Age and founder effect of *SOD1* A4V mutation causing ALS

M. Saeed, MD
Y. Yang, MS
H.-X. Deng, MD, PhD
W.-Y. Hung, PhD
N. Siddique, RN, MSN
L. Dellefave, MS, CGC
C. Gellera, PhD
P.M. Andersen, MD
T. Siddique, MD

Address correspondence and reprint requests to Prof. Teepu Siddique, Northwestern University, Feinberg School of Medicine, 303 East Chicago Avenue/Tarry 13-715, Chicago, IL 60611
t-siddique@northwestern.edu

ABSTRACT

Background: The alanine to valine mutation at codon 4 (A4V) of *SOD1* causes a rapidly progressive dominant form of amyotrophic lateral sclerosis (ALS) with exclusively lower motor neuron disease and is responsible for 50% of *SOD1* mutations associated with familial ALS in North America. This mutation is rare in Europe. The authors investigated the origin (geographic and time) of the A4V mutation.

Methods: Several cohorts were genotyped: North American patients with confirmed A4V mutation ($n = 54$), Swedish ($n = 3$) and Italian ($n = 6$) A4V patients, patients with ALS with *SOD1* non-A4V mutations ($n = 66$) and patients with sporadic ALS ($n = 96$), healthy white ($n = 96$), African American ($n = 17$), Chinese ($n = 53$), Amerindian ($n = 11$), and Hispanic ($n = 12$) subjects. High-throughput SNP genotyping was performed using *Taqman* assay in 384-well format. A novel biallelic CA repeat in exon 5 of *SOD1*, tightly linked to A4V, was genotyped on sequencing gels. Association statistics were estimated using Haploview. p Values less than 0.05 were considered significant. Age of A4V was estimated using a novel method based on r^2 degeneration with genetic distance and a Bayesian method incorporated in DMLE+.

Results: A single haplotype of 10 polymorphisms across a 5.86-cM region was associated with A4V ($p = 3.0 \times 10^{-11}$) when white controls were used, suggesting a founder effect. The strength of association of this haplotype progressively decreased when African American, Chinese, Hispanic, and Amerindian subjects were used as controls. The associated European haplotype was different from the North American haplotype, indicating two founder effects for A4V (Amerindian and European). The estimated age of A4V with the r^2 degeneration method was 458 ± 59 years (range 398–569) and in agreement with the Bayesian method (554–734 years with 80–90% posterior probability).

Conclusions: North American *SOD1* alanine to valine mutation at codon 4 descended from two founders (Amerindian and European) 400–500 years ago. *Neurology*® 2009;72:1634–1639

GLOSSARY

A4V = alanine to valine mutation at codon 4; **ALS** = amyotrophic lateral sclerosis; **FALS** = familial ALS; **LD** = linkage disequilibrium; **SDS** = sequence detection system; **SNP** = single nucleotide polymorphisms.

Amyotrophic lateral sclerosis (ALS) is characterized by motor neuron degeneration in the cerebral cortex and spinal cord resulting in progressive paralysis and death from respiratory failure within 3–5 years.^{1,2} The incidence is 1–2/100,000 population/year and is virtually constant worldwide with the exception of some areas of high incidence in the western Pacific, including Guam, western New Guinea, and the Kii Peninsula in Japan.^{3,4} A dominant form of ALS with exclusively lower motor neuron disease was first described by Sir William Osler in 1880 in the Farr family in Vermont. In the descendants of the Farr family, Siddique et al.⁵ mapped the familial ALS (FALS) gene to 21q22.1 in proximity to the gene encoding cytosolic Cu,Zn superoxide dismutase (*SOD1*). Mutations in *SOD1* account for 20% of familial cases of

Supplemental data at
www.neurology.org

Editorial, page 1628

e-Pub ahead of print on January 28, 2009, at www.neurology.org.

