



Frequency and origin of the c.2090T>G p.(Leu697Trp) *MYO3A* variant associated with autosomal dominant hearing loss

André S. Bueno¹ · Kelly Nunes¹ · Alex M. M. Dias¹ · Leandro U. Alves¹ · Beatriz C. A. Mendes² · Juliana Sampaio-Silva³ · Jeroen Smits^{4,5} · Helger G. Yntema⁶ · Diogo Meyer¹ · Karina Lezirovitz^{1,3} · Regina C. Mingroni-Netto¹

Received: 4 March 2020 / Revised: 15 March 2021 / Accepted: 8 April 2021 / Published online: 6 May 2021
© The Author(s), under exclusive licence to European Society of Human Genetics 2021

Abstract

We recently described a novel missense variant [c.2090T>G:p.(Leu697Trp)] in the *MYO3A* gene, found in two Brazilian families with late-onset autosomal dominant nonsyndromic hearing loss (ADNSHL). Since then, with the objective of evaluating its contribution to ADNSHL in Brazil, the variant was screened in additional 101 pedigrees with probable ADNSHL without conclusive molecular diagnosis. The variant was found in three additional families, explaining 3/101 (~3%) of cases with ADNSHL in our Brazilian pedigree collection. In order to identify the origin of the variant, 21 individuals from the five families were genotyped with a high-density SNP array (~600 K SNPs— Axiom Human Origins; ThermoFisher). The identity by descent (IBD) approach revealed that many pairs of individuals from the different families have a kinship coefficient equivalent to that of second cousins, and all share a minimum haplotype of ~607 kb which includes the c.2090T>G variant suggesting it probably arose in a common ancestor. We inferred that the mutation occurred in a chromosomal segment of European ancestry and the time since the most common ancestor was estimated in 1100 years (CI = 775–1425). This variant was also reported in a Dutch family, which shares a 87,121 bp haplotype with the Brazilian samples, suggesting that Dutch colonists may have brought it to Northeastern Brazil in the 17th century. Therefore, the present study opens new avenues to investigate this variant not only in Brazilians but also in European families with ADNSHL.

These authors contributed equally: André S. Bueno, Kelly Nunes

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41431-021-00891-0>.

✉ Regina C. Mingroni-Netto
renetto@ib.usp.br

- ¹ Centro de Pesquisas sobre o Genoma Humano e Células-Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil
- ² Divisão de Educação e Reabilitação dos Distúrbios da Comunicação da Pontifícia Universidade Católica de São Paulo, São Paulo, Brasil
- ³ Laboratório de Otorrinolaringologia/LIM32 – Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brasil
- ⁴ Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands
- ⁵ Department of Otorhinolaryngology, Radboud University Medical Center, Nijmegen, The Netherlands
- ⁶ Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands

Introduction

Hearing loss is one of the most frequent sensorial disorders in humans, with an incidence of 1.86 per 1000 newborns in developed countries such as the United States [1]. According to the World Health Organization, almost 466 million people are affected by hearing loss, of which 34 million are children. In developed countries, almost 80% of the hearing impairment cases have genetic causes, and the remaining 20% of deafness cases are due to environmental factors [2]. Nonsyndromic hearing loss (NSHL) corresponds to 75–80% of the deafness cases, and is transmitted with autosomal dominant inheritance in 20% of the cases, X-linked in 2–5%, and mitochondrial in about 1% [3]. To date, approximately 120 genes have been associated to NSHL. Despite the high genetic heterogeneity of hearing loss, many cases are due to mutations in a single gene, *GJB2*, which is responsible for 50% of autosomal recessive nonsyndromic hearing loss (ARNSHL), usually congenital or of early onset [4]. However, molecular investigation of autosomal dominant hearing loss, which is frequently of

later onset, is complicated by the lack of a major gene or causative variant that explains many of the cases.

The investigation of genes contributing to nonsyndromic deafness and the study of their function in the auditory system allows a better understanding of the molecular mechanisms and physiological pathways behind the disorder [5, 6].

Myosins are proteins that act as molecular motors, using the energy that results from the ATP hydrolysis to move along actin filaments [7]. While conventional myosins form bipolar filaments and are related to muscle contraction, several unconventional myosins, which do not form bipolar filaments, were related to the genetic basis of syndromic and NSHL. Unconventional myosins of the class III have a kinase N-terminal domain followed by a conserved motor domain and a specific C-terminal region [8]. There are two types of vertebrate class III myosins, MYO3A and MYO3B, coded by two different genes, both related to hearing function.

MYO3A is located at the tips of stereocilia of rat and mouse hair cells in the inner ear [9]. MYO3A is thought to regulate the length of cellular actin-rich protrusions in the presence and in the absence of cargos in the inner ear [10] and to initiate and elongate these protrusions in a tip-to-base pattern [11] so that mutations in *MYO3A* are expected to impair mechanotransduction by affecting the activity of the stereocilia [12].

Based on a genome-wide linkage analysis of 15 affected and 8 unaffected members of an Israeli family with post-lingual autosomal recessive hearing loss, Walsh et al. [13], identified a 13 cM region as strongly linked to deafness. This region, which harbors the *MYO3A* gene, was considered a novel recessive deafness locus and named DFNB30. Three different variants in *MYO3A* were found in the Israeli pedigree. The similarity of the product of the gene *MYO3A* with the protein NINAC, which plays an important role in sensory processes in flies [14], and the previous association of other myosin genes with deafness led the authors to conclude that these three variants identified in *MYO3A* were causally related to ARNSHL. Other variants, mostly leading to the loss of function of the protein, have subsequently been described in this gene as causing ARNSHL, with only one exception, a missense variant leading to dominant hearing loss [15].

