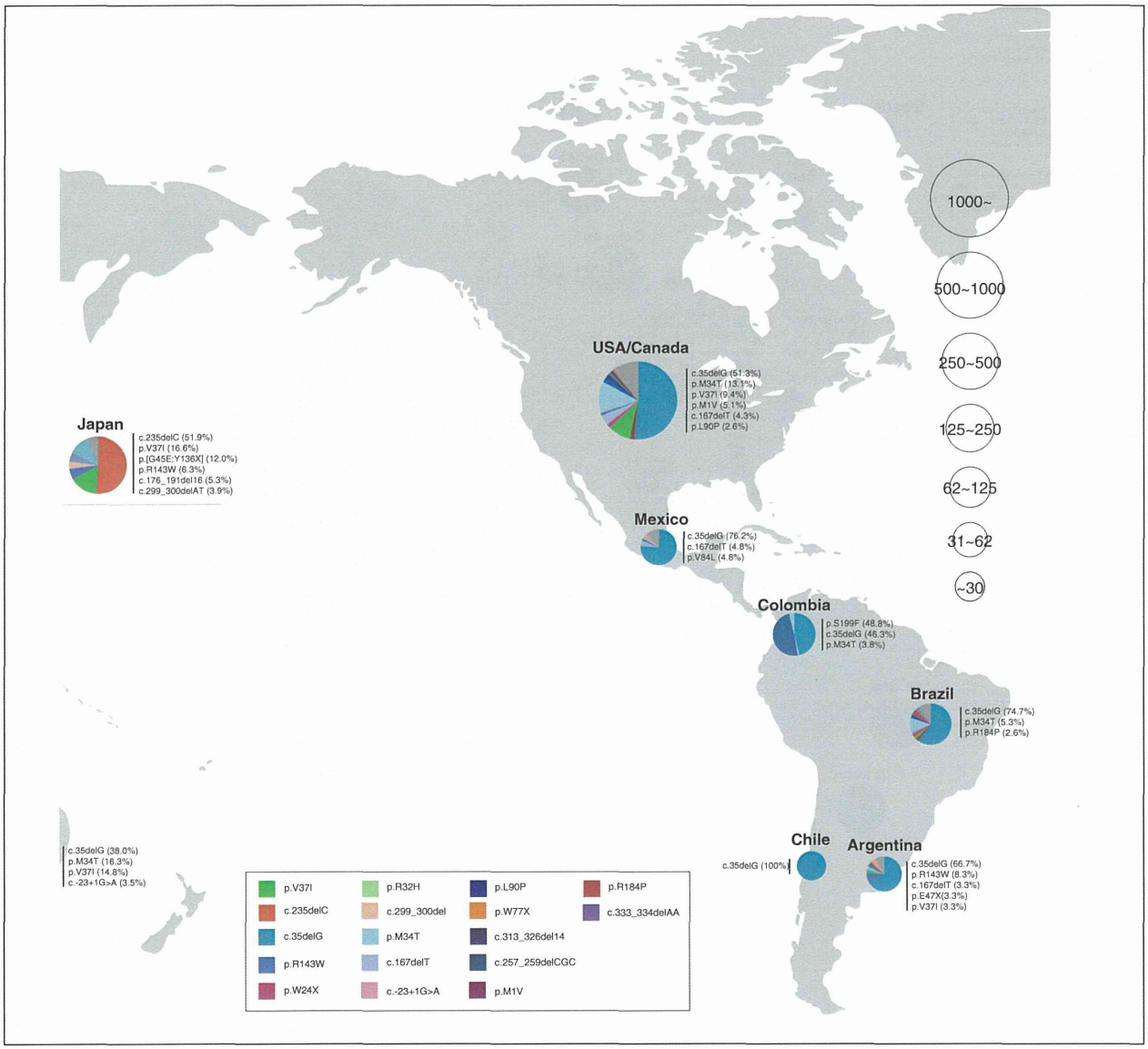

Figure 1. (continued)

European, North African, Middle East Asian, and American populations but rarely in South Asian, Southeast Asian, or East Asian countries (Figure 1). The c.35*del*G mutation is the most or second most prevalent pathogenic *GJB2* mutation in 39 of 52 countries (75%).

Haplotype analysis of the peripheral region of the *GJB2* c.35*del*G mutation indicated a founder effect rather than a mutation hot spot.[75-80] This result suggests that the c.35*del*G mutation was transferred from an ancient ancestor. However, if this mutation occurred in a common ancestor of modern humans, it would be observed in all populations; however, it is absent in Southeast Asian populations. This observation reveals that this mutation occurred in an ancestor of a modern European, Middle East Asian, or North African population. Van Laer et al[75] estimated that the mutation occurred approximately 10 000 years ago and then spread throughout Europe along 2 Neolithic population transportation routes. Dzhemileva et al[79] reported that the root of the c.35*del*G mutation is in Central Asia (Volgo-Ural region of Russia) and was transferred approximately 11 800 years ago. Najmabadi et al[78] reported that the many *GJB2* gene mutations including c.35*del*G were transferred through Iran to Europe by migration and to America by emigration. Indeed, Iranian

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

Table 2. The Spectrum of *SLC26A4* Mutant Alleles.

