A common founder effect of the splice site variant c.-23 + 1G > A in GJB2 gene causing autosomal recessive deafness 1A (DFNB1A) in Eurasia

Aisen V. Solovyev1,2 · Alena Kushniarevich3 · Elena Bliznetz4 · Marita Bady-Khoo5 · Maria R. Lalayants6,7 · Tatiana G. Markova6,7 · Gabriel Minárik8 · L’udevít Kádasi9 · Ene Metspalu3 · Vera G. Pshennikova3 · Tatiana Teryutin1,2 · Elza K. Khusnutdinova1,10,11 · Alexander Poliakov4 · Mait Metspalu3 · Olga L. Posukh5,12 · Nikolay A. Barashkov2 · Sardana A. Fedorova1

Received: 29 June 2021 / Accepted: 21 November 2021 / Published online: 27 November 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
Mutations in the GJB2 gene are known to be a major cause of autosomal recessive deafness 1A (OMIM 220290). The most common pathogenic variants of the GJB2 gene have a high ethno-geographic specificity in their distribution, being attributed to a founder effect related to the Neolithic migration routes of Homo sapiens. The c.-23 + 1G > A splice site variant is frequently found among deaf patients of both Caucasian and Asian origins. It is currently unknown whether the spread of this mutation across Eurasia is a result of the founder effect or if it could have multiple local centers of origin. To determine the origin of c.-23 + 1G > A, we reconstructed haplotypes by genotyping SNPs on an Illumina OmniExpress 730 K platform of 23 deaf individuals homozygous for this variant from different populations of Eurasia. The analyses revealed the presence of common regions of homozygosity in different individual genomes in the sample. These data support the hypothesis of the common founder effect in the distribution of the c.-23 + 1G > A variant of the GJB2 gene. Based on the published data on the c.-23 + 1G > A prevalence among 16,177 deaf people and the calculation of the TMRCA of the modified f2-haplotypes carrying this variant, we reconstructed the potential migration routes of the carriers of this mutation around the world. This analysis indicates that the c.-23 + 1G > A variant in the GJB2 gene may have originated approximately 6000 years ago in the territory of the Caucasus or the Middle East then spread throughout Europe, South and Central Asia and other regions of the world.

Introduction
The GJB2 gene (MIM #121011), which is located in the 13q12.11 chromosome region, encodes the gap junction protein connexin 26 (Cx26). The connexin 26 protein is found in the cochlea in the inner ear and is a major regulator of K+ homeostasis (Dror et al. 2009). In the absence of K+ circulation, the hair cells are unable to generate an action potential in response to sound, which leads to autosomal recessive deafness 1A (DFNB1A) (OMIM #220290). To date, more than 400 different allelic variants have been described in the GJB2 gene (Stenson et al. 2020). Many of them have a high ethno-geographic specificity in their prevalence (Chan and Chang 2014; Tsukada et al. 2015; del Castillo and del Castillo 2017), which is attributed to a founder effect observed in certain ethnic groups (Morell et al. 1998; Van Laer et al. 2001; Shahin et al. 2002; Yan et al. 2003; Balci et al. 2005; Kokotas et al. 2008; Tekin et al. 2010; Barashkov et al. 2011; Dzemileva et al. 2011; Norouzi et al. 2011; Gallant et al. 2013; Carranza et al. 2015; Bliznetz et al. 2017; Zytsar et al. 2018; Erdenechuluun et al. 2018; Posukh et al. 2019; Shinagawa et al. 2020; Zytsar et al. 2020).

The c.-23 + 1G > A splice site mutation (originally named IVS1 + 1G > A) impairs the splicing between the non-coding exon 1 and the coding exon 2 of the GJB2 gene. This mutation was first described in Caucasian populations (Denoyelle et al. 1999; Green et al. 1999). Today, the c.-23 + 1G > A mutation is found in the populations of North and South America, Australia and Europe with an average frequency of 2.4%, 8.6%, 3.8% and 2.3%.
respectively, out of all mutant GJB2 chromosomes (Supplementary Table). c.-23 + 1G > A is the second most frequent mutation of the GJB2 gene in patients with congenital deafness in South Asia (13.2%) after the c.72G > A (p.Trp24*) mutation and in the Middle East and Caucasus populations (20.0%) after the c.35delG (p.Gly12fs) mutation (Supplementary Table). c.-23 + 1G > A is the most common pathogenic GJB2 variant in Central Asia (Mongolia) and Siberia (Tekin et al. 2010; Barashkov et al. 2011; Erdenechuluun et al. 2018).