From the Davee Department of Neurology and Clinical Neurosciences (M.S., Y.Y., H.-X.D., W.-Y.H., N.S., L.D., T.S.), Feinberg School of Medicine, Northwestern University, Chicago, IL; Division of Biochemistry and Genetics (C.G.), National Neurological Institute-IRCCS, Carlo Besta, Milan, Italy; and Department of Clinical Neurosciences (P.M.A.), Umea University, Sweden.

Disclosure: The authors report no disclosures.

ALS (2% of all ALS)⁶ and TDP-43 for a small minority of cases.⁷⁻⁹ The causes of the rest (80%) of familial cases (8% of all ALS) are unknown.⁶ The etiopathogenesis of the remaining 90% of all ALS cases (sporadic ALS) is also undetermined.⁶

Collective evidence suggests a toxic gain of function of SOD1 conferred by mutations.⁶ Certain mutations affect the onset and progression of ALS.¹⁰ The molecular mechanism by which mutations in *SOD1* can convert the soluble enzyme into an aggregated form of toxic dimers and multimers associated with ALS was recently shown.¹¹ There are over 100 known *SOD1* mutations and all are dominant.¹² One of these, alanine to valine substitution on codon 4 (A4V; GCC→GTC) in exon 1¹³ was shown to be responsible for 50% of *SOD1* mutations associated with ALS in North America.^{10,13} A4V results in an aggressive form of ALS with a survival time of less than 2 years after disease onset.¹⁰ This mutation is rare in Europe. The explanation for such patterns lies in the demographic histories of populations including effects of genetic drift, migration, and natural selection. In this study we investigated the origin and founder effect of the A4V mutation in North America.

METHODS Subjects. Informed consent, approved by our respective Institutional Review Boards, to conduct genetic experiments and obtain medical records, was obtained from the study subjects. Fifty-four patients from independent families with confirmed A4V mutation were included in the study. Healthy white (n = 96), African American (n = 17), Chinese (n = 53), Hispanic (n = 12), and Amerindian (n = 11) subjects were also genotyped as control subjects. Patients with ALS from independent families having *SOD1* mutations other than A4V (n = 66) and 96 white patients with sporadic ALS were genotyped to define the extent of linkage disequilibrium (LD). Swedish (n = 3) and Italian (n = 6) A4V patients were the only available patients outside of North America and were analyzed for founder effect. Informed consent was obtained from all subjects included in this study.

Genotyping. Single nucleotide polymorphisms. Genomic DNA was extracted from whole blood according to established protocols.¹³ High-throughput single nucleotide polymorphisms (SNP) genotyping was performed using *Taqman* assay in a 384-well format on the ABI prism 7900HT sequence detection system (Applied Biosystems). In each chamber of the 384-well plate, 10 ng of genomic DNA was used. PCR primers and probes were obtained from Applied Biosystems assays-on-demand service. Reactions were run in 5 μ L volumes using *Taqman* Universal PCR master mix, primers, and probes. The amplification protocol was 95°C for 10 minutes followed by 40 cycles of 95°C for 5 seconds and 60°C for 1 minute. Genotype

data were obtained using the ABI-PRISM sequence detection system (SDS) software version 2.1.

Novel biallelic CA repeat polymorphism. In a previous sequencing of *SOD1* in FALS samples we had identified an unreported biallelic CA repeat (allele 1: 8-CA repeats/allele 2: 7-CA repeats) that strongly segregated with the A4V mutation. It is located 2,057 bp downstream of the stop codon in exon 5 of *SOD1*. Genotyping for this polymorphism was done using PCR with radiolabeled primers (F-GCTCTTAGGTTGCAAATGTTAAACTTGAT, R-GTTGGATCCCAGTGTTACACGTTGTACT) and separating the PCR products on a sequencing gel (7% urea, polyacrylamide gel). Genotypes were read as shown in figure e-1 on the *Neurology*[®] Web site at www.neurology.org. We used this biallelic polymorphism as a genetic marker for the A4V mutation in our study along with SNPs in the region.