We previously described the (c.2090T>G;p.(Leu697Trp); NM_017433.5) variant in the *MYO3A* gene segregating with nonsyndromic late-onset progressive hearing loss in 36 affected subjects from two unrelated families (referred as Families 1 and 2 in this study [16]). This variant was the second to be described as associated with autosomal dominant hearing loss (ADNSHL) in this gene, after the report of Grati et al. [15] The p.(Leu697Trp) variant alters one residue in the motor domain of the mutated protein, reducing

ATPase activity, motility, and increasing actin affinity. In COS7 cells, co-expression of MYO3A wild-type (WT) and mutated MYO3A p.(Leu697Trp) resulted in reduced filopodial elongation when compared to cells expressing WT proteins only. Similar effects were observed in hair cells from mouse cochlear explants in the presence and absence of *ESPN-1*, a MYO3A cargo actin-bundling protein. Higher actin affinity and a more stable tipward localization of MYO3A p.(Leu697Trp) in stereocilia in relation to MYO3A WT were also observed, indicating a dominant-negative effect of the mutated protein on the wild type protein [16], which explains dominant inheritance of the phenotype, in contrast to families with recessive transmission, in which usually loss of function variants were detected.

The current Brazilian population is the product of demographic processes that started 500 years ago, through the admixture of Native American populations with the first European settlers, mainly Portuguese. Later, during the sixteenth century, Africans were brought to Brazil as slaves. Starting in the eighteenth century, a large migratory process brought to Brazil other European (mainly Spaniards, Germans, French, Dutch), Middle Eastern (Syrian, Lebanese and Turkish), and Asian immigrants (Japanese), contributing to the complex scenario of admixture in Brazil [17]. This admixture process also introduced many disease-causing genetic alterations inherited from European, African, and Asian immigrants [18, 19]. Besides, a subset of rare variants of autochthonous Native American origin was combined to this inherited background [18, 20]. In this sense, when novel disease-related variants are described in the Brazilian population, it is of interest to define whether the variant is endemic (recently emerged, after the admixture process) or was introduced by a parental population (Native Americans, Europeans or Africans).

The aim of this study was to investigate the frequency, geographic source, and age of the c.2090T>G variant in *MYO3A* in the Brazilian population. In order to achieve this, we screened the variant in a collection of 101 Brazilian pedigrees presenting ADNSHL, without a conclusive molecular diagnosis. Once the variant had been identified in several families, the hypothesis that they have a common origin was tested through kinship analysis and the haplotype sharing approach. In addition, we also estimated the age of the most recent common ancestor sharing the genetic variant.

Material and methods

Patients

A collection of 101 pedigrees with presumptive ADNSHL without a conclusive molecular diagnosis was selected for the screening of the (c.2090T>G;p.(Leu697Trp); NM_017433.5)

variant in the *MYO3A* gene. This variant has already been included as associated with NSHL and deafness in the ClinVar Database (<https://www.ncbi.nlm.nih.gov/clinvar/variation/617675/>), in the Global Variome shared LOVD—Leiden Open Variation Database - Database (<https://databases.lovd.nl/shared/references/PMID:29880844> for families 1 and 2, <https://databases.lovd.nl/shared/individuals> and [https://databases.lovd.nl/shared/references/DOI:10.1038/s41431-00891-0](https://databases.lovd.nl/shared/references/DOI:10.1038/s41431-021-00891-0), for families 3, 4 and 5) and as associated with ADNSHL in The Deafness Variation Database (<https://deafnessvariationdatabase.org/hg19s?terms=10%3A26414513%3A>). All included probands had at least one affected parent with hearing loss. Of these, 56 were ascertained in the Laboratório de Otorrinolaringologia do Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo and 45 were ascertained in the Centro de Pesquisa sobre o Genoma Humano e Células Tronco, Instituto de Biociências, Universidade de São Paulo. Written informed consent was obtained from all participants or guardians. The Institutional Ethics Committee (CEP IBUSP-126/2011 FR.458729) and the Ethics Committee for Analysis of Research Projects (CEP HC-130.695) approved this study.

Audiological evaluation

Affected individuals from 101 Brazilian pedigrees underwent otological and audiological evaluation. Pure tone audiometry, both air (frequencies ranging from 250 to 8000 Hz) and bone conduction (frequencies ranging from 500 to 4000 Hz) identified thresholds. Vocal audiometry included speech reception threshold and speech recognition index. Acoustic immittance measurements were also performed, including tympanometry and acoustic reflexes thresholds.

DNA extraction

Blood samples were collected and DNA was extracted through conventional phenol/chloroform extraction or commercial kits. In some samples, DNA extraction was performed with equipment QiaSymphony (Qiagen, Hilden, Germany) and Autopure LS (Gentra Systems, Minneapolis, USA).