An extensive accumulation of c.-23 + 1G > A was found in the indigenous Yakut population of Eastern Siberia (Sakha Republic, Russia) with a proportion of 82.7% among all the mutant GJB2 chromosomes in deaf patients (Barashkov et al. 2011). A haplotype analysis of eight STR and two SNP markers showed a common origin of all chromosomes with c.-23 + 1G > A in the Yakut population; the expansion of this mutation was dated to be around 800 years ago (Barashkov et al. 2011). In deaf Mongolian and Turkish patients with the c.-23 + 1G > A mutation, one major common haplotype was also identified (Tekin et al. 2010). A recent study on the reconstruction of haplotypes with the c.-23 + 1G > A mutation revealed one common major haplotype in Yakuts, Russians, Evenks, Tuvinians, Mongols and Turks (Solovyev et al. 2017). However, this study was based on the variability of only nine SNP markers, which was insufficient to support a founder effect hypothesis on the spread of this mutation. Therefore, the issue of defining the migration routes of c.-23 + 1G > A mutation carriers requires a further in-depth study with an increased resolution. It is currently unknown whether this mutation spread across Eurasia as a result of the founder effect or if it had several local centers of origin.

The aim of this work is to determine the origin of the c.-23 + 1G > A mutation and to reconstruct the possible migration routes of its bearers based on ~700,000 SNP data of deaf patients from different regions of Eurasia.

Materials and methods

Samples and genotyping

To increase the reliability in the detection of the mutant haplotype and avoid Phasing errors in this study, we focused only on individuals with a hearing loss homozygous for the c.-23 + 1G > A mutation in the GJB2 gene. To test the hypothesis of the common origin of c.-23 + 1G > A, we collected 23 DNA samples of deaf individuals from different regions of Eurasia (14 Yakuts, 3 Tuvinians, 2 Russians, 1 Evenk, 1 Kumyk, 1 Armenian and 1 Slovak). The anthropological, linguistic and geographical affiliations of the deaf individuals are presented in Table 1. DNA was extracted from blood leukocytes using the phenol–chloroform method. Written informed consent was obtained from all participants of this study.

All 23 individuals were genotyped using an Illumina OmniExpress 730 K array according to the manufacturer’s specifications. The samples collected in Russia were genotyped at the Genoanalytica Company (Moscow, Russia). The missing data were less than 5%. The absence of cryptic relatedness corresponding with the first- and second-degree relatives in our dataset was confirmed using KING software (Manichaikul et al. 2010). A clean dataset of 693,521 autosomal variants only was obtained, including 26,626 variants located in chromosome 13.

PC and ADMIXTURE analyses

To conduct the PC and ADMIXTURE analyses, we combined the newly generated genotypes with the data from previous studies (Li et al. 2008; Behar et al. 2010, 2013; Rasmussen et al. 2010; Metspalu et al. 2011; Yunusbayev et al. 2012; Fedorova et al. 2013; Raghavan et al. 2013). Individuals with missing genotypes greater than 1.5% were excluded from the combined dataset. Only markers

| Table 1 | Anthropological, linguistic and geographical affiliation of 23 deaf patients homozygous for the c.-23+1G>A splice site mutation in GJB2 gene |
|---------|----------------------------------------------------------------------------------------------------------------------------------|
| Population | Number of deaf patients (n = 23) | Anthropological affiliation | Linguistic affiliation | Location (province/country) | Geographical region |
|----------|-------------------------------|--------------------------|---------------------|-----------------------------|---------------------|
| Evenks  | 1                             | Asian                    | Altaic/Tungusic       | Republic of Sakha (Yakutia)/Russia | Eastern Siberia |
| Yakuts   | 14                            | Asian                    | Altaic/Turkic         | Republic of Sakha (Yakutia)/Russia | Eastern Siberia |
| Tuvinians | 3                             | Asian                    | Altaic/Turkic         | Republic of Tyva/Russia      | Southern Siberia   |
| Kumyks   | 1                             | Caucasian                | Altaic/Turkic         | Republic of Dagestan/Russia  | North Caucasus    |
| Armenians | 1                             | Caucasian                | Indo-European/Armenian| Erevan/Armenia               | South Caucasus    |
| Russians | 2                             | Caucasian                | Indo-European/Slavic  | Vladimir oblast/Russia       | Eastern Europe    |
| Slovaks  | 1                             | Caucasian                | Indo–European/Slavic  | Trenčín/Slovakia             | Central Europe    |

© Springer
with a 97% genotyping rate and a minor allele frequency (MAF) > 1% were retained. The marker set was thinned by excluding SNPs in a strong LD (pairwise genotypic correlation \( r^2 > 0.4 \)) in a window of 1000 SNPs, sliding the window by 150 SNPs at a time. The final dataset included 242,229 SNPs presented in 391 individuals from 24 different world populations and in 23 individuals homozygous for c.-23 + 1G > A. The PCA was performed for the autosomal dataset using smartpca software of the EIGENSOFT package (Patterson et al. 2006). We inferred a population structure in our dataset using a model-based clustering method implemented in the ADMIXTURE software (Alexander et al. 2009). We performed the ADMIXTURE assuming \( K = 2 \) to \( K = 10 \) genetic clusters in 100 replicates and assessed the convergence between the individual runs. We discovered that the best clustering solution that met our selection criteria was \( K = 8 \).