Statistical analysis. Haplotype frequencies and association statistics for the polymorphisms were estimated using Haploview version 3.2.¹⁴ *p* Values less than 0.05 were considered significant. To estimate the age of A4V we used a Bayesian method (using Markov chain Monte Carlo method) for multipoint linkage disequilibrium mapping incorporated in the program DMLE+ version 2.2.¹⁵ Age of the mutation was also estimated based on marker-marker correlation coefficient (r^2) degeneration with genetic distance. By measuring the extent of LD (r^2) between SNPs/biallelic CA repeat and recombination rates (θ) in Morgans (based on the DeCode map) as surrogate markers of the original A4V DNA fragment, the age of the A4V mutation can be estimated. Assuming that there was 100% LD of all tested markers with the mutation at the time the mutation was introduced and that LD degenerates in proportion to recombination, age of a mutation approximates $(1 - r^2)/\theta$ in generations.

RESULTS We carried out a modified case-control association analysis of A4V patients (n = 54) and white control subjects (n = 96) with 21 polymorphisms (20 SNPs and the biallelic CA repeat) across a 21 cM (15.3 Mb) region (12 cM centromeric to *SOD1* and 9 cM telomeric) on chromosome 21 (figure e-2). Significant LD existed 0.3 Mb around the biallelic CA repeat, which is located 2,057 bp downstream of the stop codon in *SOD1* exon 5. Modest LD was also demonstrated with rs276171 (1.44 Mb from *SOD1*) and rs933130 (2.24 Mb from *SOD1*) (figure e-2).

Nine SNPs and the biallelic CA repeat (10 polymorphisms) in a 5.86 cM region (3.07 Mb) associated with A4V, most notably the CA repeat ($p = 1.72 \times 10^{-14}$) and rs276171 (3.6 cM from *SOD1*) ($p = 1.78 \times 10^{-6}$) (table e-1). A single haplotype (2221212112) of these 10 polymorphisms associated with A4V ($p = 3.0 \times 10^{-11}$). This haplotype was absent in the controls (frequency case: control = 0.23:0.0). No other haplotype showed association. The 10 polymorphisms constituting this haplotype were genotyped in subsequent cohorts for further analysis.

To determine whether the association of rs276171, 1.44 Mb away from *SOD1*, was due to LD with the A4V mutation or signified the existence of a novel ALS gene, 66 patients with *SOD1* muta-

tions other than A4V (non-A4V) were genotyped for the 10 polymorphisms constituting the A4V associated haplotype. The design resulted in pooling of several different *SOD1* mutation-haplotypes together. If a gene other than *SOD1* existed near rs7276171 then two association peaks would be observed; one within *SOD1* where all haplotypes overlapped and another near rs7276171 as observed with the A4V pool alone. However, association was found only with the SNP rs202445, located within *SOD1* at its 5' end (table e-2). Additionally, patients with sporadic ALS (n = 96) were genotyped for rs7276171 and no association was found. As genetically (non-A4V FALS) and clinically (SALS) defined ALS cohorts did not demonstrate associations outside of *SOD1*, the association signal spread over 5.86 cM with the A4V mutation indicates the region is most likely in LD with the A4V mutation. Since a single haplotype, 3.07 Mb long, carried the association signal due to LD with A4V, it indicated a founder effect for the mutation.

To determine the origin of this DNA fragment, we genotyped samples from four other major populations in North America—African American, Chinese, Hispanic, and Native American (Amerindian)—for the 10 polymorphisms constituting the associated haplotype in the white population. As the allele frequencies in the A4V population would be similar to the population closest to the origin of A4V, we expected the strength of associations to decrease if this were the case. As shown in table e-3, the genotype associations of the 10 polymorphisms for all four populations decreased in strength relative to those observed when white controls were used (table e-1). The strong haplotype association with white control subjects decreased markedly when A4V patients were compared with healthy African American ($p = 0.0022$), Chinese ($p = 0.0039$), and Hispanic control subjects ($p = 0.0172$). The association nearly disappeared when A4V cohort was compared with Amerindian subjects ($p = 0.032$), indicating that the fragment on which the A4V mutation arose was originally introduced from Amerindians who migrated from Asia into North America.