Massive parallel sequencing

DNA samples of the probands from 13 pedigrees presenting ADNSHL (with affected individuals present in at least two consecutive generations) were submitted to massive parallel sequencing of a custom panel of 100 genes designed using the *SureDesign* tool <https://earray.chem.agilent.com/suredesign/>. The DNA libraries were prepared according to *Sure Select QXT Target Enrichment System Kit for Illumina Multiplexed Sequencing* (Agilent, Santa Clara, CA) protocol. Sequencing was performed on an Illumina MiSeq System (San Diego,

CA). Sequences were aligned with the Burrows–Wheeler Aligner (BWA) (<http://bio-bwa.sourceforge.net/>). The Picard tool (<http://broadinstitute.github.io/picard/>) was used to eliminate PCR duplicates, identify indels and realign of reads. Recalibration of base qualities and variant calling were performed using Genome Analysis Tool Kit (GATK) (<https://github.com/broadinstitute/gatk>) and the annotation of variants was performed using Annovar (<http://annovar.openbioinformatics.org/en/latest/>). The strategy of sequencing was the same as described in Dias et al. [21] and further details are in a manuscript in preparation.

For each sample, variants with frequency above 1% in any one of the following databases were discarded: *1000 Genomes* Project (<http://www.1000genomes.org>), *ESP6500* (<http://evs.gs.washington.edu/EVS/>), Exome Aggregation Consortium (*ExAC*) (<http://exac.broadinstitute.org/>) and the Brazilian *ABraOM* (<http://abraom.ib.usp.br>). Next, variants in heterozygosity were selected and genotype–phenotype relationships were evaluated according to their description in *Deafness Variation Database* (<http://deafnessvariationdatabase.org/>).

Sanger sequencing

DNA samples of 88 probands of presumptively ADNSHL pedigrees (with affected individuals present in at least two consecutive generations) and relatives of the probands in the five pedigrees with the c.2090T>G variant were sequenced for the exon 19 in *MYO3A*. Relatives of probands were not analyzed unless the ones in pedigrees in which the *MYO3A* variant was detected. The following primers were used: Forward—5'-GGGAGTGACCTCATTGCCATA-3' and 5'-AGTGGCATTATGAGGTGGAAC-3' (Reverse). PCR products were purified using two enzymes: *Exonuclease I* (GE Healthcare—Little Chalfont, UK) and *Shrimp Alkaline Phosphatase* (GE Healthcare—Little Chalfont, UK). For Sanger sequencing, the purified PCR products were prepared by using BigDye® Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific—Waltham, USA). DNA sequences were analyzed with the ABI 3730 DNA Analyzer® (ThermoFisher Scientific—Waltham, USA).

SNP array genotyping

To investigate the origin of the variant in the five families, samples from 21 affected individuals carrying the variant c.2090T>G were genotyped with the high-density SNP array of Axiom Genome-Wide Human Origins (~600 K SNPs—ThermoFisher Scientific, Santa Clara, CA): seven from family 1 [16], seven from family 2 [16], five from family 3 and the probands from families 4 and 5. Data cleaning was performed according to the manufacturer's recommendation and additional filters were made with the GWASTools R package [22]; see details in [23].

Kinship coefficient, shared haplotype, and genetic variant age

The kinship coefficient was inferred through the Identity by Descent approach (IBD), using the method of moment implemented in the SNPRELATE R package and REAP software [24, 25].

To identify the shared haplotype among individuals from Brazilian families, the high-density SNP array data were phased using SHAPEITv2.r788 [26] through the duo HMM method, which takes into account the information of the reference panel (Project 1000 Genomes phase III) and the pedigrees. This method ensures that the haplotypes are consistent with the pedigree structure and any Mendelian error is reported and corrected during phasing. Finally, we compared the carrier haplotypes of individuals from different families to determine the minimum shared haplotype size.

DNA samples from eight affected individuals of Families 1 and 2 (Family 1-V:4, V:15, V:24, V:34; Family 2- III:22, III:3, IV:8, VI:24) and the Dutch proband were previously submitted to whole-exome sequencing, following the sequencing methods and data cleaning as described by Dantas et al. [16]. However, for this haplotype analyses, we did not use any variant frequency filters based on other datasets, as previously described. This dataset underwent the same approach of phase inference and identification of the shared haplotype, as described above.

The age of the variant *MYO3A* c.2090T>G was estimated using the Gamma method [27], which takes into account the size of the shared haplotype and the decay of linkage disequilibrium over time to infer the time since the most recent common ancestor. This method was specifically developed for the dating of rare variants in small samples, genotyped with high-density SNPs. The analyses were performed using R scripts [24] and were carried out including the Dutch sample and an unrelated 4th degree member of Family 1 (V:4) and one member of Family 2 (III:13).

Genetic ancestry

To determine the geographic origin of the *MYO3A* c.2090T>G variant, the local ancestry of the genomic segment that carries its variant was inferred. As parental populations, we used unrelated individuals from the Human Genetic Diversity Panel (HGDP-CEPH), from the African (Bantu, Biaka, Mandenka, Mbuti pygmy, Mozabite, San and Yoruba), European (Adygey, Basque, French, North Italian, Orcadian, Russian, Sardinian and Tuscan) and Native American (Colombian, Karitiana, Maya, Pima, Surui) groups, genotyped with the same array of SNPs as the present study (ftp://ftp.ceph.fr/hgdp_supp10). Local ancestry was inferred using RFMIX (v.1.5.4; [28]), through a discriminative modeling approach

based on a Random Forest algorithm. We used an admixture model that assumed 20 generations since the beginning of the admixture process and the windows size 0.2 cM. Global ancestry was inferred with ADMIXTURE (v.1.23; [29]), using a Bayesian model-based algorithm with a supervised tri-hybrid population reference. The SNPs were pruned for linkage disequilibrium using a threshold of $r^2 \geq 0.1$ and a 50-SNP window advancing by ten SNPs, as implemented in PLINK (v.1.9; [30]).

