by sequencing using the primers MJD52 and MJD70 on an ABI Prism 3100 automated genetic analyzer (Applied Biosystems). We used Lewontin's normalized measure of linkage disequilibrium (D'), $D' = D/D_{max}$, where $D = pAB - pApB$, $D_{max} = \min(pApb, pApB)$, p denotes frequency, and A and a and B and b are the two alternate alleles of each SNP.

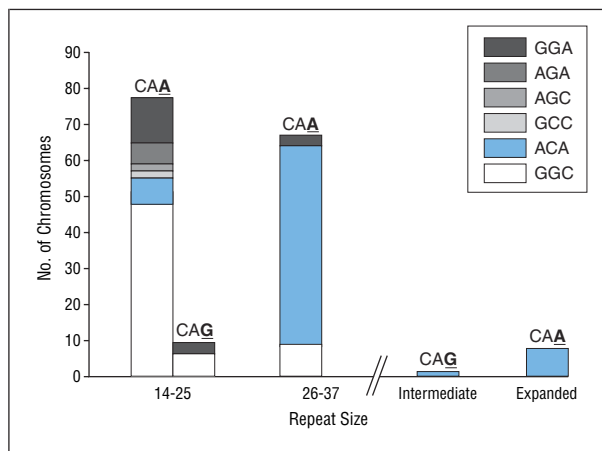


Figure 2. Distribution of CAA/CAG variant at the sixth position of the CAG repeat stretch with respect to intragenic haplotypes in the normal ($n=153$), intermediate, and expanded Machado-Joseph disease alleles.

RESULTS

CAG REPEAT AT *MJD* LOCUS

We found only 9 families with MJD in a total of 335 families with ataxia. Analysis of 263 unrelated chromosomes from the Indian population revealed 16 different alleles ranging from 14 to 37 CAG repeats; the repeat distribution was trimodal, with distinct peaks at 14, 23, and 27.¹⁰

CAG REPEAT AND INTRAGENIC POLYMORPHISMS

All 9 affected families were exclusively associated with the ACA haplotype. Analysis of 3 intragenic SNPs (Figure 1) in the normal alleles showed a significant association (Pearson $\chi^2 = 67.7, 114.4, \text{ and } 73.0; P < .001$) with respect to repeat length. The ACA and GGC alleles of the 3 SNPs were in linkage disequilibrium ($D' > 0.9$) with respect to each other. The ACA and GGC haplotypes account for more than 80% of the haplotypes observed in the Indian population (Figure 1), with an extremely skewed distribution between the larger and smaller normal alleles. Alleles with repeat sizes smaller than 26 were significantly associated with the GGC haplotype (Pearson $\chi^2 = 42.1, P < .001$), whereas large normal alleles with a repeat size of 26 or more were significantly associated with the ACA haplotype (Pearson $\chi^2 = 87.1, P < .001$).

ASSOCIATION OF HAPLOTYPES WITH CAG REPEAT SUBSTRUCTURE

More than 94% (144/153) of the normal chromosomes had a CAA allele at the sixth position, while 6% (9/153) had a CAG allele; the CAG allele is prominently associated with 19 repeats (Figure 2), very similar to the frequency reported earlier.^{10,11} This CAG variant was absent in the large normal alleles and had a flanking non-ACA haplotype.

Interestingly, of the 9 affected chromosomes, only the intermediate allele (45 repeats) had the CAG variant, while the remaining 8 expanded alleles had the CAA variant.

LOW PREVALENCE OF MJD IN THE INDIAN POPULATION

The prevalence of MJD in India (<3%) is much lower than that reported in other Asian populations (Korea, 29%¹²; Japan, 43%¹³) and the Portuguese population (84%).¹⁴ The frequency of large normal alleles with more than 31 CAG repeats (1%) was significantly lower in the present study than that reported in a comparable study in the Japanese population,¹³ which might explain the lower prevalence of MJD in India.

