Human Molecular Genetics, 1996, Vol. 5, No. 12 **2019–2022**

# A founder mutation in the γ-sarcoglycan gene of Gypsies possibly predating their migration out of India

**F. Piccolo[1], M. Jeanpierre[1], F. Leturcq[1], C. Dodé[1], K. Azibi[1,2], A. Toutain[3], L. Merlini[4], L. Jarre[5], C. Navarro[6], R. Krishnamoorthy[7], F. M. S. Tomé[8], J. A. Urtizberea[9], J. S. Beckmann[10], K. P. Campbell[11] and J.-C. Kaplan[1,*]**

[1]INSERM 129 and Service de Biochimie et Génétique Moléculaire, Hôpital Cochin, 123 boulevard de Port-Royal, 75014 Paris, France, [2]CHU Bologhine, Algiers, Algeria, [3]CHU Tours, France, [4]Istituto Ortopedico Rizzoli, Bologna, Italy, [5]Osp. Regina Margherita, Torino, Italy, [6]Hospital do Meixoeiro, Vigo, Spain, [7]INSERM 120, Paris, France, [8]INSERM 153, Paris, France, [9]AFM, 13 place de Rungis, Paris, France, [10]Généthon, Evry, France and [11]University of Iowa, College of Medicine, Iowa City, IA, USA

Received August 19, 1996 Revised and Accepted September 4, 1996

**We investigated the molecular basis of a severe form of early onset autosomal recessive muscular dystrophy with sarcoglycan (SG) deficiency in seven large Gypsy families living in different parts of Western Europe and apparently not closely related. They were linked to the LGMD2C locus (13q12) suggesting a primary defect in the γ-SG gene coding for the 35 kDa dystrophin-associated glycoprotein. All of the 18 investigated patients were homozygous for the same G→A transition in codon 283 producing the replacement of a conserved cysteine of the extra-cellular domain of the protein by a tyrosine. All affected chromosomes in homozygous and heterozygous relatives carried the same allele 5 of the intragenic marker *D13S232*. Flanking markers were studied to delineate a common ancestral haplotype, the size of which was used to compute the date of the founding mutation. We found evidence that the mutation occurred between 60 and 200 generations ago, therefore possibly predating the commonly accepted date of migration of the Gypsy ancestors out of India.**

## INTRODUCTION

Sarcoglycanopathies are muscular dystrophies due to the disruption of the dystrophin-associated sarcoglycan complex (1,2) caused by mutations in any of the four already identified sarcoglycan genes (α-,β-,γ- and δ-SG) (3–8).

We investigated the molecular basis of a severe form of early onset autosomal recessive muscular dystrophy with sarcoglycan (SG) deficiency in seven large Gypsy families living in different parts of Western Europe and apparently not closely related (see Table 1). Linkage analysis could be performed in five families and showed that the disease involved the LGMD2C locus at 13q12 suggesting a defect in the γ-sarcoglycan gene coding for the 35 kDa dystrophin-associated glycoprotein (7). A common defect responsible for a C283Y missense and consistently associated with allele 5 of the intragenic *D13S232* marker was indeed found in all affected patients and carriers.

## RESULTS

Muscle γ-SG RNA was amplified by RT–PCR and sequenced in two unrelated patients. In both cases we found a homozygous G→A transition in codon 283, changing a conserved cysteine residue in the extracellular domain of the protein to tyrosine. This mutation creates a new *Rsa*I restriction site facilitating its survey at the genomic level. In all seven families this mutation (*Rsa*I$^+$) cosegregated with the disease, all of the 18 patients being homozygous and the asymptomatic parents being heterozygous. The heterozygous *Rsa*I$^+$ siblings (15 individuals) were also asymptomatic, but some exhibited a moderate increase of creatine kinase.

The mutation was not found either in 14 chromosomes from unrelated control Gypsies, in 426 unrelated chromosomes (34 Caucasian and 392 non-tribal Indian) from normal non-Gypsy individuals, nor in 116 unrelated chromosomes (30 Caucasian, 72 North African, 14 Western Asian) from non-Gypsy unclassified LGMD2 patients. In the seven Gypsy families we found that all of the 65 mutated chromosomes (18 homozygous patients and 29 heterozygous relatives) carried allele 5 (112 bp) at the intragenic (personal data) polymorphic microsatellite *D13S232* (GDB accession ID: GDB:292119). Allele 5 is common in several distinct populations (GDB allele frequency: 0.39), including Gypsies in which it was also found independently of the C283Y mutation (Table 2).