### Identification of common chromosomal region

To search for the common chromosomal region linked to c.-23 + 1G > A, we scanned chromosome 13 of each individual \((n = 23)\), starting from this mutation position to the left and right directions until the first heterozygous allele position was reached. We did not consider the physical distances between the SNPs. To identify the longer chromosomal regions common for all the individual genomes in the sample \((n = 23)\), we developed a new method allowing the combination of two neighboring homozygous blocks that were divided by a single heterozygous position. The method minimized the significance of such heterozygous positions if one of the alleles had a majority among all the individual genomes of this position. A full description of this method is presented in Supplementary File 1.

### The modified \( f^2 \)-haplotype method and age dendrogram

To find the haplotypes from the position of the studied mutation, we modified the \( f^2 \)-haplotype analysis (Mathieson and McVean 2014). Instead of taking a random point of a doubleton, we chose the mutation position intentionally (even though it was homozygous). We then conducted pairwise comparisons of the chromosomes of the individuals to the left and right of the c.-23 + 1G > A mutation position until we reached a point with inconsistent homozygous genotypes, which provided us with an (over-)estimate of the distance to the first recombination that broke the haplotype. The approximate age of the c.-23 + 1G > A mutation was calculated by the TMRCA (time to most recent common ancestor) based on the information on the length and number of the non-matching markers in the obtained modified \( f^2 \)-haplotypes.

A full description of this method is presented in Supplementary Files 1 and 2.

### Results

#### Ancestral background

To test the hypothesis of the common origin of the c.-23 + 1G > A mutation, we collected 23 DNA samples of deaf patients from different regions of Eurasia (14 Yakuts, 3 Tuvinians, 2 Russians, 1 Evenk, 1 Kumyk, 1 Armenian and 1 Slovak) (Fig. 1a). To confirm the ethnic identity of the patients, we compared their genetic components with different populations of the world using a PCA analysis (Patterson et al. 2006) and performed a global ancestry inference with ADMIXTURE (Alexander et al. 2009). For the PC and ADMIXTURE analyses, we combined the genotypes of the patients with the data from previous studies (Li et al. 2008; Behar et al. 2010, 2013; Rasmussen et al. 2010; Metspalu et al. 2011; Yunusbayev et al. 2012; Fedorova et al. 2013; Raghavan et al. 2014). Figure 1 shows that all individuals homozygous for c.-23 + 1G > A were clustered according to their ethnic identity and geographic origin (Fig. 1b). The ADMIXTURE analysis \((K = 8)\) confirmed that the ancestral composition of each studied individual corresponded with those of a respective population (Fig. 1b). Thus, the results of both analyses (PC and ADMIXTURE) demonstrated that the genetic profiles of all samples with c.-23 + 1G > A conformed with their geographic location and ethnic origin.

#### Common chromosomal region

In this study, we used a new method that considered the likelihood of allele frequencies by the function of a beta distribution. This analysis detected the most probable extended regions of the linked alleles in the studied samples. It resembled a GWAS analysis but considered the length of the common chromosomal region, making it possible to work on a small sample. In Fig. 2a, it can be seen that the total sample had the longest common region on chromosome 13 \((R = 14.11 \text{ from position } 19,058,717 \text{ to } 21,548,026)\) flanking the c.-23 + 1G > A mutation. Long homozygosity regions were also found on chromosomes 7 and 16 but with lower values \((R = 12.3 \text{ and } R = 12.4, \text{ respectively})\) (Fig. 2a).

A detailed analysis of the homozygosity regions flanking the c.-23 + 1G > A mutation showed that their length varied between the individuals with different ancestral backgrounds (Fig. 2b). Shorter homozygous regions were found in the individuals from Eastern and Central Europe (Russian: 18.9 kb; Slovak: 296.3 kb) and from the Caucasus populations (Armenian: 29.2 kb; Kumyk: 96.6 kb) whereas these regions were longest (2.7–7.6 mb) in the Siberian individuals.
(Tuvinians, Southern Siberia). Despite the differences in the length of the homozygosity regions for each individual, all of them overlapped in the ~5.2 kb identical segment (5 SNPs: rs7329857, rs3751385, rs2274083, rs2274084 and rs7987144). All five SNP markers had an identical allelic composition (GCTCC). We noted that this identical segment (GCTCC) was typical for most of the populations presented in the NCBI database (The 1000 Genomes Project Consortium 2015).