Results

Prevalence of the *MYO3A* c.2090T>G:p.(Leu697Trp) variant and its clinical presentation

The presence of a novel missense variant c.2090T>G:p.(Leu697Trp) in *MYO3A* in two Brazilian families with late-onset nonsyndromic autosomal dominant hearing loss (Families 1 and 2 [16]) motivated us to investigate its frequency in an additional collection of 101 pedigrees with probable ADNSHL.

Massive parallel sequencing of a gene panel of 100 genes was performed in samples of 13 hearing impaired probands from unrelated pedigrees with presumptive autosomal dominant inheritance (complete results are described in a manuscript in preparation). In one of the 13 pedigrees, the only probable causative variant was the c.2090T>G:p.(Leu697Trp) in *MYO3A*. For segregation studies, we collected additional 15 samples from this family, six from affected and nine from unaffected subjects (Family 3; Fig. 1a). This variant co-segregated perfectly with the phenotype, being present in all seven affected individuals and absent in all nine unaffected individuals (Fig. 1a).

An additional collection of 88 probands with probable ADNSHL from unrelated pedigrees had their samples screened by Sanger sequencing for the same variant. The *MYO3A* c.2090T>G was present in two probands of this collection (Families 4 and 5; Fig. 1b/c). The only samples available for molecular studies in Families 4 and 5 were those from the probands.

Audiological evaluations of individuals from Families 3, 4, and 5 are shown in Fig. S1. Hearing loss can be described as nonsyndromic sensorineural bilateral and progressive in all of these families, ranging from mild to severe. The age of onset in Family 3 varied from 9 to 20 years old (average of onset was 15). In Family 4, the age of onset varied from 48 to 60 years (average of 54) and in Family 5, the age of onset varied from 14 to 18 years (average of 16).

Summing up, the c.2090T>G:p.(Leu697Trp) variant in *MYO3A* gene explains 3/101 (3%) of cases with presumptive ADNSHL without previous conclusive molecular diagnosis in our cohort.

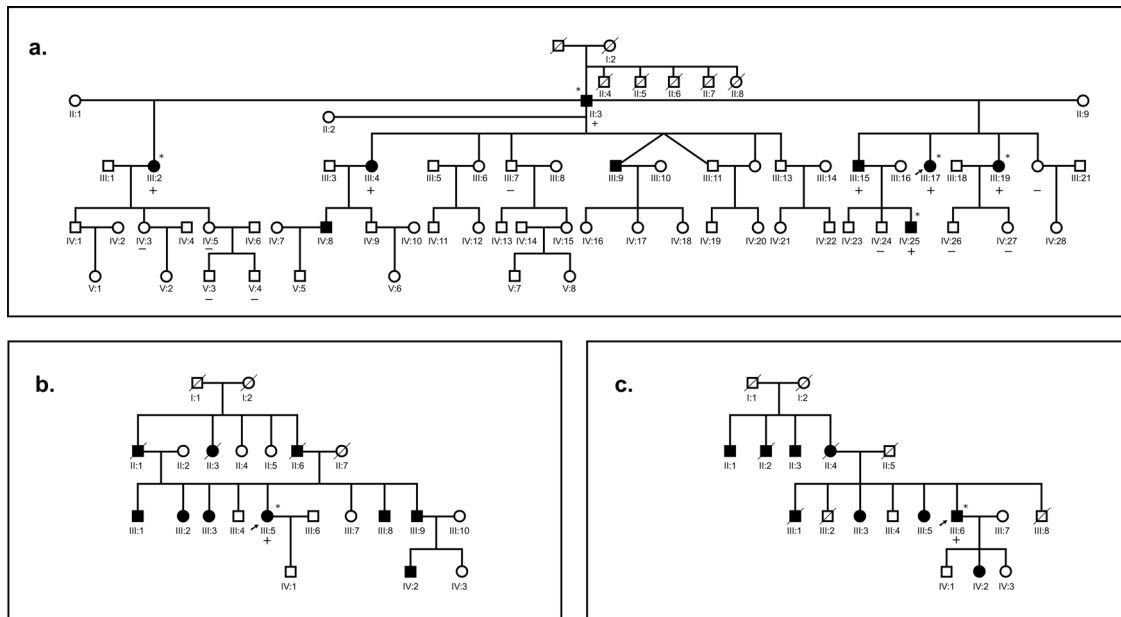


Fig. 1 Segregation of *MYO3A* variant in three pedigrees. Pedigrees from families 3 to 5 showing individuals whose phenotype correlates with the segregation of the c.2090T>G:p.(Leu697Trp) in *MYO3A*. Sanger sequencing was performed in individuals with a plus sign (+) and with a minus sign (-). Individuals indicated with a plus sign are heterozygotes with the c.2090T>G:p.(Leu697Trp) variant; individuals

indicated with a minus sign do not carry the variant. Asterisks indicate the individuals who were genotyped by SNP arrays. The probands are indicated with an arrow. Massive parallel sequencing was performed with the sample from the proband of Family 3. **a** Pedigree of family 3. **b** Pedigree of family 4. **c** Pedigree of family 5.