There was no difference between the Swedish and the Italian haplotype, indicating a common European origin. However, this European haplotype was different (1211211122) from the North American haplotype (2221212112). All Europeans were homozygous for allele1 of the CA repeat. However, only 18% of North American A4V patients were homozygous for allele 1 and most carried allele 2 (82%). Since the CA repeat is located only 11.8 kbp from the A4V mutation within *SOD1* and LD with A4V extends up to 3.07 Mb, recombination is un-

likely. This indicated different origins of the A4V mutation. Thus we partitioned our North American A4V cohort into two groups: those homozygous for the CA repeat allele 1 and carriers of allele 2. The remaining 9 SNPs were used to carry out associations. Haplotype (9-SNPs) statistics of North American A4V subjects homozygous for allele 1 of CA repeat with European A4V patients showed absence of association suggesting similarity of genetic background. Significant differences in SNPs (rs7276171, $p = 7.04 \times 10^{-5}$ and rs10154126, $p = 0.0038$) and haplotypes (221212112; $F = 0.24$, $p = 0.03$ and 121211122; $F = 0.06$, $p = 0.02$) were found when North American A4V patients carrying the CA repeat allele 2 were compared with European A4V patients. These findings suggest that A4V was introduced into North America from Europeans (18%) and Amerindians (82%) separately.

To estimate when A4V was introduced into the white population, two independent methods were used. The Bayesian method for multipoint LD mapping incorporated in the program DMLE+ applied the white haplotype frequency data for the 10 associated polymorphisms in the 5.86 cM region to calculate the posterior probability of the time of onset of A4V mutation. The estimated age of A4V, using this method, was 14 to 61 generations (271 to 1,231 years, assuming 20-year generation time) with 50 to 99% posterior probability (figure e-3). With 80 to 90% probability, A4V was transmitted 554 to 734 years ago.

To apply the r^2 degeneration method, we used the CA repeat as a marker for the A4V mutation and calculated r^2 values for the four distant SNPs viz rs7276171, rs2834171, rs3827181, and rs933130 (figure 1). Since rs1537096 is located relatively close to the CA repeat and is in substantial LD with it, it was also taken to be a reference point. Therefore, the r^2 degeneration calculation was repeated for the four distant SNPs with rs1537096 to confirm the results obtained with the CA repeat. As shown in the table, the age of A4V mutation based on r^2 degeneration with recombination distance was calculated to be 458 ± 59 years (range 398–569).

DISCUSSION This study demonstrates that the high prevalence of the A4V mutation in North American patients with FALS is due to a founder effect. This mutation has had two separate origins in the course of human evolution, one in Europe and the other in Amerindians, who probably carried it into North America from Asia. This fits in with our observation that most of the A4V families in our databank originate from the western New York state region. The Amerindian contribution (82%) to the

Figure 1 Novel method to estimate age of *SOD1* alanine to valine mutation at codon 4 using r^2 degeneration with genetic distance (Morgans)

prevalence of A4V in North America far outweighs that of the European contribution (18%). It is likely that A4V was introduced into the white population from the Amerindians about 400 to 500 years ago at about the time of the Jamestown and Plymouth landings. In our databank of over 500 FALS cases, none are Amerindian. It is possible that for reasons unknown the mutation became extinct in the Amerindian population. The other possibility is that this difference may be due to ascertainment bias or sociocultural factors. Indeed, a third possibility remains that ALS does not manifest in Amerindians in spite of the A4V mutation due to a protective genetic factor. This inhibitory factor may not have been transmitted to the white population along with the A4V mutation. Clearly, population-based studies that would take into account biologic and sociocultural

factors would help in the understanding of the reasons for this discrepancy.