Country	No. of Screened Patients	No. of Mutant *SLC26A4* Alleles	Variant Alleles, No. (%)																														References
			c.919-2A>G	p.H723R	c.1705+5G>A	c.1001+G>A	p.V138F	p.T416P	p.L236P	p.G209V	p.V239D	p.L445W	p.R409H	p.L597S	p.S90L	p.Q514K	p.T410M	p.Y530H	p.N392Y	p.E384G	p.E29Q	c.1197*del*T	p.M147V	c.1149+3A>G	p.A372V	p.V659L	p.R470H	p.Y375S	p.G672E	c.1341*del*G	p.K715N	c.1652*ins*T	p.R24L	Others	
Japan	100	151	14 (9.3)	77 (51.0)	8 (5.3)	2 (1.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (2.0)	0 (0)	3 (2.0)	0 (0)	0 (0)	0 (0)	2 (1.3)	0 (0)	1 (0.6)	3 (2.0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (2.0)	0 (0)	35 (23.1)	Miyagawa et al, 2014[7]
China	2352	612	377 (61.6)	71 (11.6)	7 (1.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.1)	10 (1.6)	1 (0.1)	0 (0)	0 (0)	12 (2.0)	0 (0)	13 (2.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0.2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	118 (19.3)	Yuan et al, 2012[58]
Korea	56	78	24 (30.7)	47 (60.2)	1 (1.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (2.6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (3.8)	Shin et al, 2012[8]
Taiwan	101	150	115 (76.7)	10 (6.7)	1 (0.6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (2.7)	0 (0)	1 (0.6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (2.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	15 (10)	Wu et al, 2010[59]
Mongolia	135	43	25 (62.5)	4 (7.5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2.3)	0 (0)	0 (0)	3 (6.9)	0 (0)	1 (2.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (4.6)	1 (2.3)	1 (2.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	8 (18.6)	Dai et al, 2008[60]
Pakistan	536	46	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	16 (35.6)	0 (0)	0 (0)	0 (0)	5 (10.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2.2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (4.3)	0 (0)	2 (4.3)	19 (41.3)	Anwar et al, 2009[61]
Iran	80	16	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (12.5)	2 (12.5)	1 (6.0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (6.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	10 (62.5)	Kahrizi et al, 2009[62]
Turkey	333	6	0 (0)	2 (33.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (33.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (33.3)	0 (0)	0 (0)	0 (0)	0 (0)	Tekin et al, 2003[63]
France	100	63	0 (0)	0 (0)	0 (0)	2 (3.2)	2 (3.2)	1 (1.6)	1 (1.6)	9 (14.3)	0 (0)	4 (6.4)	1 (1.6)	0 (0)	0 (0)	0 (0)	1 (1.6)	4 (6.4)	0 (0)	0 (0)	1 (1.6)	0 (0)	4 (6.4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	33 (52.3)	Albert et al, 2006[64]
Germany	3	6	0 (0)	0 (0)	0 (0)	0 (0)	4 (66.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (16.7)	0 (0)	1 (16.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Borck et al, 2003[65]
UK	56	39	0 (0)	0 (0)	0 (0)	8 (20.5)	1 (2.6)	5 (12.8)	9 (23.1)	0 (0)	0 (0)	0 (0)	1 (2.6)	0 (0)	0 (0)	0 (0)	1 (2.6)	1 (2.6)	0 (0)	5 (12.8)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	8 (20.5)	Coyle et al, 1998[66]
Spain	67	36	0 (0)	0 (0)	0 (0)	3 (8.3)	2 (5.6)	2 (5.6)	0 (0)	1 (2.8)	0 (0)	2 (5.6)	1 (2.8)	0 (0)	0 (0)	5 (13.9)	1 (2.8)	0 (0)	0 (0)	0 (0)	1 (2.8)	2 (5.6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	18 (50.0)	Pera et al, 2008[67]
Czech Republic	303	34	0 (0)	0 (0)	0 (0)	2 (5.9)	6 (18.0)	2 (5.9)	1 (2.9)	1 (2.9)	0 (0)	3 (8.8)	2 (5.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	16 (47.1)	Pourova et al, 2010[68]
Denmark	109	157	0 (0)	0 (0)	0 (0)	10 (6.5)	27 (17.3)	32 (20.5)	22 (14.1)	2 (1.3)	1 (0.6)	6 (3.8)	2 (1.3)	3 (1.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	8 (5.1)	10 (6.5)	3 (1.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	31 (20)	Rendtorff et al, 2013[69]
USA	58	32	0 (0)	0 (0)	0 (0)	7 (21.9)	1 (3.1)	5 (15.6)	2 (6.2)	2 (6.1)	0	0 (0)	0 (0)	3 (9.3)	0 (0)	0 (0)	0 (0)	2 (6.1)	0 (0)	0 (0)	1 (3.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (9.3)	0 (0)	0 (0)	0 (0)	0 (0)	4 (12.4)	Campbell et al, 2001[70]
Brazil	23	14	0 (0)	0 (0)	1 (7.1)	1 (7.1)	1 (7.1)	0 (0)	0 (0)	0 (0)	0 (0)	1 (7.1)	2 (14.2)	0 (0)	0 (0)	0 (0)	1 (7.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	7 (50)	de Moraes et al, 2013[71]
Australia	364	46	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2.2)	3 (6.5)	0 (0)	0 (0)	0 (0)	0 (0)	4 (8.7)	4 (8.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (4.4)	0 (0)	0 (0)	3 (6.5)	3 (6.5)	1 (2.2)	0 (0)	0 (0)	0 (0)	0 (0)	25 (54.3)	Dahl et al, 2013[17]

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

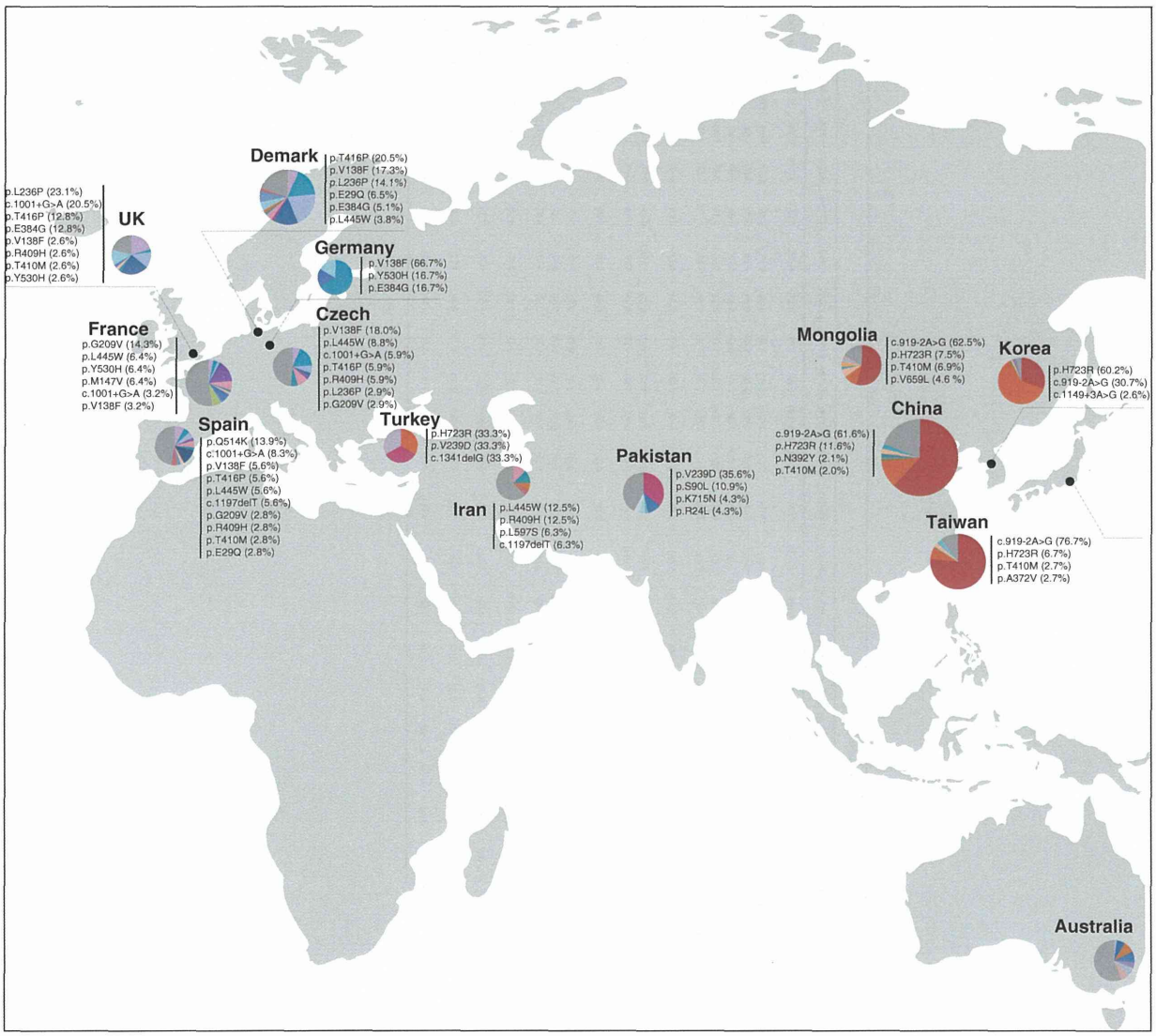

Figure 2. The spectrum of *SLC26A4* mutations.
A larger circle indicates a larger number of mutated alleles.