Relatedness and origin of the *MYO3A* c.2090T>G:p.(Leu697Trp) variant

Among the five families segregating this variant, four (Families 1, 2, 4 and 5) reported coming from the state of Minas Gerais, located in Southeast region of Brazil. Family 3 subjects claimed to come from the state of Paraíba in the Brazilian northeast region. The geographical locations, from which the most ancient known ancestor came, in each pedigree, are shown in Fig. 2.

The presence of the *MYO3A* c.2090T>G:p.(Leu697Trp) in five pedigrees with similar phenotypes and the clustering of 4 families within a geographic region (Fig. 2) led us to hypothesize that this variant had a common origin. To distinguish a scenario of common ancestry from one of distinct mutational events, we investigated the haplotypic context of the c.2090T>G variant. To do so, a subset of 21 affected individuals from the five families was genotyped with a high-density SNP array (Table S3). IBD analysis was performed and 133/210 pairs had kinship coefficient ≥ 0.02 , equivalent to that of second cousins or even more closely related (Table S1). Based on the kinship coefficient, Families 2 and 3 were the most closely related, followed by Families 1 and 3, and Families 1 and 2. The proband from Family 5 is the one with the lowest relatedness with all the others (Fig. 3).

The comparison of haplotypes surrounding the *MYO3A* variant revealed that all 21 genotyped individuals share a 607.098 bp haplotype block (containing 154 genotyped

SNPs). The large shared haplotype indicates that this variant probably arose in a common ancestor to all carriers of the variant (Tables S1 and S2).

Given that the history of Brazilian populations involves extensive admixture, with large contributions from Native American, European, and African individuals, we sought to investigate the ancestry of the carrier individuals and of the chromosomal segment in which the variant was present. On average, the individuals carrying the variant have 80.53% of European ancestry ($\pm 7.35\%$), 9.99% of African ancestry ($\pm 5.68\%$) and 9.47% Native American ancestry ($\pm 6.50\%$). However, when we examined the chromosomal segment carrying the variant using a local ancestry approach, we found that in all individuals the segment is of European ancestry (Fig. 4).

To further investigate if this variant had European origin we searched whether it was present in eight public databases: Deafness Variation Database (version 8.2 - <http://deafnessvariationdatabase.org/>), Clinvar (<http://clinvar.ncbi.nlm.nih.gov/>), ExAC Browser (release 1.0 <http://exac.broadinstitute.org/>), The Human Gene Mutation Database (<http://www.hgmd.cf.ac.uk/ac/index.php>), Database of Genomic Variations (<http://dgv.tcag.ca/dgv/app/home>), Ensemble (<https://www.ensembl.org/index.html>), LOVD (version 3.0 build 22 - <https://databases.lovd.nl/shared/genes>) and CLINVAR (<https://www.ncbi.nlm.nih.gov/clinvar/>). We found that the variant had been reported in the LOVD in an individual of Dutch origin. In fact, the individual in which the variant was detected belongs to a



Fig. 2 Brazilian map showing the geographical origin of the ancestors of each family. Brazilian map highlighting the localization of city of the most ancient known ancestor from families 1–5. In light gray, the states of Paraíba (PB) and Minas Gerais (MG) are highlighted. In dark gray, the Northern region and the *Mata* region of the state of Minas Gerais are indicated. Four of the five ancestors came from the state of Minas Gerais: Family 1-Visconde do Rio Branco; Family 2-Monte Azul; Family 4-Ubá; Family 5-Monte Azul. The only pedigree whose oldest known ancestor came from the Northeastern region of Brazil is also indicated: Family 3-from Piancó, in the state of Paraíba.

family with neurosensorial progressive hearing loss. The Dutch subject reported an age of onset around 12 and needed her first hearing aid at the age of 21. There was also reported that other family members needed hearing aids around the age of 25–40 and experienced progressive HL.

To check whether families share a common ancestor, we compared the haplotype of eight Brazilian samples with the Dutch sample. The samples shared a minimum haplotype of 87,121 bp. Based on linkage disequilibrium decay over the time, we inferred the age of the most recent common ancestor between them in 44.3 generations (CI = 31.5–57.1). Assuming that one generation is equivalent to 25 years, we estimated that this corresponds to approximately 1100 years (CI = 775–1425).

Discussion

The first variant in *MYO3A* identified as causative of ADNSHL (p.Gly488Glu; rs145970949:G>A) was reported by Grati et al. [15], in an African American family. The ancestry of the chromosomal segment containing the mutation was not investigated, so the geographic origin of the mutation remains uncertain.

The variant c.2090T>G:p.(Leu697Trp) was the second variant with dominant transmission described in *MYO3A* gene, described in two families (Families 1 and 2), and the dominant-negative effect was proposed to explain pheno-