## DISCUSSION

On the basis of linguistic, socio-cultural, morphological, serological and HLA haplotype evidence (9,10) the Romani Gypsies of Europe are believed to originate from Northern Indian populations who arrived in Europe around AD 1100 (9). Due to almost complete endogamy they form a genetically isolated community, with increased incidence of autosomal recessive diseases.

*To whom correspondence should be addressed

**Table 1.** The seven Gypsy families with LGMD2C

| Family | Country | Total no. patients | No. patients studied | Mutation in γ-SG gene | Muscle α-SG (N = +++) | Muscle γ-SG (N = +++) | Phenotype |
|---|---|---|---|---|---|---|---|
| F1550 | France | 9 (4[a]) | 3 | C283Y | + | 0 | Severe[b] |
| F1708 | Italy | 5 | 4 | C283Y | + | 0 | Severe[b] |
| F2256 | France | 3 (1[a]) | 2 | C283Y | 0/+ | 0 | Severe[b] |
| F2500 | Italy | 1 | 1 | C283Y | 0 | nd | Early symptoms[c] |
| F2528 | Spain | 12 (1[a]) | 6 | C283Y | 0 | 0 | Severe[b] |
| F2572 | France | 7 (6[a]) | 1 | C283Y | nd | nd | Severe[b] |
| F2877 | France | 1 | 1 | C283Y | nd | nd | Severe[d] |

[a]Deceased patients.
[b]Wheel-chair bound before 14 years.
[c]Currently 3 years old (ref. 18).
[d]Currently 10 years old.
nd, not done.

**Table 2.** γ-SG haplotypes in different populations

| Haplotype | Number of independent chromosomes | | | | |
|---|---|---|---|---|---|
| | Gypsies | | Non-Gypsies | | |
| | with LGMD2C (7)[a] | without LGMD2C (10) | Caucasians (24) | North Africans (48) | Non-tribal Indians (36) |
| *Rsa*I+/allele 5[b] | 100% | 0% | 0% | 0% | 0% |
| *Rsa*I–/allele 5[b] | 0% | 30% | 58% | 19% | 25% |
| RsaI+/other allele[b] | 0% | 0% | 0% | 0% | 0% |

*Rsa*I+ = C283Y mutation; *Rsa*I– = wild type.
[a]One chromosome per consanguineous family.
[b]Marker *D13S232*.

The seven Gypsy families with severe autosomal recessive muscle dystrophy of the LGMD2C type live in France, Italy and Spain. They are not closely related since they are known to have settled in these countries several generations ago. The finding of a common C283Y mutation, not yet described in another population background, consistently associated with allele 5 of the intragenic *D13S232* marker, suggests that it is a private mutation arising in a γ-SG gene already carrying this allele. A comparable situation has already been described with the Tunisian γ-SG mutation which is consistently associated with allele 3 (122 bp) of the same *D13S232* marker (7,11).

How old is the C283Y mutation? This mutation occurred on an ancestral haplotype that has been scrambled through successive meiotic recombinations except for the most proximal markers. The size of the original haplotype remaining in a given family can be deduced by comparing the mutation bearing haplotypes in distinct families. It is correlated to the time of occurrence of the mutation and could be used as a genetic clock: the smaller the remaining haplotype, the older the mutation.