ORIGIN OF EXPANDED CHROMOSOMES FROM LARGE NORMAL ALLELES

The occurrence of the ACA haplotype in 40% of the normal chromosomes in the present study, the highest among studied populations, indicates the antiquity of this haplotype in the Indian population. Moreover, the association of the ACA haplotype with large normal alleles suggests that the expanded alleles originated from the pool of large normal alleles.

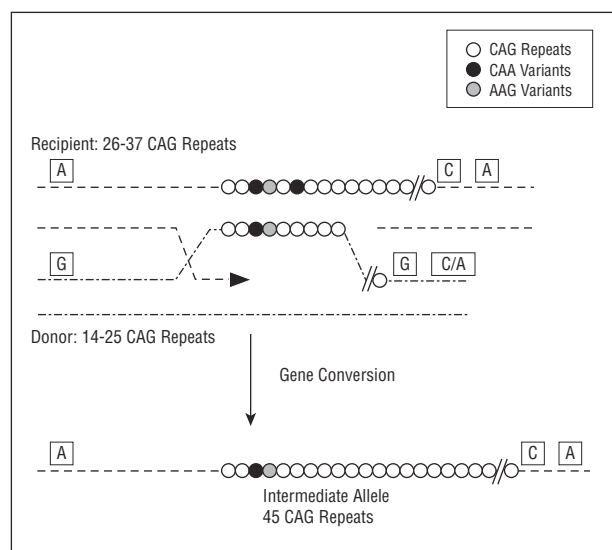


Figure 3. The proposed model for the mechanism of gene conversion involved in the origin of the intermediate allele at the *MJD* locus.

MECHANISM OF ORIGIN OF THE INTERMEDIATE ALLELE

Although the intermediate allele had the same flanking ACA haplotype, it had a CAG variant at the sixth position that was not present in the large normal and expanded alleles. This suggests that the intermediate allele might have arisen through an interallelic gene conversion event, with an ACA allele with a CAA variant at the sixth position as the recipient and a smaller allele with a CAG variant acting as the donor (all CAG variants are present in alleles with fewer than 26 repeats) (**Figure 3**). Since this event would involve a smaller and a larger normal allele, an intermediate length could be the result. By an analogous mechanism, large expanded alleles could be generated by an event involving two ACA alleles, but this needs to be validated in other populations. The possibility of recombination cannot be determined, as the flanking haplotype of the donor and the recipient is same.

Because the mechanism of expansion involves larger normal alleles, the haplotype associated with the expanded allele might depend on the frequency of the haplotypes associated with large normal alleles in a given population. This might also explain why different haplotypes are associated with expanded alleles in other populations (**Table**).⁸

INTRODUCTION OF THE FOUNDER HAPLOTYPE FOR MJD INTO THE AZORES

Comparison of 16 non-Portuguese families with MJD from across the world revealed the ACA haplotype in the vast majority, suggesting that this is the ancestral haplotype associated with the affected chromosomes.⁷ However, the ACA and GGC haplotypes occur with almost equal frequency in expanded chromosomes in the Azores, despite the rarity of the ACA haplotype in the normal population (**Table**). In contrast, the ACA haplotype is very common in India, especially in large normal alleles (>26 repeats). This indicates that large normal alleles with the ACA haplotype may have been introduced in the Azores through admixture between the Indian and Portuguese populations, which could have served as founders of the MJD mutation in the Azorean population. This is corroborated by well-documented historical evidence related to the Moorish sea trade and to maritime links between Portugal and South Asia.¹⁵ The Portuguese had

Table. Haplotypes in Normal and Expanded Alleles in the Indian and Azorean Populations*†

No. of CAG Repeats	Haplotype, %			
	ACA		GGC	
	Indian	Azorean	Indian	Azorean
14-21	14	5	50	68
22-26	36.5	0	55.5	81
27-37	95.2	7	4.8	71
Expanded	100	47	0	45