Using a modified case-control association analysis we mapped a large region (<10 cM) on either side of *SOD1*, in known A4V patients compared to white control subjects. This helped us to identify the extent of LD with A4V and therefore essentially map the size of the DNA fragment introduced into the white population. The association signal was spread over a surprisingly long stretch (>3 Mb), questioning the existence of LD that is known to normally degenerate in less than 60 kb.¹⁶ The effect was due to our modified approach with selection of a pure group of A4V patients as cases. We verified this by the absence of associations distant from *SOD1* in non-A4V and sporadic ALS cohorts. Each *SOD1* mutation is expected to have its own unique haplotype. Pooling these differing haplotypes together was expected to eliminate association except near *SOD1*. We found this to be the case when we pooled all non-A4V patients together as cases and carried out association with white controls.

The existence of a single associated haplotype for A4V patients against white controls demonstrated a founder effect. This haplotype was most closely matched to Amerindians, indicating its origin in this population. It differed most from white subjects, suggesting its later introduction in them. When compared with A4V patients, the allele 2 of the CA repeat showed high frequency in North American A4V patients, Amerindians, and Chinese subjects, and demonstrated absence of association in the Chinese and the Amerindian populations. Thus, the Amerindians, who are likely to be descendants of people who migrated from Siberia to North America by crossing the Strait of Bering, probably carried it into North America. Time of origin analysis using two independent methods confirmed with high probability that A4V was introduced into the white

Table Estimation of the age of *SOD1* A4V mutation using novel r^2 degeneration with genetic distance (Morgans) method

Marker	rs7276171	rs2834171	rs3827181	rs933130	Measure
CA repeat	0.175	0.053	0.053	0.073	r^2
	3.61	4.03	4.53	4.66	Relative cM (θ)
	22.9	23.5	20.9	19.9	Age $[(1 - r^2) \times 100/\theta]$ (generations)
	457	470	418	398	Age of A4V (y)
rs1537096	0.024	0.008	0.11	0.039	r^2
	3.43	3.85	4.35	4.48	Relative cM (θ)
	28.5	25.8	20.5	21.5	Age $[(1 - r^2) \times 100/\theta]$ (generations)
	569	515	409	429	Age of A4V (y)

A4V = alanine to valine mutation at codon 4.

Figure 2 Origin of the *SOD1* alanine to valine mutation at codon 4

Migration of haplotypes into North America based on significance of association (<http://en.wikipedia.org/wiki/Image:BlankMap-World-noborders.png>).

population approximately 400 to 500 years ago. Here we demonstrate a novel method of calculating age of a mutation based on the LD statistic, r^2 . The calculation is simple and rapid. Its consistency is demonstrated by a narrow SD obtained with multiple markers and its agreement with results from DMLE+.

Interestingly, all European A4V patients (Swedish and Italian) were homozygous for allele 1 of the CA repeat and there was no difference between the Swedish and Italian haplotypes. The most prevalent European haplotype was different from the North American haplotype. Further analysis showed that 18% of North American A4V patients were also homozygous for allele 1 of the CA repeat and had genetic backgrounds similar to the European A4V patients. Since recombination is unlikely at this locus due to its location within *SOD1*, it indicated that A4V of European origin was introduced into North America separately with subsequent migrations. Thus the *SOD1* A4V mutation has had two independent founders: one in Europe and one in the admixed North American white population, where the latter took place ~400 to 500 years ago (figure 2) (<http://en.wikipedia.org/wiki/Image:BlankMap-World-noborders.png>).

In a previous study, using a multigenerational family-based approach and by genotyping four microsatellite markers in ~3 Mb region (30.1 to 32.8

Mb) on chromosome 21, it was postulated that A4V had occurred independently at least twice over the course of human evolution.¹⁷ Our study confirms this observation and traces the origin of the A4V geographically and temporally. It was shown that the apparent contradiction of the existence of *SOD1* D90A mutation in both dominant and recessive forms is due to founder effects with a single founder for the recessive pedigrees and several for the dominant pedigrees.¹⁸ All D90A recessive families maintained a consistent inheritance pattern and had milder disease.¹⁸ This study suggested that a tightly linked protective factor was most likely responsible for the milder recessive form of the disease.¹⁸ Our study traces the origin of *SOD1* A4V mutation to Amerindians, who appear to rarely contract ALS, and the hypothesis most worthy of further exploration is that there exists a protective genetic factor for *SOD1* A4V mutation as well. If this is true then it will be an important disease-modifying factor, awaiting identification.

