Meta Gene 15 (2018) 60–64

Contents lists available at ScienceDirect

Meta Gene

journal homepage: www.elsevier.com/locate/mgene

Prevalence and ancestral origin of the c.1987delC *GAA* gene mutation causing Pompe disease in Central Mexico

Adriana Grijalva-Pérez[a], Carmen Esmer[b], Sandra Romero-Hidalgo[c], Arturo Rojo-Dominguez[d], Sandra Rosas-Madrigal[e], Rafael Velazquez-Cruz[c], Teresa Villarreal-Molina[c], Victor Acuña-Alonzo[f], Antonio Bravo-Oro[b], Alessandra Carnevale[c,*]

[a] *FES-Zaragoza UNAM, Mexico*
[b] *Hospital Central "Dr. Ignacio Morones Prieto" San Luis Potosí, Mexico*
[c] *Instituto Nacional de Medicina Genómica, Mexico*
[d] *Universidad Autónoma Metropolitana, Unidad Cuajimalpa, Mexico*
[e] *CONACYT-Instituto Nacional de Medicina Genómica, Mexico*
[f] *Escuela Nacional de Antropología e Historia, Mexico*

ARTICLE INFO

Keywords:
Pompe Disease
GAA mutation
Ancestry
Component

ABSTRACT

Pompe Disease (PD) is a rare autosomal recessive disorder caused by lysosomal enzyme acid-alpha-glucosidase (GAA) deficiency, which leads to lysosomal glycogen accumulation, swelling, cellular damage and dysfunction of cardiac, respiratory and muscular tissue. In Mexico, < 100 PD cases have been diagnosed. Recently, three apparently unrelated classic infantile PD cases coming from three different isolated communities of Central Mexico were diagnosed at the same hospital in the State of San Luis Potosi, and were found to be homozygous for a novel frameshift GAA gene mutation (c.1987delC). The purpose of the study was to determine mutation carrier frequencies in these localities, to alert the authorities on the high risk of PD in this region, to gain insight on the structural effects of the mutation by protein modeling, and to disclose the ancestral origin of the mutation. Quantitative PCR analyses showed that 24/307 (7.8%) subjects from two of the communities were heterozygous for the mutation. A genetic counseling campaign and newborn screening were organized with the help of local authorities. Tridimensional modeling predicted the mutant c.1987delC the second of the 3 GAA domains would be improperly folded and the third domain missing. Global ancestry analysis of six individuals from these communities (5 mutation carriers and 1 non-carrier) showed that > 95% of their genetic component was of Native Mexican origin, while < 4% was of European origin. The *GAA* mutation was found to lie within a segment of Native Mexican ancestry, probably Mayan-Zapotec.

1. Introduction

Glycogen storage disease type II (OMIM 232300) or Pompe Disease (PD) is an autosomal recessive disorder caused by mutations of the acid-glucosidase (GAA) lysosomal enzyme (acid maltase, EC 3.2.1.20, Swiss) (Raben et al., 2002). GAA enzyme deficiency leads to lysosomal glycogen accumulation and consequently lysosomal swelling, cellular damage and tissue dysfunction (Van der Ploeg and Reuser, 2008). Two broad phenotypes can be distinguished based on the age of onset of symptoms, extent of organ involvement, and clinical course, infantile onset and late onset PD. (Van der Ploeg and Reuser, 2008) Classic infantile PD typically presents during the first few weeks of life with hypotonia, progressive weakness, macroglossia, hepatomegaly and hypertrophic cardiomyopathy, with a debilitating and frequently fulminating course, associated with feeding problems and severe respiratory failure (Kishnani et al., 2006; van Capelle et al., 2016; Chu et al., 2016).

PD is caused by pathogenic mutations in both *GAA* alleles (Hermans et al., 2004). The Pompe Disease Mutation Database (www.pompecenter.nl) lists over 400 different causal sequence variants in the *GAA* gene. While most mutations are rare and have been found in only a few patients, a subset of mutations are frequent in certain populations due to founder effects. The most common mutation is the intronic c.32–13 T > G causing late onset PDM; the c.525delT deletion is common in the Dutch population, c.2238G > C is common in Taiwan and China; and tc.2560C > T is found in up to 60% of PD cases of African descent (Becker et al., 1998; Liu et al., 2014; MENA Pompe Working Group et al., 2015).

* Corresponding author.
E-mail address: acarnevale@inmegen.gob.mx (A. Carnevale).

https://doi.org/10.1016/j.mgene.2017.11.005
Received 5 September 2017; Received in revised form 6 November 2017; Accepted 15 November 2017
Available online 16 November 2017

We previously described three apparently unrelated children with classical infantile Pompe disease, homozygous for a novel frameshift *GAA* gene mutation (c.1987delC), who lived in 3 different isolated communities of the Huastec region in Central Mexico (Esmer et al., 2013). The purpose of the present study was to determine the prevalence of mutation carriers in these localities to alert the authorities on the high risk of PD in this region, to gain insight on the structural effects of the mutation by protein modeling, and to disclose the ancestral origin of the mutation.

2. Material and methods

2.1. Participants

Participants were recruited on site by an open announcement made in the three Huastec region communities (Papatlaco, Emiliano Zapata and Tetlalpa). All individuals willing to participate, and whose parents and grandparents had been born in the same community were included. Participants recruited from Papatlaco and Emiliano Zapata were apparently unrelated to the index case of each community, however most individuals recruited from Tetlalpa referred to be second or third degree relatives of the child with PD from this community.