(continued)

patients with hearing loss carried many moderately frequent *GJB2* mutations (Figure 1).

Lineages associated with haplogroup IJ in the Y-haplotype tree based on Karafet et al[74] (Figure 4) revealed the same distribution pattern as that of the c.35*del*G mutation. Haplogroup I is a major haplotype observed in European populations, and the other European major haplogroup (R) is distributed in Europe, North Eurasia, and India. The *GJB2* c.35*del*G mutation has not been found in any Indian or South Asian populations; therefore, we speculate that the c.35*del*G lineage may be haplogroup I, which was transferred to North Africa and Europe from the Middle East. Haplogroup J was observed in Middle Eastern, North African, and European populations. Based on their analysis, haplogroups I and J were divided into 2 groups from 1 ancestral lineage. Our cluster data indicated that many European, Middle Eastern, and North African populations were characterized by the c.35*del*G mutation and grouped into 1 cluster. This cluster was clearly distinguishable from the other cluster that included the East Asian, South Asian, and Sub-Saharan African populations.

The c.235delC and c.299_300delAT Mutations

The *GJB2* c.235*del*C mutation is 1 of the most common genetic causes of hearing loss in East Asian countries (Japan, Korea, and China) and is observed in low frequency in Middle Eastern and East European countries. The c.299_300*del*AT mutation has a similar distribution to that of c.235*del*C. It is

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

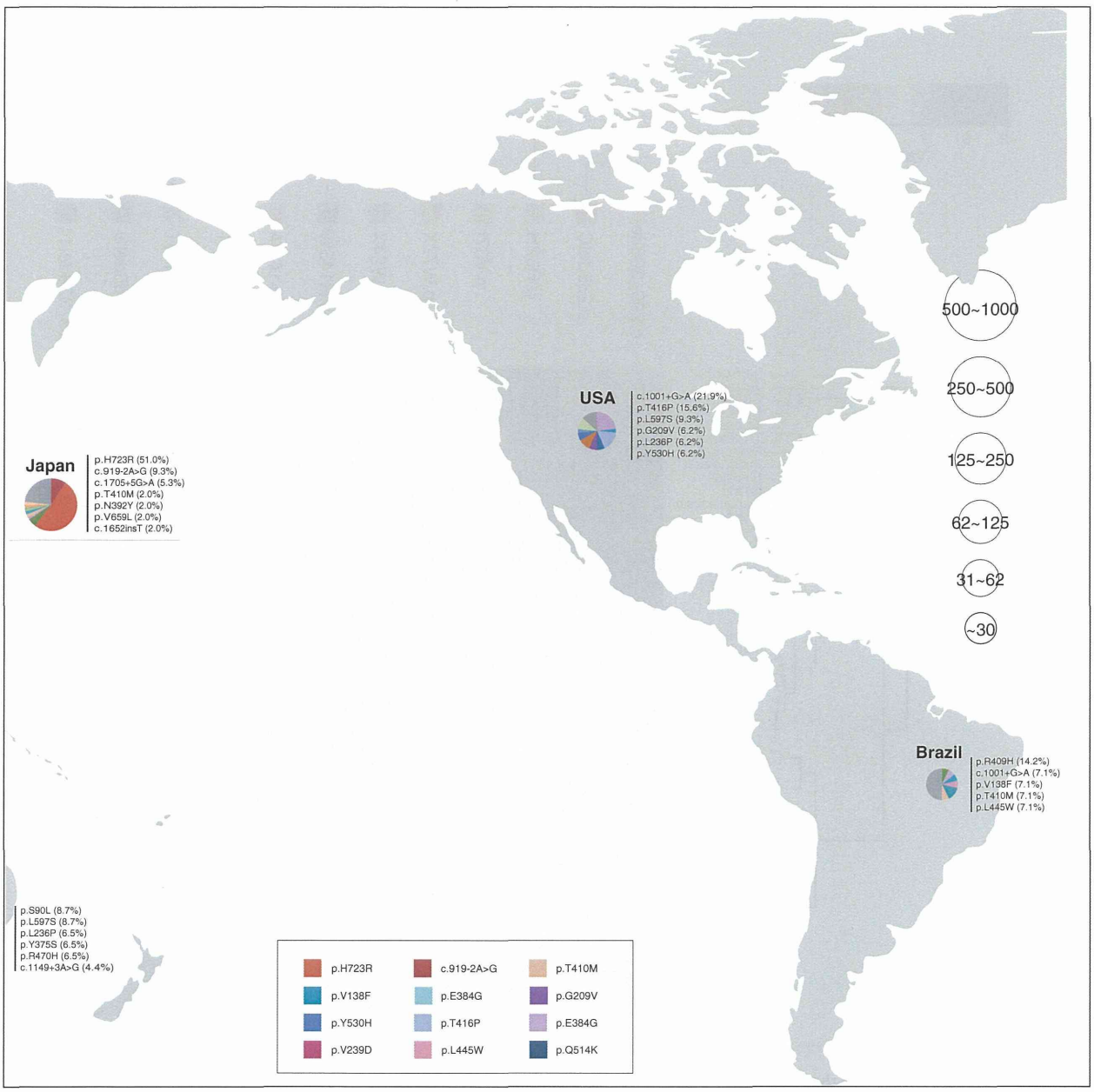

Figure 2. (continued)

interesting that neither of these mutations were observed in Africa. The lineages associated with haplogroup NO* on the Y-haplotype tree are distributed mainly in East Asian countries. Haplogroup O was the most common haplogroup in China, Japan, and Korea and was suspected to derive from the Han-Chinese lineage (Figure 4). Haplogroup N was observed in Siberian and North Eurasian populations. Haplogroups N and O were separated into 2 lineages in Central Asia. Our cluster data showed that East Asian populations (Japan, China, and Korea), which were characterized by the c.235*del*C mutation, were grouped into 1 cluster, and the Mongolian population was located just outside this cluster. The c.235*del*C mutation also arose from a common founder and has a roughly estimated age of approximately 11 500 years.[80,81] Some authors have hypothesized that the c.235*del*C mutation arose near Baikal Lake or the Altai-Sayan region (central region of Eurasia) and then spread into East Asia.[80,82] A previous systematic review and meta-analysis of the c.235*del*C mutation