*Data for the Azorean population are from Maciel et al.⁸

†In normal alleles in the Indian population there were 161 chromosomes.

extensive settlements in India, and though their population was limited, extensive interaction with the local communities was encouraged. By 1815, when the population was formally recorded, there were several families of known mixed descent in Goa and several thousand individuals.¹⁶ In addition, slaves were traded from Goa, and a significant number of sailors on Portuguese ships were from the Indian and African coasts. Indian sailors and soldiers were also used extensively in the Portuguese territories further East.¹⁵ The maritime importance of the Azores was critical, and soldiers, sailors, and slaves of all regions passed through the islands. Thus, it is reasonable to suggest that a chromosome with the ACA haplotype may have been introduced into Portugal and the Azores by admixture with the South Asian population, and that a subsequent mutation in this haplotype background could have spread worldwide through Portuguese activity. However, this theory needs more careful validation, as maritime links between the Indian subcontinent, Southeast Asia, China, and Africa predate the Portuguese. Data on the ACA haplotype in the normal populations of these regions are not available.

Given the wide expanse of Portuguese movements, it would be worthwhile to compare haplotypes along these maritime routes rather than based on geographical proximity.¹⁷ It would also be useful to study the gene structures of these populations to understand the differences in disease and symptoms.

We were not able to determine whether there was one founder for the MJD mutant allele that through drift or a population bottleneck skewed to such a high proportion of the affected alleles in the Portuguese population. It is also possible that there have been recurrent mutations in a predisposed background. In addition, since there are other haplotypes that are also associated with expanded alleles in the Azorean population, the possibilities could coexist. Moreover, since there has been admixture from other populations with the Portuguese,¹⁸ this could also explain the presence of multiple founders.

In conclusion, our study suggests that the expansion-prone large normal alleles with the ACA haplotype could have been introduced into the Portuguese population through admixture with South Asian populations. Our analysis also underscores the importance of population history in defining the genomic structure of a locus to understand the molecular mechanism of repeat instability.

Accepted for Publication: May 26, 2004.

Correspondence: Mitali Mukerji, PhD, Institute of Genomics and Integrative Biology, Council for Scientific and Industrial Research, Mall Road, Delhi 110007, India (mitali@igib.res.in).

Author Contributions: *Study concept and design:* Mittal, Srivastava, Satish Jain, and Mukerji. *Acquisition of data:* Mittal, Srivastava, Satish Jain, and Mukerji. *Analysis and interpretation of data:* Mittal, Satish Jain, Sanjeev Jain, and Mukerji. *Drafting of the manuscript:* Mittal, Srivastava, Sanjeev Jain, and Mukerji. *Critical revision of the manuscript for important intellectual content:* Mittal, Srivastava, Satish Jain, Sanjeev Jain, and Mukerji. *Administrative, technical, and material support:* Sanjeev Jain. *Study supervision:* Satish Jain, Sanjeev Jain, and Mukerji.

Funding/Support: This work was supported by grants from the Department of Biotechnology, Government of India, to the Program on Functional Genomics, Institute of Genomics and Integrative Biology, New Delhi, and from the Council for Scientific and Industrial Research.

Acknowledgments: We thank Samir K. Brahmachari, PhD, for providing intellectual support during the course of this investigation; Vani Brahmacharifar, PhD, critical evaluation of the manuscript; and Sangeeta Sharma, BSc, for help with the GeneScan and sequence analysis.