**Mutation age**
We could not use the LD structure to estimate the age of the mutation as it would have been biased by the Yakuts, the major group in our sample. Instead, we used a modified f2-haplotype method (Mathieson and McVean 2014) for the calculation of the time to the most recent common ancestor (TMRCA). This method is suitable for the hypothetical estimation of the closest common ancestor between...
two haplotypes. We determined the lengths of the modified \( f_2 \)-haplotypes by a pairwise comparison of each individual from the position of c.-23 + 1G > A to the position where both individuals had different homozygous alleles (Fig. 3). The samples from the Caucasus (Kum, Arm) and Europe (Rus2) had short \( f_2 \)-haplotypes that were equidistant from the other studied samples and composing the European cluster. Conversely, the longest haplotypes were found in the Siberian cluster (Yak1–2, Yak4–5, Yak8, Yak13, Evk), which also included one Russian sample (Rus1). The intermediate cluster comprised the Slovak (Slv), the Yakut (Yak3, Yak9–10) and the Tuvinian (Tuv1–3) samples where the Tuvinians formed a small group on their own.

Using both the genetic and physical lengths as well as the number of heterozygous positions that our genotyping resolution allowed, we calculated the approximate haplotype age. The TMRCA calculation results based on the modified \( f_2 \)-haplotypes provided an approximate age of the divergence.
of the mutant haplotypes but not of the individuals in general. The obtained results indicated the more distant mutant \( f_2 \)-haplotypes to be in the Caucasus (Kum, Arm) and Europe (Rus2) samples (Fig. 4). The time to the common ancestor for all studied samples was estimated to be around 245 generations ago (if one generation is considered to be 25 years, then the approximate age of the c.-23 + 1G > A mutation is \( \sim 6125 \) years).

**Discussion**

For the reconstruction of the haplotypes with the c.-23 + 1G > A splice site mutation in the \( GJB2 \) gene, we collected DNA samples from 23 deaf individuals with this variant in the homozygous state. The donors of the samples had different anthropological and linguistic affiliations from four regions of Eurasia: Siberia (Yakuts, Tuvinians and Evenk), the Caucasus (Kumyk and Armenian) and Eastern and Central Europe (Russians and Slovak) (Fig. 1a). To confirm the geographic location and ethnic origin reported by the sample donors, as the first stage of our study we performed the PC and ADMIXTURE analyses in the context of worldwide populations. In the PC analysis, all samples with the c.-23 + 1G > A mutation clustered according to their ethnic identity and geographic origin (Fig. 1b). The ADMIXTURE analysis (\( K = 8 \)) confirmed that the ancestral composition of each studied individual corresponded with a respective population (Fig. 1c). The results of both analyses (PC and ADMIXTURE) demonstrated that the genetic profiles of all samples with c.-23 + 1G > A agreed with their geographic location and ethnic origin and excluded the likelihood of a distant relationship between the observed patients (Fig. 1b, c).
For the following analysis, we tested the hypothesis of a common origin of the c.-23 + 1G > A mutation. To do this, we performed a homozygosity analysis on chromosome 13 for each sample starting from the point of mutation. All samples had an identical segment spanning ~ 5.2 kb (from position 20,762,929 to 20,768,144). It was not possible to define the origin of c.-23 + 1G > A by this identical segment as it flanked only five SNP markers, which were similar to most populations in the NCBI database. During the analysis of the chromosomes, we discovered extended homozygosity blocks in all studied genomes that were interrupted by a single heterozygous position (Supplementary File 1). We developed a new method that allowed for the possibility of minimizing the significance of such interruptions by heterozygous positions to identify the longer chromosomal regions common to all individual genomes in the sample (Supplementary File 1). This approach allowed us to locate a long region (~ 2.4 Mb) spanning the studied mutation on chromosome 13 that was shared by all the individual genomes in the sample (Supplementary File 1). This approach allowed us to locate a long region (~ 2.4 Mb) spanning the studied mutation on chromosome 13 that was shared by all the individual genomes in the sample (Fig. 2). These results indicated that all the individual genomes in the studied sample might have inherited chromosomal regions with the c.-23 + 1G > A mutation from one founder chromosome in the distant past. It should be noted though that the results of this new approach might be biased toward the largest group in the sample analyzed, a Yakut population.

The results of the modified f2-haplotype analysis showed that individuals from the populations of the Caucasus and Europe (Kum, Arm and Rus2) carried a greater number of divergent haplotypes than the other studied individuals (Fig. 3). These facts suggest that this mutation spread among the populations of the Caucasus and Europe earlier than in Siberia. Presuming that the mutation had a common origin, we calculated the approximate age of this mutation using the TMRCA approach. The TMRCA calculation based on the modified f2-haplotype showed that the most recent common ancestor with the c.-23 + 1G > A mutation lived not less than ~ 6000 years ago (Fig. 4).