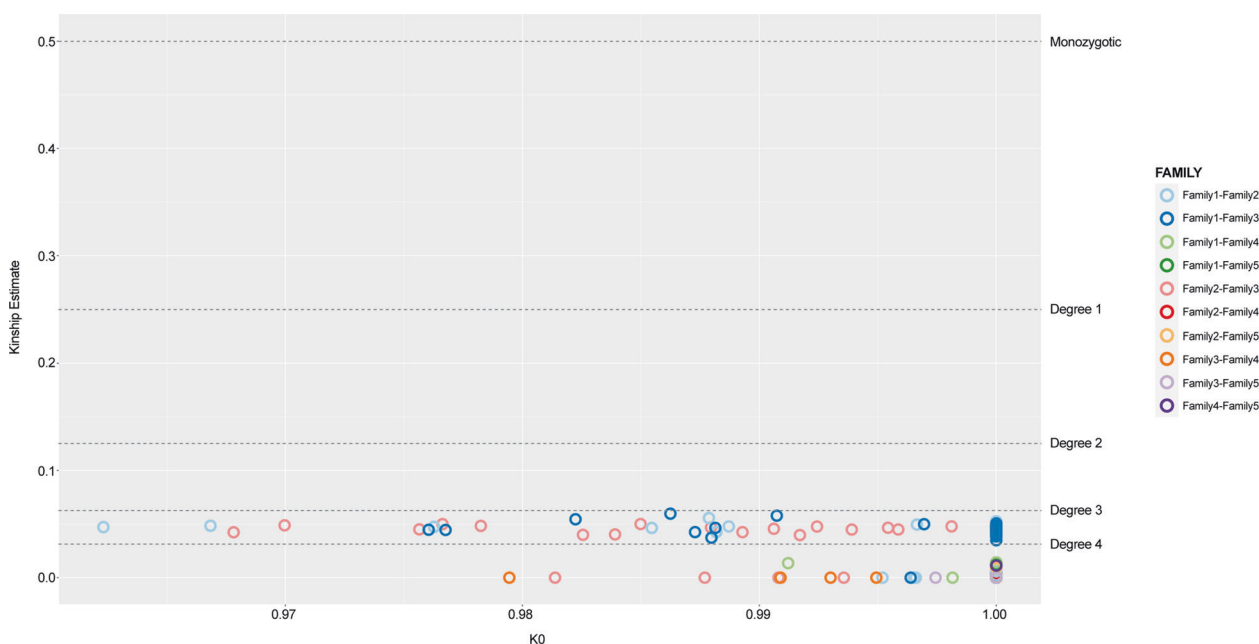


Fig. 3 IBD analysis showing relatedness between 210 pairs of individuals from different families. k_0 is the probability of non-sharing identical-by-descent alleles. The dashed lines correspond to the kinship

thresholds for Monozygotic twins, Degree 1 (parent–offspring; full siblings); Degree 2 (half siblings; uncle–nephew; grandfather–grandson), and Degree 3 (first cousins).

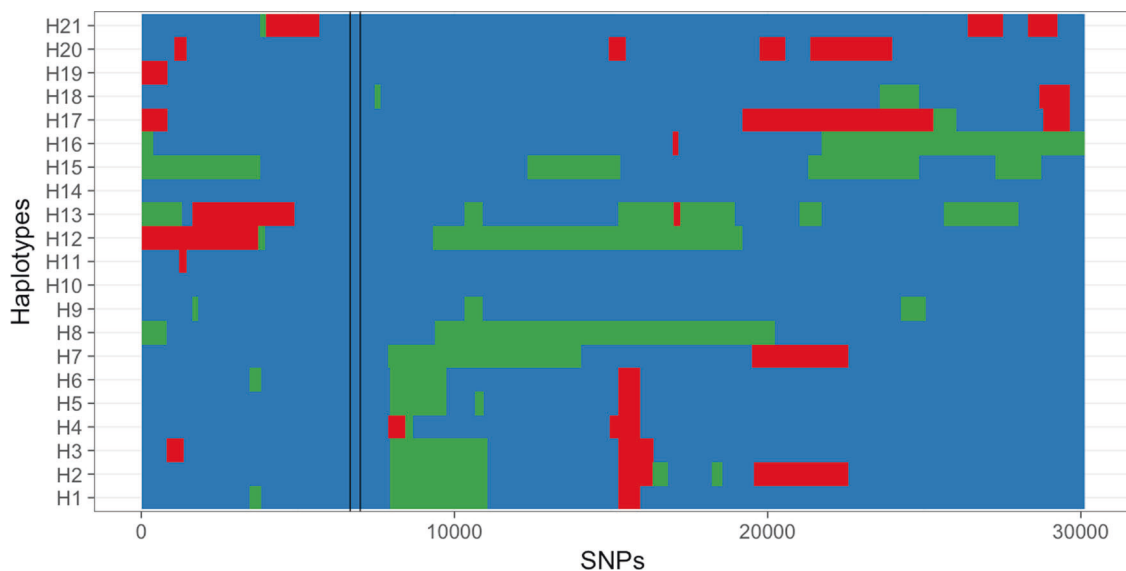


Fig. 4 Variant origin by local ancestry inference. Axis Y represent the phased chromosomes carrying the variant c.2090T>G:p.(Leu697Trp) in *MYO3A* gene. Axis X represents the 30,117 SNPs genotyped on the chromosome 10. The colors, green, blue, and red

represent the African, European, and Native American ancestry, respectively. The vertical black lines point to the variant region, delimiting the minimum shared haplotype.

typic expression in heterozygotes, as indicated by all biochemical data [16].

The c.2090T>G:p.(Leu697Trp) variant in *MYO3A* explained about 3% (3/101) of the Brazilian pedigrees which we analyzed because of presumptive ADNSHL since three additional families were found in this collection (Families 3, 4, and 5). The presence of this variant in five different families that were geographically clustered within Brazil suggested a common origin.

This hypothesis gained strength as we found a 607 kb shared haplotype around the variant when comparing members of the five families. Furthermore, the IBD analysis was performed in pairs of individuals from different pedigrees and identified a high kinship coefficient, between probands and between different pairs of individuals, among the 210 pairs (Table S1). This result suggests at least a common great-grandfather connecting the families, although interviews and careful pedigree inspection in the five families did not allow detection of known linking members between them.