REFERENCES

1. Kawaguchi Y, Okamoto T, Taniwaki M, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet.* 1994;8:221-228.
2. Maciel P, Gaspar C, DeStefano AL, et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am J Hum Genet.* 1995;57:54-61.
3. Maruyama H, Nakamura S, Matsuyama Z, et al. Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. *Hum Mol Genet.* 1995;4:807-812.
4. Takiyama Y, Sakoe K, Nakano I, Nishizawa M. Machado-Joseph disease: cerebellar ataxia and autonomic dysfunction in a patient with the shortest known expanded allele (56 CAG repeat units) of the MJD1 gene. *Neurology.* 1997;49:604-606.
5. van Alfen N, Sinke RJ, Zwarts MJ, et al. Intermediate CAG repeat lengths (53,54) for MJD/SCA3 are associated with an abnormal phenotype. *Ann Neurol.* 2001;49:805-807.
6. Padiath QS, Srivastava AK, Roy S, Jain S, Brahmachari SK. Identification of a novel 45 repeat unstable allele associated with a disease phenotype at the MJD1/SCA3 locus. *Am J Med Genet B Neuropsychiatr Genet.* 2005;133:124-126.
7. 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;68:523-528.
8. Maciel P, Gaspar C, Guimaraes L, et al. Study of three intragenic polymorphisms in the Machado-Joseph disease gene (MJD1) in relation to genetic instability of the (CAG)_n tract. *Eur J Hum Genet.* 1999;7:147-156.
9. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16:1215.
10. Limprasert P, Nouri N, Heyman RA, et al. Analysis of CAG repeat of the Machado-Joseph gene in human, chimpanzee and monkey populations: a variant nucleotide is associated with the number of CAG repeats. *Hum Mol Genet.* 1996;5:207-213.
11. Igarashi S, Takiyama Y, Cancel G, et al. Intergenerational instability of the CAG repeat of the gene for Machado-Joseph disease (MJD1) is affected by the genotype of the normal chromosome: implications for the molecular mechanisms of the instability of the CAG repeat. *Hum Mol Genet.* 1996;5:923-932.
12. Lee WY, Jin DK, Oh MR, et al. Frequency analysis and clinical characterization of spinocerebellar ataxia types 1, 2, 3, 6, and 7 in Korean patients. *Arch Neurol.* 2003;60:858-863.
13. 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;63:1060-1066.
14. Silveira I, Lopes-Cendes I, Kish S, et al. Frequency of spinocerebellar ataxia type 1, dentatorubropallidolusian atrophy, and Machado-Joseph disease mutations in a large group of spinocerebellar ataxia patients. *Neurology.* 1996;46:214-218.
15. Subrahmanyam S. *The Portuguese Empire in Asia, 1500-1700: A Political and Economic History.* London, England: Longman Group Ltd; 1993.
16. *A General Statistical and Historical Report on Portuguese India, to Which Is Appended an Account of Contents Suppressed in 1835, Extracted From Official Documents in 1850 by Capt. Joaquim Jose Cicilia K.C.; Chief Secretary, Portuguese India.* Microfilm I 1253/Zug 1984, Index 590. Located at: British Library, London, England.
17. Santos C, Lima M, Montiel R, et al. Genetic structure and origin of peopling in the Azores islands (Portugal): the view from mtDNA. *Ann Hum Genet.* 2003;67:433-456.
18. Martins S, Matama T, Guimaraes L, et al. Portuguese families with dentatorubropallidolusian atrophy (DRPLA) share a common haplotype of Asian origin. *Eur J Hum Genet.* 2003;11:808-811.