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

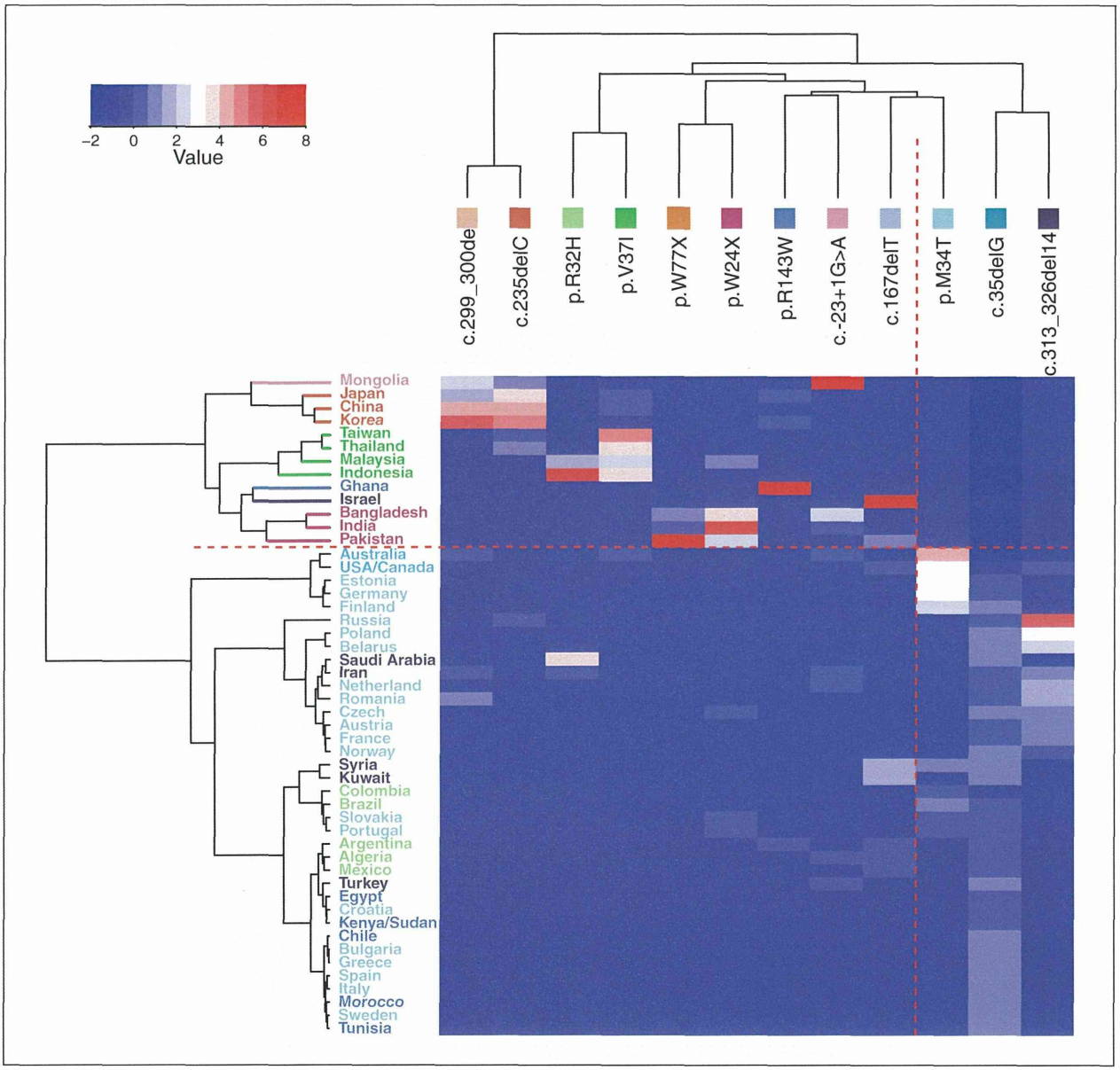

Figure 3. Cluster analysis for the *GJB2* mutations.
A cluster analysis of the *GJB2* mutation allele frequencies shown in Table 1 and Figure 1 was performed by calculating the Euclidean distance and using Ward's clustering method to elucidate the similarities between ethnic populations.

also suggested that the c.235*del*C mutation increases the risk for NSHL in East Asian populations.[83]

The c.-23+1G>A Mutation

c.-23+1G>A was the most common *GJB2* mutation in Mongolia. A previous study also showed a higher prevalence of c.-23+1G>A in the Yakut population of Eastern Siberia located near Mongolia.[84] The c.-23+1G>A mutation was the second most prevalent in India and Bangladesh in the South Asian region. Some patients from the Middle East and Europe and others in a Czech population with the c.-23+1G>A allele had a higher prevalence of this mutation[85]; thus, some migration has occurred through Central Asia to South Asia, the Middle East, and Europe. However, because a noncoding exon mutation analysis is not always included in all reports, it is difficult to estimate the worldwide distribution of this mutation.