To elucidate the origin of this mutation and reconstruct the possible migration routes of its carriers, we analyzed the published data regarding the prevalence of c.-23 + 1G > A among 16,177 deaf people around the world (Supplementary Table). This analysis showed that c.-23 + 1G > A is present with various frequencies on almost all continents of the world (from 0.6% in East Asia to 58.1% in Siberia and Central Asia), except for Southeast Asia and Sub-Saharan Africa (Supplementary Table). Based on the data of the extensive accumulation of c.-23 + 1G > A, there are three possible geographic regions from which this mutation could begin to spread: Central Asia, the Caucasus or Middle East and South Asia (Fig. 5). It is unlikely that the center of origin was
migration routes of the carriers of c.23 + 1G > A across the world. In the pie charts: a proportion of the c.23 + 1G > A mutation is shown by a red color, the other mutations in the GJB2 gene are shown by a blue color. The black dots represent the countries with information on the proportions of c.23 + 1G > A among mutant GJB2 chromosomes in different world populations and the supposed migration routes are presented in Fig. 5.

It is interesting that the prevalence of c.23 + 1G > A in the world is comparable with the hypothetical routes of the expansion of Indo-European languages (~4000 to ~2500 BC) across Europe, the Middle East, the Caucasus and South Asia (Tassi et al. 2017). The world prevalence of c.23 + 1G > A corresponds with the expansion of populations of a European descent, which are widely presented in North America and Australia, and also had a significant impact on the genetic structure of modern populations of South America and Siberia. The extremely high frequency of c.23 + 1G > A among indigenous populations in Central Asia and Siberia may represent a random population effect (genetic drift and a bottleneck) and can be explained by the genetic structure and demographic events in the history of these populations.

**Conclusion**

Data on ~700,000 SNP markers of 23 deaf individuals from different regions of Eurasia support the hypothesis of a common origin of the c.23 + 1G > A mutation in the GJB2 gene in the studied sample. It presumably originated in the territory of
the Caucasus or Middle East and only then spread throughout Europe, South Asia, Central Asia and Siberia. To determine the age of the c.-23 + 1G > A mutation and the initial center of its occurrence with a greater accuracy, it is necessary to analyze a greater number of samples with this mutation from different populations worldwide with high-resolution methods of the whole genome sequences.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00439-021-02405-w.

Acknowledgements We thank all patients and blood sample donors who have contributed to this study. The authors are grateful to Anatoly N. Alekseev, Chandana Basu Mallik and Georgii P. Romanov for helpful comments on previous versions of this manuscript.

Author contributions Conceptualization: AVS, AK, NAB and SAF. Funding acquisition: SAF, AVS, VGP, OLP, and EM. Methodology: AVS and SAF. Project administration: SAF, NAB, MM, EKH, and AP. Resources: EB, M.B-Kh, MRL, TGM, GM, LM, and FMT. Visualization: AVS. Writing—original draft: AVS and NAB. Writing—review and editing: OLP, AK, and SAF.

Funding This study was supported by the Yakut Science Centre of Complex Medical Problems project: “Study of the genetic structure and burden of hereditary pathology of populations of the Republic of Sakha (Yakutia)”, the Ministry of Science and Higher Education of the Russian Federation (FSRSG-2020-0016), the Russian Foundation for Basic Research (#18-05-00305_Artica, # 20-015-00328_A, #19-34-60023_Perspektiva), the Russian State Budget program No 0259-2021-0012 and IUT 24 0259-2021-0014, the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012) and IUT 24 Research Funding of the Institutional Estonian Ministry of Education and Research.

Code availability Bioinformatic analyses were done with publicly available software packages as described in the methods. Custom codes to aid in variant filtering are available upon request.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval This study was approved by the local Biomedical Ethics Committee at the Yakut Scientific Center of Complex Medical Problems, Yakutsk, Russia (Yakutsk, Protocol No. 50, 24 December 2019).

Consent to participate Informed consent for genetic analyses was obtained from all individual participants included in this study.

Consent for publication All the authors agreed that this manuscript be submitted to the journal of Human Genetics for publication.

References

Abidi O, Boulouiz R, Nahili H, Imken L, Rouba H, Chaïk A, Barakat A (2008) The analysis of three markers flanking GJB2 gene suggests a single origin of the most common 35delG mutation in the Moroccan population. Biochem Biophys Res Commun 377(3):971–974. https://doi.org/10.1016/j.bbrc.2008.10.086

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19(9):1655–1664. https://doi.org/10.1101/gr.094052.109

Balci B, Gerçeker FO, Aksoy S, Sennaroğlu G, Kalay E, Sennaroğlu L, Dinçer P (2005) Identification of an ancestral haplotype of the 35delG mutation in the GJB2 (connexin 26) gene responsible for autosomal recessive non-syndromic hearing loss in families from the Eastern Black Sea Region in Turkey. Turk J Pediatr 47(3):213–221

Barashkov NA, Dzhemileva LU, Fedorova SA, Teryutin FM, Posukh OL, Fedotova EE, Lobov SL, Khusnutdinova EK (2011) Autosomal recessive deafness 1A (DFNB1A) in Yakut population isolate in Eastern Siberia: extensive accumulation of the splice site mutation IVS1+1G>A in GJB2 gene as a result of founder effect. J Hum Genet 56(9):631–639. https://doi.org/10.1038/jhg.2011.72

Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rossset P, Parik J, Rootsi S, Chaupey G, Kutuev I, Yudkovsky G, Khusnutdinova EK, Balanovsky O, Semino O, Pereira L, Comas D, Gurwitz D, Bonne-Tamir B, Parmitt T, Hammer MF, Skorecki K, Vilkens R (2010) The genome-wide structure of the Jewish people. Nature 466(7303):238–242. https://doi.org/10.1038/nature09103

Behar DM, Metspalu M, Baran Y, Kopelman NM, Yunusbayev B, Gladstein A, Tzur S, Sahakyan H, Bahanmehr A, Yepiskoposyan L, Tambets K, Khusnutdinova E, Kushnirevich A, Balanovsky O, Balanovsky E, Kovacevic L, Marjanovic D, Mihalov E, Kouvatsi A, Triantaphyllidis C, King RJ, Semino O, Antonio Torroni A, Hammer MF, Metspalu E, Skorecki K, Rossset S, Herperin E, Villemes R, Rosenberg NA (2013) No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. Hum Biol 85(6):859–900. https://doi.org/10.3378/027.085.0604

Bliznetz EA, Lalayants MR, Markova TG, Balanovsky OP, Balanovskaya EV, Skhalyakhko RA, Peshchekhova EA, Nitkina NV, Voronin SV, Kudyashova EK, Glootov OS, Polyakov AV (2017) Update of the GJB2/DFNB1 mutation spectrum in Russia: a founder Ingush mutation del(GJB2-D13S175) is the most frequent among other large deletions. J Hum Genet 62(8):795–795. https://doi.org/10.1038/jhg.2017.42

Carranza C, Menendez I, Herrera M, Castellanos P, Amado C, Maldonado F, Rosales L, Escobar N, Guerra M, Alvarez D, Foster J 2nd, Guo S, Blanton SH, Bademci G, Tekin M (2015) A Mayan founder mutation is a common cause of deafness in Guatemala. Clin Genet 89(4):461–465. https://doi.org/10.1111/cge.12676

Chan DK, Chang KW (2014) GJB2-associated hearing loss: systematic review of worldwide prevalence, genotype, and auditory phenotype. Laryngoscope 124(2):E34-53. https://doi.org/10.1002/lary.24332

del Castillo FJ, del Castillo I (2017) DFNB1 non-syndromic hearing impairment: diversity of mutations and associated phenotypes. Front Mol Neurosci 10:428. https://doi.org/10.3389/fnmol.2017.00428

Denoyelle F, Marlyn S, Weil D, Moatti D, Chauvin P, Garabédian EN, Petit C (1999) Clinical features of the prevalent form of childhood deafness, DFNB1, due to a Connexin–26 gene defect: implications for genetic counseling. Lancet 353(9161):1298–1303. https://doi.org/10.1016/S0140-6736(98)11071-1

Dror AA, Avraham KB (2009) Hearing loss: mechanisms revealed by genetics and cell biology. Annu Rev Genet 43:411–437. https://doi.org/10.1146/annurev-genet-102108-134135

Dzhemileva LU, Posukh OL, Barashkov NA, Fedorova SA, Teryutin FM, Akhmetova VL, Khusiatova IM, Khussainova RI, Lobov SL, Khusnutdinova EK (2011) Haplotype diversity and reconstruction of ancestral haplotype associated with the c.35delG mutation in the GJB2 (Cx26) gene among the volgo-ural populations of Russia. Acta Naturae 3(3):52–63
Erdenechuluun J, Lin YH, Ganbat K, Bataakhduu D, Makhbal Z, Tsai CY, Lin YH, Chan YH, Hsu CJ, Hsu WC, Chen PL, Wu CC (2018) Unique spectra of deafness-associated mutations in Mongolians provide insights into the genetic relationships among Eurasian populations. PLoS ONE 13(12):e0209797. https://doi.org/10.1371/journal.pone.0209797

Fedorova SA, Reidla M, Metspalu M, Rootsi S, Tambets K, Trofinova N, Zhadanov SI, Kashani BH, Olivieri A, Voevoda MI, Osipova LP, Platono FA, Tomsky MI, Khusnutdinova EK, Torroni A, Villesens R (2013) Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasis. BMC Evol Biol 13:127. https://doi.org/10.1186/1471-2148-13-127

Gallant E, Franchise L, Tsai EA, Berman M, Zhao Y, Feeting H, Kaur M, Deardorff MA, Wilkens A, Clark D, Hakonarson H, Rehm HL, Krantz ID (2013) Homozygosity for the V37I GJB2 mutation in fifteen probands with mild to moderate sensorineural hearing impairment: further confirmation of pathogenicity and haplotype analysis in Asian populations. Am J Med Genet A 161A(9):2148–2157. https://doi.org/10.1002/ajmg.a.36042

Green GE (1999) Carrier rates in the midwestern United States for GJB2 mutations causing inherited deafness. JAMA 281(23):2211. https://doi.org/10.1001/jama.281.23.2211