A total of 352 individuals were included in the study: 191 out of 810 inhabitants from Papatlaco, 116/187 from Emiliano Zapata and 45/490 from Tetlalpa. In addition, 200 individuals from non-isolated nearby towns in the Huastec region were recruited as controls. Informed consent was obtained from all subjects. The study was approved by the Research and Ethics Committees of all participant Institutions, and by local Health authorities of the State of San Luis Potosí.

2.2. High resolution amplicon melting analysis

Genomic DNA samples were extracted from peripheral whole blood using the Gentra Puregene Genomic DNA Isolation Kit (Qiagen). Exon 14 of GAA gene was amplified from genomic DNA by polymerase chain reaction (PCR) in a LightCycler® 480 device (Roche Diagnostics, Penzberg, Germany). High resolution melting reactions were performed with approximately 50 ng of DNA, 0.5 μM each primer (GAA F-gACTCTgCCCTCCCAgAAAT, GAA R-gATCgCCCACCTgCCAT), the labeled single probes (GAA LC640-CACAgCTCCTCTgAggTgTTgCCC–PH; GAA FL-AgCTgggTCCAgCgCAC–FL), and LightCycler® 480 genotyping master mix according to the manufacturer's protocol. Auto-call genotyping data determined by differentiation of the normalized and temperature-adjusted melting curves for analyzed samples were obtained using LightCycler 480 Software. All samples were tested in duplicate, and at least one heterozygous and five different normal controls were included in each run. Genotypes of all c.1987delC heterozygous and 80 non-carrier individuals were confirmed by Sanger sequencing.

2.3. Structural modeling

Molecular modeling was performed with the Molecular Operating Environment package (MOE, chemcomp.com) based on publicly available structures from the Protein Data Bank (PDB, rcsb.org). Both the GAA wild type sequence (NCBI Reference Sequence: NC_000017.11), and the GAA c.1987delC mutant sequence were used for modeling. Alignments of both wt and mutant sequences were performed with Clustal Omega (Sievers et al., 2011) and were adjusted based on structure to favor the location of indels on external loops. Two hundred initial structures were built randomly, selecting compatible main chain segments and amino acid conformations from MOE protein fragment libraries, in order to model indel regions and side chain replacements, respectively. Energy of initial models was minimized in MOE using CHARMM27 force field (MacKerell et al., 2004). Minimization was completed when the average energy gradient reached 0.05 kcal/(mol Å). The best structure according to MOE folding parameters was further refined until the energy gradient was below 0.01 kcal/(mol Å). The resulting models were checked with MOE for stereochemical quality.

2.4. Ancestral origin of the mutation

Genotypes obtained from Affymetrix SNP 6.0 microarrays were analyzed in six individuals: four first degree relatives of the index case from Tetlalpa (mother, father, a carrier sister and a non-carrier brother); one carrier individual from Papatlaco and one carrier individual from Emiliano Zapata. In addition, SNP 6.0 microarray genotype data from 279 Native Mexican individuals (55 Totonac, 86 Nahuan, 50 Zapotec and 88 Mayan), as well as 112 European and 113 African individuals from the Hap Map project were included as parental or reference populations.

Genotype data quality control (QC) was checked in PLINK (Purcell et al., 2007). SNPs with any of the following criteria were excluded from the analysis: minor allele frequency < 5%, call rate < 5%, mendelian inconsistencies, and/or linkage disequilibrium (r2 > 0.2). A total of 64,767 SNPs meeting quality control criteria common across the six studied individuals and reference populations were used for the analysis.

Global ancestry was estimated using ADMIXTURE software (Alexander et al., 2009). Reference populations were used to generate multidimensional scaling (MDS) plots using PLINK (Purcell et al., 2007). Relatedness among the six individuals was tested by pairwise identity by descent (IBD) estimations. Local ancestry was estimated across chromosome 17 using PCAdmix (Brisbin et al., 2012) based on phased-trio data as described by Moreno-Estrada et al. (Moreno-Estrada et al., 2014) A 2.5 Mb segment comprising the c.1987delC mutation and 75 surrounding SNPs was selected, and the putative ancestral origin of the mutation was determined using STRUCTURE (Pritchard et al., 2000) assuming K = 4.

3. Results

3.1. Genotype analysis

Altogether, 31 of the 352 subjects sampled from these 3 nearby and isolated communities within the Huastec region were found to be heterozygous *GAA* frameshift c.1987delC mutation carriers. No homozygous individuals were found. As shown in Table 1, 5.2% of tested individuals in Papatlaco, 12% in Emiliano Zapata and 15.5% in Tetlalpa were heterozygous mutation carriers. However, it is important to point out that most individuals recruited from Tetlalpa referred to be relatives of the PD index case from this community. No mutation carriers were found among the 200 controls recruited from other non-isolated communities of the Huastec region.

3.2. Protein modeling

The PDB file 3LPO was used as template. This template has 42% identity to the query sequence and 18 indels. The wildtype GAA model has three domains. The central domain is an alpha-beta barrel with eight parallel strands surrounded by eight helixes with loops of different lengths, and the other two domains are beta-sandwiches (Fig. 1).

Table 1
Prevalence of mutation carriers in Papatlaco and Emiliano Zapata.

	Papatlaco	E. Zapata	Total	Controls
Non carriers	181 (94.7%)	102 (87.9%)	288 (93.8%)	200 (100%)
Carriers	10 (5.2%)	14 (12%)	24 (7.8%)	0 (0%)
Total	191	116	307	200