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

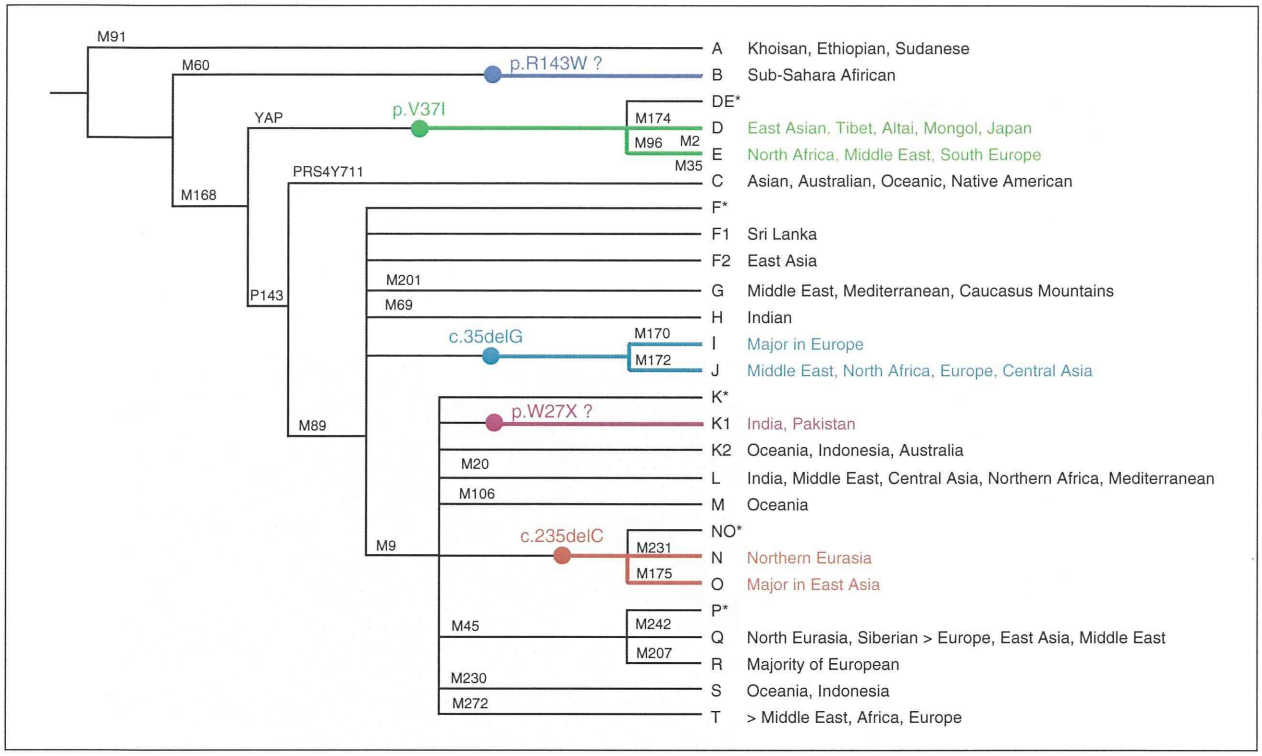

Figure 4. Suspected origin of the *GJB2* mutations.
Suspected origin of *GJB2* mutations are marked on the human Y-chromosomal haplogroup tree, which is applicable to human migration investigations. The *GJB2* mutation clustering results in each ethnic population and the distribution of the *GJB2* mutations were compared to the proposed human migration routes and the distribution presumed from the Y-chromosome haplotype tree.

The p.W24X and p.W77X Mutations

The *GJB2* p.W24X mutation is the most common cause of hearing loss in South Asian countries (India, Pakistan, and Bangladesh) and is also found in Middle Eastern and East European populations at a moderate frequency and at a low frequency in Southeast Asian populations, but it was not observed in African or East Asian populations. A possible founder effect has been reported for this mutation in an Indian population, and the age of the p.W24X mutation was calculated to be 7880 years.[86,87] In contrast, the p.W77X mutation is restricted to South Asian countries (India, Pakistan, and Bangladesh) and Iran. Lineages associated with haplogroup H in the Y-haplotype tree were distributed only in South Asian countries, and haplogroup R was distributed in South and Central Asia and Europe (Figure 4). Haplogroup R has presumed roots in Indo-European language speakers. This observation suggests that the ancestral population of modern Indo-European language speakers carried the p.W24X mutation and that p.W77X occurred in a subgroup of this lineage after migration into South Asian countries. Our cluster analysis indicated that the South Asian populations (India, Pakistan, and Bangladesh) were characterized by the p.W24X and p.W77X mutations and grouped into 1 cluster, similar to the Southeast Asian and East Asian populations; thus, further analyses are required to understand the origins of these mutations.

The p.V37I Mutation

The *GJB2* p.V37I mutation is 1 of the most common causes of hearing loss in Southeast Asian and Oceanic countries (Thailand, Malaysia, Indonesia, and Taiwan). This mutation was also identified at a moderate frequency in Japanese, Chinese, Mongolian, Australian, and American patients with hearing loss. In addition, the *GJB2* p.V37I mutation was also found in North African countries at a moderate frequency (Egypt, Algeria, Morocco, Kenya/Sudan, and Tunisia) and at a low frequency in East European and Middle Eastern countries. It is interesting that this mutation was not observed in Sub-Saharan African (Ghana) or South Asian countries (India, Pakistan, and Bangladesh). The p.V37I mutation is also frequently found in control subjects from these countries (carrier rate: 1.20% in Japanese,[4] 1.35% in Korean,[88] and 1.4% in Chinese[3]). The p.V37I mutation has been frequently associated with mild to moderate hearing loss[4,6] and was first described as a

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

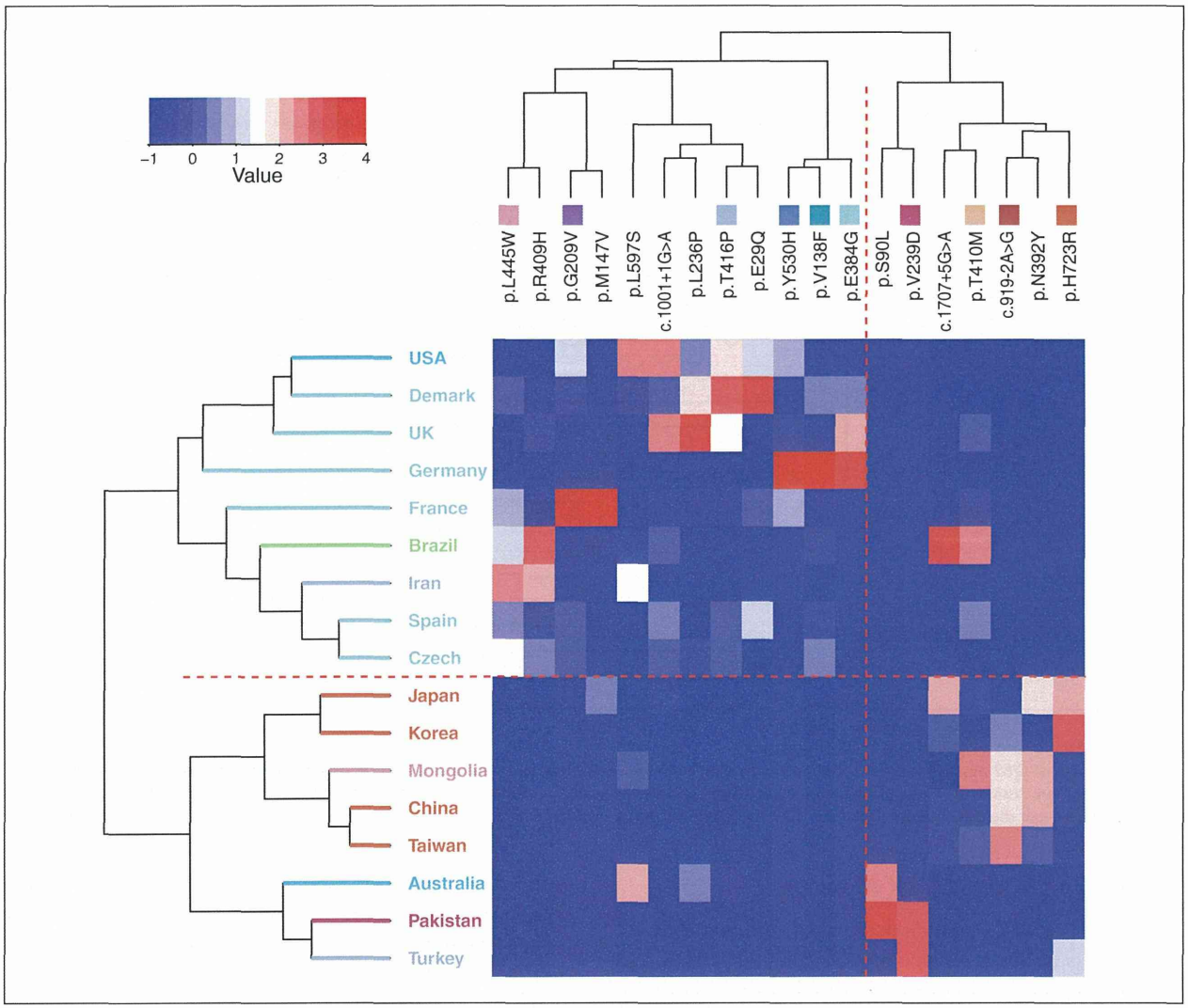

Figure 5. *SLC26A4* mutation clustering analysis.
A cluster analysis of the *SLC26A4* mutation allele frequencies shown in Table 2 and Figure 2 was performed by calculating the Euclidean distance and using Ward's clustering method to elucidate the similarities between ethnic populations.