Attempts have already been made to estimate the age of a mutation from linkage disequilibrium of flanking markers (12,13). Here we take advantage of the high consanguinity of Gypsies to estimate the age of the mutation from the size of the remaining haplotype expressed as a recombination fraction using flanking markers. If $g$ generations have elapsed since the appearance of the mutation, the probability that no recombination occurred in a fraction $t$ on each side of the mutation is $(1–t)^g$. Deriving $g$ from $t$ is a typical Bayesian problem. We computed that the size of the remaining original haplotype on each side of a mutation is approximately $1/g$ when $g$ is >20. For example if $t$ = 1 cM, $g$ is ~100. The number of families does not affect the maximum likelihood value, but affects the confidence interval which is inversely correlated to the number of families. Our preliminary data using microsatellites flanking the intragenic marker *D13S232* show a minimal size of the ancestral haplotype (Fig. 1). Computation suggested that the most likely number of generations elapsed since the appearance of the mutation is 110 (95% confidence interval: 60–200) (Fig. 2). Assuming that a generation is represented by 20 years (14), this would indicate that the C283Y mutation in the γ-SG gene is at least 1200 years old. If the genetic clock had been reset by a bottleneck around the time of wandering (fixation of one haplotype in a small population), the age of the mutation would be even older. We therefore assume that the C283Y mutation predates the commonly accepted date of migration of Gypsies out of Northern India. This hypothesis may be tested by investigating the Indian tribes from which Gypsies are believed to have originated (9).

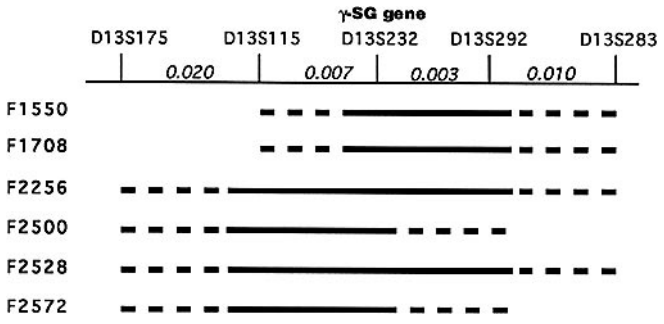


**Figure 1.** Fraction of the original haplotype remaining in six Gypsy families. *D13S232* is an intragenic marker of the γ-sarcoglycan gene. Fraction of recombination between markers is indicated in italics.

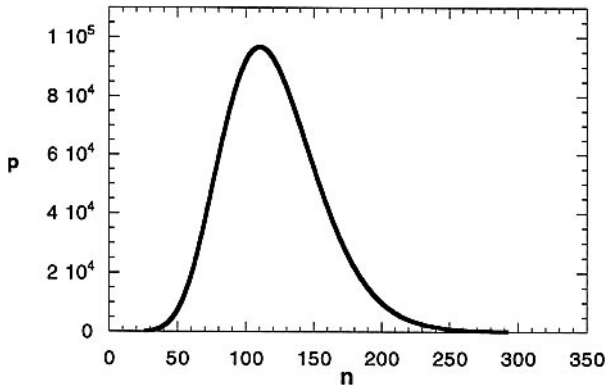


**Figure 2.** Probability density function (p) of number of generations (n) elapsed since the occurrence of the C(283)Y mutation in the γ-sarcoglycan gene in Gypsy ancestors, computed from haplotypes depicted in Figure 1, using a Bayesian algorithm.

## MATERIALS AND METHODS

### Family data

The diagnosis was ascertained on the following criteria: Duchenne-like clinical presentation, males and females equally affected, recessive mode of inheritance, high serum creatine kinase level, normal dystrophin and decreased level of α- and γ-SG in muscle specimens (15).

### Genotyping

DNA was prepared from peripheral blood lymphocytes and amplified using polymorphic microsatellite (*D13S175*, *D13S115*, *D13S232*, *D13S292* and *D13S283*) primer pairs listed in the Genome Database (GDB), or from the Généthon marker collection as described (16).

### RT–PCR and mutation analysis

Total RNA was extracted from muscle biopsies using a RNAsol kit (Bioprobe) and cDNA synthesis was performed using random hexanucleotide primers. The coding sequence of the γ-SG gene was amplified as previously reported (7). The PCR products were purified by electrophoresis in 0.8% low-melting-point agarose gels and sequenced directly with the ABI DyeDeoxy Terminator Cycle Sequencing kit, using the PCR primers. Both strands of PCR products were sequenced using an ABI 377 sequencer. The C283Y creates an *Rsa*I restriction site visualised on 4% agarose gel after genomic DNA amplification for 35 cycles at 94/59/72°C for 60/60/60 s with the following primers: 5′-CCTGTCTGTGGCCG-GTGTGA and 5′-GCGTTTACTTCCCATCCACGCTGC.