REFERENCES

1. Caplan LR. *Posterior Circulation Disease: Clinical Findings, Diagnosis, and Management*. Boston, Mass: Blackwell Science; 1996.
2. Berger MS, Wilson CB. Intracranial dissecting aneurysms of the posterior circulation. *J Neurosurg*. 1984;61:882-894.
3. Fisher CM, Ojemann RG, Roberson GH. Spontaneous dissection of cervico-cerebral arteries. *Can J Neurol Sci*. 1978;5:9-19.
4. Le Tu PT, Zuber M, Meder JF, Mas JL. Dissection isolee de l'artere cerebrale posterieure. *Rev Neurol*. 1996;152:542-547.
5. Sasaki O, Koizumi T, Ito Y, Sorimachi T, Koike T, Tanaka R. Dissecting aneurysm of the posterior cerebral artery treated with proximal ligation. *Surg Neurol*. 1992;37:394-401.
6. Pozzati E, Padovani R, Fabrizi A, Sabatini L, Gaist G. Benign arterial dissections of the posterior circulation. *J Neurosurg*. 1991;75:69-72.
7. Pozzati E, Andreoli A, Limoni P, Casmiro M. Dissecting aneurysms of the vertebrobasilar system: study of 16 cases. *Surg Neurol*. 1994;41:119-124.
8. Ohki M, Nakajima M, Sato K, et al. A case of dissecting aneurysm associated with mixed connective tissue disease. *No To Shinkei*. 1994;46:855-858.
9. Ciceri E, Klucznik RP, Grossman RG, Rose JE, Mawad ME. Aneurysms of the posterior cerebral artery: classification and endovascular treatment. *AJNR Am J Neuroradiol*. 2001;22:27-34.
10. Berthier E, Bourrat C. Dissecting aneurysm of the posterior cerebral artery: case report and review of the literature. *Cerebrovasc Dis*. 1993;3:56-59.
11. Maillo A, Diaz P, Morales F. Dissecting aneurysm of the posterior cerebral artery: spontaneous resolution. *Neurosurgery*. 1991;29:291-294.
12. Pessin MS, Kwan ES, DeWitt LD, Hedges TR III, Gale DG, Caplan LR. Posterior cerebral artery stenosis. *Ann Neurol*. 1987;21:85-89.
13. Georgiadis AL, Yamamoto Y, Kwan ES, Pessin MS, Caplan LR. Anatomy of sensory findings in patients with posterior cerebral artery territory infarction. *Arch Neurol*. 1999;56:835-838.
14. Pessin MS, Lathi ES, Cohen MB, Kwan ES, Hedges TR III, Caplan LR. Clinical features and mechanism of occipital infarction. *Ann Neurol*. 1987;21:290-299.
15. Yamamoto Y, Georgiadis AL, Chang H-M, Caplan LR. Posterior cerebral artery territory infarcts in the New England Medical Center Posterior Circulation Registry. *Arch Neurol*. 1999;56:824-832.
16. Chaves C, Estol C, Esnaola M, et al. Spontaneous intracranial internal carotid artery dissection. *Arch Neurol*. 2002;59:977-981.
17. Caplan LR. Worsening in ischemic stroke patients: is it time for a new strategy? *Stroke*. 2002;33:1443-1445.
18. De Georgia M, Belden J, Pao L, Pessin M, Kwan E, Caplan LR. Thrombus in vertebrobasilar dolichoectatic artery treated with intravenous urokinase. *Cerebrovasc Dis*. 1999;9:28-33.

Correction

Errors in Table and Acknowledgment. In the article by Mittal et al titled "Founder Haplotype for Machado-Joseph Disease in the Indian Population," published in the April 2005 issue of the ARCHIVES (2005;62:637-640), in the **Table** at the bottom of page 639, the haplotype percentages in both Indian columns were incorrect. The table is reprinted here with the correct values.

Table. Haplotypes in Normal and Expanded Alleles in the Indian and Azorean Populations*†

No. of CAG Repeats	Haplotype, %			
	ACA		GGC	
	Indian	Azorean	Indian	Azorean
14-21	10.3	5	56.4	68
22-26	24.3	0	55.7	81
27-37	82.7	7	9.6	71
Expanded	100	47	0	45

*Data for the Azorean population are from Maciel et al.⁸

†In normal alleles in the Indian population there were 161 chromosomes.

In addition, a word was misplaced in the acknowledgments on page 640 near the top of column 2. The acknowledgments should have read as follows: "Acknowledgment: We thank Samir K. Brahmachari, PhD, for providing intellectual support during the course of this investigation; Vani Brahmachari, PhD, for critical evaluation of the manuscript; and Sangeta Sharma, BSc, for help with the GeneScan and sequence analysis."