polymorphism.[89] Later, it was described as a potential pathogenic missense mutation,[90,91] and Bruzzone et al[92] confirmed that the p.V37I mutation can impair channel activity. Although the most recurrent *GJB2* mutations exhibit severe phenotypes, the p.V37I and the p.M34T and p.L90P mutations, which are the second most prevalent mutations in some Caucasian countries, have milder phenotypes.[6] Because most studies of the *GJB2* mutation spectrum include a severe-to-profound cohort, these milder phenotype mutations may not be detected in a deaf study cohort.

Lineages associated with haplogroup DE* in the Y-haplotype tree revealed the same distribution pattern as that of p.V37I (haplogroup DE* was observed in East Asian, North African, Middle Eastern, and South European countries but not in Sub-Saharan African or South Asian countries). In those reports, haplogroup DE* was separated from other populations at an early stage of human migration (Figure 4) and was distinguished from other lineages by carrying haplogroups C and F* (which included haplogroups F-T). In our cluster analysis, Southeast Asian populations, characterized by the p.V37I mutation, were grouped inside the cluster that included East Asian, South Asian, and Sub-Saharan African populations and was distinguishable from other countries. This result also supports the notion that the genetic backgrounds of these populations were distinguishable from those of other populations.

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

The p.R143W Mutation

The *GJB2* p.R143W mutation is 1 of the most common causes of hearing loss in Sub-Saharan African populations (Ghana) and is observed in East Asian, European, and Middle Eastern populations at moderate to low frequencies. It is interesting that this mutation was not observed in Southeast Asian, South Asian, or North African populations. Lineages associated with haplogroup B in the Y-haplotype tree were ancestors of a Sub-Saharan population and could be associated with the origin of the p.R143W mutation; however, this lineage is restricted to Africa and was presumed to be separated at a very early stage of human migration (Figure 4). Thus, it is impossible to explain the high mutation frequency in Ghana and the low to moderate frequencies in East Asian, Middle Eastern, and European populations based on the Y-chromosome lineage. In our cluster results, the Ghanan population was clustered with the East and Southeast Asian populations and distinguishable from North African populations, suggesting that Ghanan and East and Southeast Asian populations separated during an early stage of human expansion from Africa. A possible explanation for this inconsistency in the p.R143W and Y-chromosome distributions may be the occurrence of this mutation in each of the different ancestral lineages. Nevertheless, a haplotype analysis of the region in the vicinity of this mutation in Ghana and other countries is necessary to make this conclusion.

Other Specific Mutations

The c.167*del*T mutation is found in the Ashkenazi Jewish population. Moreover, some specific pathogenic mutations occur in specific areas such as p.S99F in Colombia, c.257_259*del*CGC in Iran and Turkey, and c.313_326*del*14bp in Eastern Europe. In addition, the c.176_191*del*16bp mutation occurs in Japan and China, whereas this mutation is rare in the other regions. The p.G45E/Y136X mutation is the third most prevalent mutation in the Japanese population; however, there are no reports in other countries.

Prevalent SLC26A4 Mutations

The *SLC26A4* mutation was the second most frequent mutation in patients with NSHL. However, there have been only a limited number of studies on the frequency of the *SLC26A4* mutation performed on patients with NSHL compared to those with *GJB2* mutations. Studies on NSHL have revealed biallelic *SLC26A4* mutations in 2% to 3.5% of Caucasian patients,[17,68,93] but in 5.5% to 12.6% of East Asian patients.[58,94-96] The high prevalence (82%-97.9%) of *SLC26A4* mutations in patients with EVA is compatible with the high prevalence of *SLC26A4* mutations reported in East Asians. These frequencies are also higher than those reported in Caucasian populations (20% in the USA,[95] 40.0% in France,[64] and 27% in Spain[67]). Compared to *GJB2* mutations, there have been fewer reports on the mutation spectra and fewer mutated *SLC26A4* alleles identified in Caucasian populations than in East Asian populations.

We summarized the prevalent *SLC26A4* mutations in each ethnic population in Table 2 and Figure 2.

The p.H723R and c.919-2A>G mutations were the most common in the Asian population. p.H723R was predominant in Japan (51.0% frequency) and Korea (60.2% frequency). The frequencies of c.919-2A>G were 61.6%, 76.7%, and 62.5% in China, Taiwan, and Mongolia, respectively. p.V239D was the most frequent mutation in Turkey (33.3%) and Pakistan (35.6%). However, these mutations were not detected in the Caucasian population.

The c.1001+1G>A, p.V138F, p.T416P, p.L236P, and p.G209V mutations were prevalent in the Caucasian population. The c.1001+1G>A mutation was the most or second most prevalent mutation in 4 of 7 European countries and the United States (range, 7.1%-20.5%). The p.V138F mutation was the most prevalent in countries with Caucasian populations and was predominant in Germany (66.7%), Czechoslovakia (18.0%), and Denmark (17.3%). The p.T416P, p.L236P, and p.G209V mutations were mainly found in Denmark (20.0%), the United Kingdom (23.1%), and France (14.3%), respectively, and these mutations were found at a moderate frequency in Europe and the United States. Most of the mutations found in the Caucasian population were not found in the Asian or the Middle Eastern populations. It was evident that the *SLC26A4* mutation spectrum found in the Asian population was quite different from that in the Caucasian population.

Haplotype analyses in previous studies confirmed the founder effect of p.H723R and c.919-2A>G.[94,97] We also performed a cluster analysis of the standardized allele frequencies of the *SLC26A4* mutations to elucidate the similarities between the ethnic populations shown in Table 2 by calculating the Euclidean distance and using Ward's clustering method (Figure 5). The results of the *SLC26A4* cluster analysis were quite similar to those of the *GJB2* cluster analyses for the Japanese, Korean, Chinese, Taiwanese, and Mongolian populations, which are characterized by the p.H723R, p.N392Y, c.919-2A>G, p.T410M, and c.1707+5G>A mutations, and were grouped into 1 cluster. This result clearly indicates the similarities in the genetic backgrounds between the East Asian population and the *GJB2* c.235*del*C distribution. The *GJB2* mutation analysis results indicate that haplogroup NO* may be an ancestor of these mutations. In contrast, most of the European populations were grouped into 1 large cluster (Figure 5). It is interesting that the European populations were divided into 2 clusters at the bottom of the clustering tree. One cluster included Denmark, the United Kingdom, the United States, and Germany, whose populations were characterized by

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

p.E29Q, p.T416P, p.L236P, c.1001+G>A, and p.L597S mutations, and the other included the French, Spanish, Czech, and Iranian populations, characterized by the p.G209V, p.R409H, and p.L445W mutations. This distribution pattern was quite similar to the Y-chromosome haplotype distribution for modern European populations. The ancestor of the northern part of Europe (including Danish, British, and German populations) was presumed to be haplogroup R and that for the southern part of Europe (including many Mediterranean countries) was presumed to be haplotype I. This observation could be the reason for the differences among the 2 groups of European populations.