### Statistical analysis

Since all patients were offspring of consanguineous matings (first cousin or first cousin once removed), each mutation was assumed to have been inherited through a recent heterozygous parent and only one haplotype was taken into consideration in each family. The C283Y mutation occurred on an ancestor haplotype that could be reconstructed by comparing the distinct family haplotypes. For any generation number *g* elapsed since the appearance of the mutation (actually the earliest common ancestor) the probability of observing on one side of a mutation the remaining original haplotype in a given interval *t* can be directly calculated as $(1–t)^g$. The likelihood of any generation number is computed following Bayes' theorem by dividing each of these probabilities by the sum of all probabilities. Corrections have been made for allele frequencies assuming that present allele frequencies in Gypsies reflect past frequencies. Since we assume that the seven families share common ancestors they could not be considered as independent, but this non-independence only affects the confidence interval, not the likelihood peak, and could at least be partially corrected following a Luria–Delbrück type of analysis (17). In order to assess algorithm efficiency, we have used computer generated haplotypes obtained from simulations using various realistic parameters and given numbers of generation, to calculate backward the probability distribution of these generation numbers. The algorithm appeared robust to a reasonable variation in demographic parameters (not shown). These computation features have been implemented in a dedicated computer program ABEL.

## ACKNOWLEDGEMENTS

We wish to thank A. Barois, D. Bonneau, B. Echenne, C. Guillou, J. C. Lambert, M. Mora, S. Muralitharan, I. Penisson, O. Boespflug-Tanguy, J. M. Vallat and J. M. Warter for providing clinical information and samples from their patients, N. Deburgrave and S. Llense for their invaluable technical assistance and the Association Française contre les Myopathies (AFM) for financial support. K.P.C. is an Investigator of the Howard Hughes Medical Institute. F.P. and K.A. are fellows of the AFM.

## REFERENCES

1. Campbell, K.P. (1995) Three muscular dystrophies: loss of cytoskeleton-extracellular matrix linkage. *Cell*, **80**, 675–679.
2. Ozawa, E., Yoshida, M., Suzuki, A., Mizuno, Y., Hagiwara, Y. and Noguchi, S. (1995) Dystrophin-associated proteins in muscular dystrophy. *Hum. Mol. Genet.*, **4**, 11711–11716.
3. Roberds, S.L., Leturcq, F., Allamand, V., Piccolo, F., Jeanpierre, M., Anderson, R.D., Lim, L.E., Lee, J.C., Tomé, F.M.S., Roméro, N.B., Fardeau, M., Beckmann, J.S., Kaplan, J.C. and Campbell, K.P. (1994) Missense mutations in the adhalin gene linked to autosomal recessive muscular dystrophy. *Cell*, **78**, 625–633.