Kokotas H, Van Laer L, Grigoriadou M, Iliadou V, Economides J, Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM (2014) Demography and the age of rare variants. Nature 488(7410):359–363. https://doi.org/10.1038/nature11335

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM (2008) Worldwide human relationships inferred from genome-wide patterns of variation. Science 319(5866):1100–1104. https://doi.org/10.1126/science.1153717

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26(22):2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Mathieson I, McVean G (2014) Demography and the age of rare variants. PLoS Genet 10(8):e1004528. https://doi.org/10.1371/journal.pgen.1004528

Metspalu M, Romantschuk MG, Yupinshayev B, Chaubey G, Mallick CB, Tkachenko I, Metspalu E, Remm M, Pitchappan R, Singh L, Thangaraj K, Villesens R, Kivisild T (2011) Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. Am J Hum Genet 89(6):731–744. https://doi.org/10.1016/j.ajhg.2011.11.010

Morell RJ, Kim HJ, Hood LJ, Goforth L, Friderici K, Fisher R, Van Camp G, Berlin CI, Oddoux C, Ostrer H, Heats B, Friedman TB (1998) Mutations in the connexin 26 gene (GJB2) among Ashkenazi Jews with nonsyndromic recessive deafness. N Engl J Med 339(21):1500–1505. https://doi.org/10.1056/NEJM199811193392103

Nnorouzi V, Azhi H, Fattahi Z, Esteghamat F, Bazazzadegan N, Nishimura C, Nikzar N, Jalalvand K, Kahrizi K, Smith RJH, Najmabadi H (2011) Did the GJB2 35delG mutation originate in Near Eastern or Middle Eastern ancestry? Genes (basel) 2(3):270–281. https://doi.org/10.3390/genes2030270

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. Plos Genet 2(12):e190. https://doi.org/10.1371/journal.pgen.0020190

Posukh OL, Zytarsk MV, Bady-Khoo MS, Danilchenko VY, Maslova EA, Barashkov NA, Bondar AA, Morozov IV, Maximov VN, Voevoda MI (2019) Unique mutational spectrum of the GJB2 gene and its pathogenic contribution to deafness in Tuvinians (Southern Siberia, Russia): a high prevalence of rare variant c.516G>C (p.Trp172Cys). Genes (basel) 10(6):429. https://doi.org/10.3390/genes10060429

Raghavan M, Skoglund P, Graf K, Metspalu M, Albrectsen A, Moltek I, Rasmussen S, Stafford TW Jr, Orlando L, Metspalu E, Karmin M, Tambets K, Rootsi S, Mägi R, Campos PF, Balanovska E, Balanovsky O, Khusnutdinova E, Litvinov S, Osipova LP, Fedorova SA, Voevoda MI, DeGiorgio M, Sicheritz-Ponten T, Brunnak S, Demeshchenko S, Kivisild T, Villesens R, Nielsen R, Jakobsson M, Williams LE (2013) No evidence from genome-wide data of a Khazar Origin for the Ashkenazi Jews. Hum Biol 85(6):859–900. https://doi.org/10.3378/027.085.0604

RamShankar M, Girirajan S, Dagan O, Ravi Shankar HM, Jalvi R, Ranagaseyy R, Avraham KB, Anand A (2003) Contribution of a new GJB2 mutation and founder effect to non-syndromic hearing loss in India. J Med Genet 40(5):e68. https://doi.org/10.1136/jmg.40.5.e68

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrectsen A, Moltek I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MTP, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson Andrew S, Gledhill Andrew R, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Gronnow B, Meldgaard M, Andreassen C, Fedorova SA, Osipova LP, Higham TFG, Ramsey CB, Tyo H, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Ponten T, Villesens R, Nielsen R, Krogh A, Wang J, Williams LE (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463(7282):757–762. https://doi.org/10.1038/nature08835

Rothrock CR, Murgia A, Sartorato EL, Leonardi E, Wei S, Lebeis SL, Yu LE, Elfenbein JL, Fisher RA, Friderici KH (2003) Connexin 26 35delG does not represent a mutational hotspot. Hum Genet 113(1):18–23. https://doi.org/10.1007/s00439-003-0944-2

Shahin H, Walsh T, Sobe T, Lynch E, King MC, Avraham KB, Kanaan M (2002) Genetics of congenital deafness in the Palestinian population: multiple connexin 26 alleles with shared origins in the Middle East. Hum Genet 110(3):284–289. https://doi.org/10.1007/s00439-001-0674-2

Shinagawa J, Moteki H, Nishio SY, Noguchi Y, Usami SI (2020) Haplotypic analysis of GJB2 mutations: Founder effect or mutational hot spot? Genes (basel) 11(3):250. https://doi.org/10.3390/genes11030250