As previously described, most of the families carrying the mutation live in Minas Gerais (Southeastern Brazilian region), a state with a history of attracting immigrants to work on the exploitation of ore. However, Monte Azul (Families 2 and 5), located in *Sertão Mineiro* region, and Visconde do Rio Branco (Family 1), along with Ubá (Family 4), in the *Mata* region (Fig. 2), were not included in the ore exploitation regions, being basically cattle raising and agricultural regions.

Based on the historical records, a plausible dispersal migratory route would be Paraíba (Northeastern Brazilian

region) toward the North of Minas Gerais (Southeastern Brazilian region), and finally toward the *Mata* region. The North of Minas Gerais, also known as *Sertão Mineiro*, (Fig. 2) was settled from the sixteenth century by migratory waves from both the Brazilian Northeastern and Southeastern regions [31]. While the North of Minas Gerais has a history of three centuries of continuous settlement, the *Mata* region (Fig. 2) has a history of more recent occupation, having been populated only in the late eighteenth century, when agricultural expansion attracted immigrants, including from the *Sertão Mineiro*, to work in the countryside [32]. The most likely route of the mutation would be from the Northeastern region toward *Sertão Mineiro* and, later, toward the *Mata* region.

A second question addressed was when and where this genetic variant originated. Native American, African, and European descendants lived in the regions where the five families came from. Ancestry analyses revealed that the individuals carrying the mutation present majority of European ancestry (80.53%). Our local ancestry analysis showed that the mutation was in a shared chromosome segment of European origin. In addition, we found the variant described in the LOVD database in a Dutch individual. We identified that this individual shares a ~87 kb haplotype with individuals from the Brazilian families, suggesting a single origin for the variant and having the estimate of the most recent common ancestor among them ~1100 years.

We contacted other research groups with large samples of pedigrees with ADNSHL in Italy (Dr. Girotto and Dr. Morgan, personal communication) and Spain (Dr. Moreno-

Pelayo, personal communication) and they reported they have never detected this variant, which suggests it is rare in Europe.

From 1630–1654 the Brazilian Northeast, especially the states of Pernambuco (1630–1654), Paraíba, Rio Grande do Norte (1634–1654) and Ceará (1637–1654), were part of the Dutch colony. Historical accounts suggest that around 7000 Dutch soldiers and sailors landed in Brazil in 1630 [33]. Since many of these settlers left descendants in the region, it is possible that the variant originated in the Netherlands and was introduced in Northeastern Brazil during the colonial period and dispersed to the Southeast during the Brazilian agricultural expansion in the eighteenth century.

The finding of rare disease-causing variants in the Brazilian population with clear continental origins is recurrent, and in recent years examples have been found involving SPOAN syndrome (associated to *KLC2* gene mutation, with a European origin) [18], Santos Syndrome (associated to *WNT7A* gene mutation; with a Native American origin) [18], Spinocerebellar ataxia (involving the expansion ATTCT in the *ATXN10*; with Native American origin) [19], and Li-Fraumeni Syndrome (involving *P53* gene; with European origin) [20].

In conclusion, in the present study, we found support for a common origin for the *MYO3A* c.2090T>G:p.(Leu697Trp) variant and determined the time since the most recent common ancestor. The description of the same variant in the Netherlands and the 87 Kb haplotype shared with individuals from Brazilian families point to a likely Dutch origin. All these data, added to the fact that the mutation was present in 3% of ADNSHL families from our cohort, indicate that there are probably many other hearing loss families in Brazil, or maybe outside Brazil, that carry the same variant. The identification of frequent mutations in geographically defined regions is of enormous relevance, because it helps to define more efficient and cheaper strategies of molecular investigation, with immediate applications to the genetic counseling of the families with affected individuals. Our findings reinforce the elevated potential of the Brazilian large families in revealing novel candidate genes or mechanisms of mutation leading to hearing loss.

Acknowledgements The authors are indebted to Humberto C. Marcolino and to Maria Teresa B. M. Auricchio, for technical assistance. We thank Dr. Alfredo Tabith Jr. and all DERDIC staff for clinical assistance and Dr. Jeanne Oiticica for clinical supervision. We deeply thank Dr. Hans K. P. van Amstel, Dr. Rolph Pfundt, for the valuable information about the Dutch proband, Anna Morgan and Dr. Giorgia Giroto for information about Italian pedigrees and Dr. Miguel Angel Moreno Pelayo, about Spanish pedigrees with ADNHL. We are also grateful to FAPESP, CAPES, CNPq and NIH for their financial support. We thank all family members for their participation in the study.

Funding This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - CEPID Human Genome and Stem Cell Research Center 2013/08028-1) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES. KN and DM were supported by United States National Institute of Health (NIH, R01 GM-075091).