4. Piccolo, F., Roberds, S.L., Jeanpierre, M., Leturcq, F., Azibi, K., Beldjord, C., Carrié, A., Récan, D., Chaouch, M., Reghis, A., El Kerch, F., Sefiani, A., Voit, T., Merlini, L., Collin, H., Eymard, B., Beckmann, J.S., Romero, N., Tomé, F.M.S., Fardeau, M., Campbell, K.P. and Kaplan, J.-C. (1995) Primary adhalinopathy: a common cause of autosomal recessive muscular dystrophy of variable severity. *Nature Genet.*, **10**, 243–245.
5. Bönnemann, C.G., Modi, R., Noguchi, S., Mizuno, Y., Yoshida, M., Gussoni, E., McNally, E.M., Duggan, D.J., Angelini, C., Hoffman, E.P., Ozawa, E. and Kunkel, L.M. (1995) β-sarcoglycan (A3b) mutations cause autosomal recessive muscular dystrophy with loss of the sarcoglycan complex. *Nature Genet.*, **11**, 266–273.
6. Lim, L.E., Duclos, F., Broux, O., Bourg, N., Sunada, Y., Allamand, V., Meyer, J., Richard, I., Moomaw, C., Slaughter, C., Tomé, F.M.S., Fardeau, M., Jackson, C.E., Beckmann, J.S. and Campbell, K.P. (1995) β-sarcoglycan: characterization and role in limb-girdle muscular dystrophy linked to 4q12. *Nature Genet.*, **11**, 257–285.
7. Noguchi, S., McNally, E., Ben Othmane, K., Hagiwara, Y., Mizuno, Y., Yoshida, M., Yamamoto, H., Bönnemann, C.G., Gussoni, E., Denton, P.H., Kyriakides, T., Middleton, L., Hentati, F., Ben Hamida, M., Nonaka, I., Vance, J.M., Kunkel, L.M. and Ozawa, E. (1995) Mutations in the dystrophin-associated protein γ-sarcoglycan in chromosome 13 muscular dystrophy. *Science*, **270**, 819–822.
8. Nigro, V., Piluso, G., Belsito, A., Politano, L., Puca, A.A., Paparella, S., Rossi, E., Viglietto, G., Esposito, M.G., Abbondanza, C., Medici, N., Molinari, A.M., Nigro, G. and Puca, G.A. (1996) Identification of a novel sarcoglycan gene at 5q33 encoding a sarcolemmal 35 kDa glycoprotein. *Hum. Mol. Genet.*, **5**, 1179–1186.
9. de Pablo, R., Vilches, C., Moreno, M.E., Rementeria, M.C., Solis, R. and Kreisler, M. (1992) Distribution of HLA antigens in Spanish Gypsies: a comparative study. *Tissue Antigens*, **40**, 187–196.
10. Mastana, S.S. and Papiha, S.S. (1992) Origin of the Romany Gypsies—Genetic evidence. *Z. Morph. Anthropol*, **79**, 43–51.
11. Ben Othmane, K., Speer, M.C., Stauffer, J., Blel, S., Middleton, L., Ben Hamida, C., Etribi, A., Loeb, D., Hentati, F., Roses, A.D., Ben Hamida, M., Pericak-Vance, M.A. and Vance, J.M. (1995) Evidence for linkage disequilibrium in chromosome 13-linked Duchenne-like muscular dystrophy (LGMD2C). *Am. J. Hum. Genet.*, **57**, 732–734.
12. Morral, N., Bertranpetit, J., Estivill, X., Nunes, V., Casals, T., Gimenez, J., Reis, A., Varon, M.R., Macek, M.J. , Kalaydjieva, L. *et al* (1994) The origin of the major cystic fibrosis mutation (ΔF508) in European populations. *Nature Genet.*, **7**, 169–75.
13. Risch, N., de, L.D., Ozelius, L., Kramer, P., Almasy, L., Singer, B., Fahn, S., Breakefield, X. and Bressman, S. (1995) Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nature Genet.*, **9**, 152–159.
14. Bittles, A.H., Mason, W.M., Greene, J. and Rao, N.A. (1991) Reproductive behavior and health in consanguineous marriages. *Science*, **252**, 789–94.
15. Jung, D., Leturcq, F., Sunada, Y., Duclos, F., Tomé, F.M.S., Moomaw, C., Merlini, L., Azibi, K., Chaouch, M., Slaughter, C., Fardeau, M., Kaplan, J.C. and Campbell, K.P. (1996) Absence of γ-sarcoglycan (35 DAG) in autosomal recessive muscular dystrophy linked to chromosome 13q12. *FEBS Lett.*, **381**, 15–20.
16. Guilford, P., Dodé, C., Crozet, F., Blanchard, S., Chaib, H., Levilliers, J., Levi, A.F., Weil, D., Weissenbach, J., Cohen, D. *et al.* (1995) A YAC contig and an EST map in the pericentromeric region of chromosome 13 surrounding the loci for neurosensory nonsyndromic deafness (DFNB1 and DFNA3) and limb-girdle muscular dystrophy type 2C (LGMD2C). *Genomics*, **29**, 163–9.
17. Hästbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A. and Lander, E. (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genet.*, **2**, 204–11.
18. Morandi, L., Barresi, R., Di Blasi, C., Jung, D., Sunada, Y., Confalonieri, V., Dworzak, F., Mantegazza, R., Antozzi, C., Jarre, L., Pini, A., Gobbi, G., Bianchi, C., Cornelio, F., Campbell, K.P. and Mora, M. (1996) Clinical heterogeneity of adhalin deficiency. *Ann. Neurology*, **39**, 196–202.


## NOTE ADDED IN PROOF

Since submission of this paper we found three additional Gypsy families from France, Spain and Germany with muscular dystrophy associated with the C283Y mutation in the γ-SG gene.