Received April 19, 2007. Accepted in final form November 18, 2008.

REFERENCES

1. Tandan R, Bradley WG. Amyotrophic lateral sclerosis, part 1: clinical features, pathology, and ethical issues in management. *Ann Neurol* 1985;18:271–280.
2. Tandan R, Bradley WG. Amyotrophic lateral sclerosis, part 2: etiopathogenesis. *Ann Neurol* 1985;18:419–431.

3. Bobowick AR, Brody JA. Epidemiology of motor neuron diseases [review]. *N Engl J Med* 1973;288:1047–1055.
4. Nelson LM. Epidemiology of ALS. *Clin Neurosci* 1995;3:327–331.
5. Siddique T, Figlewicz DA, Pericak-Vance MA, et al. Linkage of a gene causing familial amyotrophic lateral sclerosis to chromosome 21 and evidence of genetic-locus heterogeneity. *N Engl J Med* 1991;324:1381–1384.
6. Siddique T, Deng HX. Genetics of amyotrophic lateral sclerosis. *Hum Mol Genet* 1996;5:1465–1470.
7. Kabashi E, Valdmanis PN, Dion P, et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet* 2008;40:572–574.
8. Sreedharan J, Blair IP, Tripathi VB, et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* 2008;319:1668–1672.
9. Van Deerlin VM, Leverenz JB, Bekris LM, et al. TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *Lancet Neurol* 2008;7:409–416.
10. Juneja T, Pericak-Vance MA, Laing NG, Dave S, Siddique T. Prognosis in familial amyotrophic lateral sclerosis: progression and survival in patients with glu100gly and ala4val mutations in Cu,Zn superoxide dismutase. *Neurology* 1997;48:55–57.
11. Deng HX, Shi Y, Furukawa Y, et al. Conversion to the amyotrophic lateral sclerosis phenotype is associated with intermolecular linked insoluble aggregates of SOD1 in mitochondria. *Proc Natl Acad Sci USA* 2006;103:7142–7147.
12. Radunovic A, Leigh PN. ALSO Database: database of SOD1 (and other) gene mutations in ALS on the Internet: European FALS Group and ALSOD Consortium. *Amyotroph Lateral Scler Other Motor Neuron Disord* 1999;1:45–49.
13. Deng HX, Hentati A, Tainer JA, et al. Amyotrophic lateral sclerosis and structural defects in Cu,Zn superoxide dismutase. *Science* 1993;261:1047–1051.
14. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–265.
15. Reeve JP, Rannala B. DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics* 2002;18:894–895.
16. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001;29:229–232.
17. Rosen DR. A shared chromosome-21 haplotype among amyotrophic lateral sclerosis families with the A4V SOD1 mutation. *Clin Genet* 2004;66:247–250.
18. Al-Chalabi A, Andersen PM, Chioza B, et al. Recessive amyotrophic lateral sclerosis families with the D90A SOD1 mutation share a common founder: evidence for a linked protective factor. *Hum Mol Genet* 1998;7:2045–2050.

New Guidelines Examine Safety of Women with Epilepsy and Pregnancy

New evidence-based practice guidelines developed by the American Academy of Neurology in full collaboration with the American Epilepsy Society show the relative safety for women with epilepsy to become pregnant, but caution against taking one particular epilepsy drug, which can cause birth defects. The guidelines were published in the April 27, 2009, online issue of *Neurology*[®] and were presented at the AAN's Annual Meeting in Seattle. The guidelines were also published electronically in *Epilepsia*, the journal of the International League Against Epilepsy. They represent an update of the 1998 guideline, "Management Issues for Women with Epilepsy."