Solovyev AV, Barashkov NA, Bady-Khoo MS, Zytars MV, Posukh OL, Romanov GP, Raffaïlom AN, Sazonov NN, Alexeey AN, Dzhemileva LU, Khusnutdinova EK, Fedorova SA (2017) Reconstruction of SNP haplotypes with mutation c.-23+1G>A in human GJB2 (Chromosome 13) in some populations of Eurasia. Russ J Hum Genet 5:76–88. https://doi.org/10.1134/S1399000417010014

Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, Hayden M, Heywood S, Millar DS, Phillips AD, Cooper DN (2020) The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. Hum Genet 139(10):1197–1207. https://doi.org/10.1007/s00439-020-01999-3

Tassi F, Vai S, Ghirotto S, Lari M, Modi A, Pili E, Brunelli A, Susca RR, Budnik A, Labuda D, Alberti F, Faluzae-Fox C, Reich D, Caramelli D, Barbujani G (2017) Genome diversity in the Neo-lithic Globular Amphora culture and the spread of Indo-European languages. Proc R Soc B 284(1867):20171540. https://doi.org/10.1098/rspb.2017.1540

Tekin M, Xia XJ, Erdenetungalragul R, Cengiz FB, White RW, Radnaabazar J, Dangaasuren B, Tastan H, Nance WE, Pandya A (2012) GJB2 mutations in Mongolia: complex alleles, low frequency, and reduced fitness of the deaf. Ann Hum Genet 76(2):155–164. https://doi.org/10.1111/j.1469-1809.2010.00640.x
The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. Nature 526(7571):68–74. https://doi.org/10.1038/nature15393

Tsukada K, Nishio S, Hattori M, Usami S (2015) Ethnic-specific spectrum of GJB2 and SLC26A4 mutations: their origin and a literature review. Ann Otol Rhinol Laryngol 124(1 suppl):61S-76S. https://doi.org/10.1177/0003489415575060

Van Laer L, Coucke P, Caethoven G, Flothmann K, Prasad SD, Chamberlin GP, Houseman M, Taylor GR, Van de Heyning CM, Fransen E, Rowland J, Cucci RA, Smith RJ, Van Camp G (2001) A common founder for the 35delG GJB2 gene mutation in connexin 26 hearing impairment. J Med Genet 38(8):515–518. https://doi.org/10.1136/jmg.38.8.515

Yan D, Park HJ, Ouyang XM, Pandya A, Doi K, Erdenetungalag R, Du LL, Matsushiro N, Nance WE, Griffith AJ, Liu XZ (2003) Evidence of a founder effect for the 235delC mutation of GJB2 (connexin 26) in East Asians. Hum Genet 114(1):44–50. https://doi.org/10.1007/s00439-003-1018-1

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

**Authors and Affiliations**

Aisen V. Solovyev1,2 - Alena Kushniarevich3 - Elena Bliznetz4 - Maria Bady-Khoo5 - Maria R. Lalayants6,7 - Tatiana G. Markova6,7 - Gabriel Minárik8 - L’udevít Kádasi9 - Ene Metspalu3 - Vera G. Pshennikova2 - Fedor M. Teryutin1,2 - Elza K. Khusnutdinova1,10,11 - Alexander Poliakov4 - Mait Metspalu3 - Olga L. Posukh5,12 - Nikolay A. Barashkov2 - Sardana A. Fedorova1

1 Laboratory of Molecular Biology, Institute of Natural Sciences, M.K. Ammosov North-Eastern Federal University, Kulakovskogo 46, 677010 Yakutsk, Russia
2 Laboratory of Molecular Genetics, Yakut Science Centre of Complex Medical Problems, Yaroslavskogo 6/3, 677019 Yakutsk, Russia
3 Estonian Biocentre, Institute of Genomics, University of Tartu, Riia 23b, 51010 Tartu, Estonia
4 Research Centre for Medical Genetics, Moskvorechye st., 1, 115478 Moscow, Russia
5 Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Prospekt Lavrentyeva 10, 630090 Novosibirsk, Russia
6 National Research Centre for Audiology and Hearing Rehabilitation, Leninsky prospect 123, 117513 Moscow, Russia
7 Russian Medical Academy of Continuing Professional Education, 2/1, Barikadnaya str, 123995 Moscow, Russia
8 Institute of Molecular Biomedicine, Comenius University Faculty of Medicine, Sasinkova, 4811 08 Bratislava, Slovakia
9 Institute of Molecular Physiology and Genetics, Slovak Academy of Sciences, Dubravskova cesta 9, 840 05 Bratislava, Slovakia
10 Laboratory of Human Molecular Genetics, Institute of Biochemistry and Genetics, Ufa Federal Research Centre of the Russian Academy of Sciences, October Avenue, 71, 450054 Ufa, Russia
11 Department of Genetics and Fundamental Medicine, Bashkir State University, Zaki Validi, 32, 450076 Ufa, Russia
12 Novosibirsk State University, Pirogova st 1, 630090 Novosibirsk, Russia