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Morton CC, Nance WE. Newborn hearing screening—a silent revolution. *N Engl J Med*. 2006;354:2151–64.
- Shearer AE, Hildebrand MS, Smith RJH. Hereditary hearing loss and deafness overview. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, et al., editors. *GeneReviews*. Seattle (WA): University of Washington; 1993-2019.
- Smith RJ, Bale JF Jr, White KR. Sensorineural hearing loss in children. *Lancet*. 2005;365:879–90.
- Denoyelle F, Marlin S, Weil D, Moatti L, Chauvin P, Garabédian EN, et al. Clinical features of the prevalent form of childhood deafness, DFNB1, due to a connexin-26 gene defect: implications for genetic counselling. *Lancet*. 1999;353:1298–303.
- Friedman TB, Griffith AJ. Human nonsyndromic sensorineural deafness. *Annu Rev Genom Hum Genet*. 2003;4:341–402.
- Raviv D, Dror AA, Avraham KB. Hearing loss: a common disorder caused by many rare alleles. *Ann N Y Acad Sci*. 2010;1214:168–79.
- Kalhammer G, Bähler M. Unconventional myosins. *Essays Biochem*. 2000;35:33–42.
- Montell C, Rubin GM. The *Drosophila* ninaC locus encodes two photoreceptor cell specific proteins with domains homologous to protein kinases and the myosin heavy chain head. *Cell*. 1988;52:757–72.
- Schneider ME, Dosé AC, Salles FT, Chang W, Erickson FL, Burnside B, et al. A new compartment at stereocilia tips defined by spatial and temporal patterns of myosin IIIa expression. *J Neurosci*. 2006;26:10243–52.
- Raval MH, Quintero OA, Weck ML, Unrath WC, Gallagher JW, Cui R, et al. Impact of the motor and tail domains of class III Myosins on regulating the formation and elongation of actin protrusions. *J Biol Chem*. 2016;291:22781–92.
- Salles FT, Merritt RC Jr, Manor U, Dougherty GW, Sousa AD, Moore JE, et al. Myosin IIIa boosts elongation of stereocilia by transporting espin 1 to the plus ends of actin filaments. *Nat Cell Biol*. 2009;11:443–50.
- Walsh VL, Raviv D, Dror AA, Shahin H, Walsh T, Kanaan MN, et al. A mouse model for human hearing loss DFNB30 due to loss of function of myosin IIIA. *Mamm Genome*. 2011;22:170–7.
- Walsh T, Walsh V, Vreugde S, Hertzano R, Shahin H, Haika S, et al. From flies' eyes to our ears: mutations in a human class III myosin cause progressive nonsyndromic hearing loss DFNB30. *Proc Natl Acad Sci USA*. 2002;99:7518–23.
- Dosé AC, Burnside B. Cloning and chromosomal localization of a human class III myosin. *Genomics* 2000;67(Aug):333–42.
- Grati M, Yan D, Raval MH, Walsh T, Ma Q, Chakchouk I, et al. *MYO3A* causes human dominant deafness and interacts with protocadherin 15-CD2 isoform. *Hum Mutat*. 2016;37:481–7.

16. Dantas VGL, Raval MH, Ballesteros A, Cui R, Gunther LK, Yamamoto GL, et al. Characterization of a novel MYO3A missense mutation associated with a dominant form of late onset hearing loss. *Sci Rep.* 2018;8:8706.
17. Salzano FM, Sans M. Interethnic admixture and the evolution of Latin American populations. *Genet Mol Biol.* 2014;37(1 Suppl):151–70.
18. de Farias AA, Nunes K, Lemes RB, Moura R, Fernandes GR, Melo US, et al. Origin and age of the causative mutations in KLC2, IMPA1, MED25 and WNT7A unravelled through Brazilian admixed populations. *Sci Rep.* 2018;8:16552.
19. Paskulin DD, Giacomazzi J, Achatz MI, Costa S, Reis RM, Hainaut P, et al. Ancestry of the Brazilian TP53 c.1010G>A (p.-Arg337His, R337H) founder mutation: clues from haplotyping of short tandem repeats on chromosome 17p. *PLoS ONE.* 2015;10:e0143262.
20. Bampi GB, Bisso-Machado R, Hünemeier T, Gheno TC, Furtado GV, Veliz-Otani D, et al. Haplotype Study in SCA10 families provides further evidence for a common ancestral origin of the mutation. *Neuromolecular Med.* 2017;19:501–9.
21. Dias A, Lezirovitz K, Nicastro FS, Mendes B, Mingroni-Netto RC. Further evidence for loss-of-function mutations in the CEA-CAM16 gene causing nonsyndromic autosomal recessive hearing loss in humans. *J Hum Genet.* 2019;64:257–60.
22. Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I, et al. GWAS Tools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics.* 2012;28:3329–31.
23. Nunes K, Zheng X, Torres M, Moraes ME, Piovezan BZ, Pontes GN, et al. HLA imputation in an admixed population: An assessment of the 1000 Genomes data as a training set. *Hum Immunol.* 2016;77:307–12.
24. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 2012;28:3326–8.
25. Thornton T, Tang H, Hoffman TJ, Ochs-Balcom HM, Baan BJ, Risch N. Estimating kinship in admixed populations. *Am J Hum Genet.* 2012;91:122–38.
26. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2011;9:179–81.
27. Gandolfo LC, Bahlo M, Speed TP. Dating rare mutations from small samples with dense marker data. *Genetics.* 2014;197:1315–27.
28. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet.* 2013;93:278–88.
29. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
30. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
31. Mata-Machado B. História do Sertão Noroeste de Minas Gerais (1690-930). Belo Horizonte: Imprensa Oficial; 1991.
32. Fundação João Pinheiro. História do café das matas de minas (1808-2015). Belo Horizonte: Fundação João Pinheiro; 2018. <http://fjp.mg.gov.br/index.php/docman/dectec-2018/872-cafe-portal-com-capas/file>.
33. Mello EC. O negócio do Brasil. São Paulo: Companhia das Letras; 2011.