The p.V239D mutation was the most common mutation in Pakistan and Turkey. It is unfortunate that no reports have described *SLC26A4* mutations in Africa, the roots of humans; thus, future studies are required to define the origin of the mutation and the *SLC26A4* mutation distributions worldwide.

GJB2 *and* SLC26A4 *Mutation Origins*

In this review article, we summarized the 2 major causes of hearing loss, the *GJB2* and *SLC26A4* gene mutation spectra, in many ethnic populations and also performed clustering analysis for the *GJB2* and *SLC26A4* gene mutations. We also performed a comparative analysis between the clustering analysis results and the Y-chromosomal haplogroup analysis results, which revealed human migration routes.

The combination of the results for *GJB2* and *SLC26A4* shows that many mutation distributions are well explained by founder effects in ancient human lineages, as predicted from Y-chromosome haplotype analysis. p.R143W and p.V37I mutations in *GJB2* are spread widely across the globe and are speculated to have occurred at a very early stage in human migration and have been passed down to descendants for a very long time. The p.T410M mutation in *SLC26A4* is also observed in many ethnic populations and may also have occurred at a very early stage in human migration.

Most of the common mutations, such as c.35*del*G, c.235*del*C, and p.W24X of the *GJB2* gene and p.H723R, c.919-2A>G, p.V239D, p.V138F, p.T416P, p.L236P, and p.L445W, are clearly separated into 2 large subgroups: 1 includes the c.235*del*C and p.W24X mutations of *GJB2* and the p.H723R, c.919-2A>G, and p.V239D mutations of *SLC26A4* observed in the East Asian, South Asian, and Southeast Asian populations, whereas the other includes the c.35*del*G mutation of *GJB2* and the p.V138F, p.T416P, p.L236P, and p.L445W mutations of *SLC26A4* observed in the North African, European, Middle Eastern, and North Eurasian populations. This disequilibrium in the distribution of these mutations reveals that these gene mutations occurred after the branching off of each ancestral lineage. It is interesting that many previous reports proposed the origin of these mutations to be in the Middle East or the southern part of Central Asia, areas proposed to contain the roots of many populations belonging to haplogroup F* to T of the Y-chromosome haplogroup.

On the other hand, there are a number of very restricted mutations such as *GJB2*: p.S99F in the Colombian, c.257_259*del*CGC in the Iranian and Turkish, c.313_326*del*14bp in the Eastern European, c.176_191*del*16bp in the Japanese and Chinese, and p.G45E/Y136X in the Japanese populations, as well as many *SLC26A4* mutations. This restricted distribution of these mutations might reflect the fact that these mutations occurred more recently in our ancestors after migration. Haplotype analysis of the region in the vicinity of these mutations is necessary to confirm this conclusion.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by a Health and Labour Sciences Research Grant for Research on Rare and Intractable Diseases [H24-Nanchitou(Nan)-Ippan-032] and Comprehensive Research on Disability Health and Welfare (H25-Kankaku-Ippan-002) from the Ministry of Health, Labour and Welfare, Japan (S.U.), and by a Grant-in-Aid for Scientific Research (A) (22249057) from the Ministry of Education, Culture, Sports, Science and Technology, Japan (S.U.).

References

1. Morton NE. Genetic epidemiology of hearing impairment. *Ann N Y Acad Sci*. 1991;630:16-31.
2. Azaiez H, Chamberlin GP, Fischer SM, et al. *GJB2*: the spectrum of deafness-causing allele variants and their phenotype. *Hum Mutat*. 2004;24:305-311.
3. Dai P, Yu F, Han B, et al. *GJB2* mutation spectrum in 2,063 Chinese patients with nonsyndromic hearing impairment. *J Transl Med*. 2009;7:26.
4. Tsukada K, Nishio S, Usami S; Deafness Gene Study Consortium. A large cohort study of *GJB2* mutations in Japanese hearing loss patients. *Clin Genet*. 2010;78:464-470.
5. Chan DK, Chang KW. *GJB2*-associated hearing loss: systematic review of worldwide prevalence, genotype, and auditory phenotype. *Laryngoscope*. 2014;124:E34-E53.
6. Snoeckx RL, Huygen PL, Feldmann D, et al. *GJB2* mutations and degree of hearing loss: a multicenter study. *Am J Hum Genet*. 2005;77:945-957.
7. Miyagawa M, Nishio SY, Usami S; Deafness Gene Study Consortium. Mutation spectrum and genotype-phenotype correlation of hearing loss patients caused by *SLC26A4* mutations in the Japanese: a large cohort study. *J Hum Genet*. 2014;59:262-268.
8. Shin JW, Lee SC, Lee HK, Park HJ. Genetic screening of *GJB2* and *SLC26A4* in Korean cochlear implantees:

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

experience of Soree Ear Clinic. *Clin Exp Otorhinolaryngol.* 2012;5(suppl 1):S10-S13.
9. Wu CC, Hung CC, Lin SY, et al. Newborn genetic screening for hearing impairment: a preliminary study at a tertiary center. *PLoS One.* 2011;6:e22314.
10. Tekin M, Xia XJ, Erdenetungalag R, et al. *GJB2* mutations in Mongolia: complex alleles, low frequency, and reduced fitness of the deaf. *Ann Hum Genet.* 2010;74:155-164.
11. Snoeckx RL, Djelantik B, Van Laer L, Van de Heyning P, Van Camp G. *GJB2* (connexin 26) mutations are not a major cause of hearing loss in the Indonesian population. *Am J Med Genet A.* 2005;135:126-129.
12. Zainal SA, Md Daud MK, Abd Rahman N, Zainuddin Z, Alwi Z. Mutation detection in *GJB2* gene among Malays with non-syndromic hearing loss. *Int J Pediatr Otorhinolaryngol.* 2012;76:1175-1179.
13. Wattanasirichaigoon D, Limwongse C, Jariengprasert C, et al. High prevalence of V37I genetic variant in the connexin-26 (*GJB2*) gene among non-syndromic hearing-impaired and control Thai individuals. *Clin Genet.* 2004;66:452-460.
14. Padma G, Ramchander PV, Nandur UV, Padma T. *GJB2* and *GJB6* gene mutations found in Indian probands with congenital hearing impairment. *J Genet.* 2009;88:267-272.
15. Santos RL, Wajid M, Pham TL, et al. Low prevalence of connexin 26 (*GJB2*) variants in Pakistani families with autosomal recessive non-syndromic hearing impairment. *Clin Genet.* 2005;67:61-68.
16. Bajaj Y, Sirimanna T, Albert DM, Qadir P, Jenkins L, Bitner-Glindzicz M. Spectrum of *GJB2* mutations causing deafness in the British Bangladeshi population. *Clin Otolaryngol.* 2008;33:313-318.
17. Dahl HH, Ching TY, Hutchison W, Hou S, Seeto M, Sjahalam-King J. Etiology and audiological outcomes at 3 years for 364 children in Australia. *PLoS One.* 2013;8:e59624.
18. Bazazzadegan N, Nikzat N, Fattahi Z, et al. The spectrum of *GJB2* mutations in the Iranian population with non-syndromic hearing loss—a twelve year study. *Int J Pediatr Otorhinolaryngol.* 2012;76:1164-1174.
19. Brownstein Z, Avraham KB. Deafness genes in Israel: implications for diagnostics in the clinic. *Pediatr Res.* 2009;66:128-134.
20. Tekin M, Arici ZS. Genetic epidemiological studies of congenital/prelingual deafness in Turkey: population structure and mating type are major determinants of mutation identification. *Am J Med Genet A.* 2007;143A:1583-1591.
21. Imtiaz F, Taibah K, Ramzan K, et al. A comprehensive introduction to the genetic basis of non-syndromic hearing loss in the Saudi Arabian population. *BMC Med Genet.* 2011;12:91.
22. Al-Sebeih K, Al-Kandari M, Al-Awadi SA, et al. Connexin 26 gene mutations in non-syndromic hearing loss among Kuwaiti patients. *Med Princ Pract.* 2014;23:74-79.
23. Al-Achkar W, Moassass F, Al-Halabi B, Al-Ablog A. Mutations of the connexin 26 gene in families with non-syndromic hearing loss. *Mol Med Rep.* 2011;4:331-335.
24. Posukh O, Pallares-Ruiz N, Tadinova V, Osipova L, Claustres M, Roux AF. First molecular screening of deafness in the Altai Republic population. *BMC Med Genet.* 2005;6:12.
25. Minarik G, Tretinarova D, Szemes T, Kadasi L. Prevalence of DFNB1 mutations in Slovak patients with non-syndromic hearing loss. *Int J Pediatr Otorhinolaryngol.* 2012;76:400-403.
26. Danilenko N, Merkulava E, Siniauskaya M, et al. Spectrum of genetic changes in patients with non-syndromic hearing impairment and extremely high carrier frequency of 35*del*G *GJB2* mutation in Belarus. *PLoS One.* 2012;7:e36354.
27. Teek R, Kruustuk K, Zordania R, et al. Prevalence of c.35*del*G and p.M34T mutations in the *GJB2* gene in Estonia. *Int J Pediatr Otorhinolaryngol.* 2010;74:1007-1012.
28. Wiszniewski W, Sobieszczanska-Radoszewska L, Nowakowska-Szyrwinska E, Obersztyn E, Bal J. High frequency of *GJB2* gene mutations in Polish patients with prelingual nonsyndromic deafness. *Genet Test.* 2001;5:147-148.
29. Sansović I, Knezević J, Musani V, Seeman P, Barisić I, Pavelić J. *GJB2* mutations in patients with nonsyndromic hearing loss from Croatia. *Genet Test Mol Biomarkers.* 2009;13:693-699.
30. Seeman P, Malikova M, Raskova D, et al. Spectrum and frequencies of mutations in the *GJB2* (Cx26) gene among 156 Czech patients with pre-lingual deafness. *Clin Genet.* 2004;66:152-157.
31. Popova DP, Kaneva R, Varbanova S, Popov TM. Prevalence of GBJ2 mutations in patients with severe to profound congenital nonsyndromic sensorineural hearing loss in Bulgarian population. *Eur Arch Otorhinolaryngol.* 2012;269:1589-1592.
32. Radulescu L, Martu C, Birkenhager R, Cozma S, Ungureanu L, Laszig R. Prevalence of mutations located at the dfnb1 locus in a population of cochlear implanted children in eastern Romania. *Int J Pediatr Otorhinolaryngol.* 2012;76:90-94.
33. Janecke AR, Hirst-Stadlmann A, Gunther B, et al. Progressive hearing loss, and recurrent sudden sensorineural hearing loss associated with *GJB2* mutations—phenotypic spectrum and frequencies of *GJB2* mutations in Austria. *Hum Genet.* 2002;111:145-153.
34. Santos RL, Aulchenko YS, Huygen PL, et al. Hearing impairment in Dutch patients with connexin 26 (*GJB2*) and connexin 30 (GJB6) mutations. *Int J Pediatr Otorhinolaryngol.* 2005;69:165-174.
35. Marlin S, Feldmann D, Blons H, et al. *GJB2* and GJB6 mutations: genotypic and phenotypic correlations in a large cohort of hearing-impaired patients. *Arch Otolaryngol Head Neck Surg.* 2005;131:481-487.
36. Zoll B, Petersen L, Lange K, et al. Evaluation of Cx26/*GJB2* in German hearing impaired persons: mutation spectrum and detection of disequilibrium between M34T (c.101T>C) and -493*del*10. *Hum Mutat.* 2003;21:98.
37. D'Andrea P, Veronesi V, Bicego M, et al. Hearing loss: frequency and functional studies of the most common connexin26 alleles. *Biochem Biophys Res Commun.* 2002;296:685-691.
38. Rabionet R, Zelante L, Lopez-Bigas N, et al. Molecular basis of childhood deafness resulting from mutations in the *GJB2* (connexin 26) gene. *Hum Genet.* 2000;106:40-44.
39. Pampanos A, Economides J, Iliadou V, et al. Prevalence of *GJB2* mutations in prelingual deafness in the Greek population. *Int J Pediatr Otorhinolaryngol.* 2002;65:101-108.
40. Nogueira C, Coutinho M, Pereira C, Tessa A, Santorelli FM, Vilarinho L. Molecular investigation of pediatric Portuguese patients with sensorineural hearing loss [published online September 25, 2011]. *Genet Res Int.* doi:10.4061/2011/587602.
41. Carlsson PI, Karltorp E, Carlsson-Hansen E, Ahlman H, Moller C, Vondobeln U. *GJB2* (connexin 26) gene mutations among hearing-impaired persons in a Swedish cohort. *Acta Otolaryngol.* 2012;132:1301-1305.

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

42. Löppönen T, Väisänen ML, Luotonen M, et al. Connexin 26 mutations and nonsyndromic hearing impairment in northern Finland. *Laryngoscope*. 2003;113:1758-1763.
43. Siem G, Fagerheim T, Jonsrud C, et al. Causes of hearing impairment in the Norwegian paediatric cochlear implant program. *Int J Audiol*. 2010;49:596-605.
44. Homøe P, Koch A, Rendtorff ND, et al. *GJB2* (connexin-26) mutations are not frequent among hearing impaired patients in east Greenland. *Int J Audiol*. 2012;51:433-436.
45. Putcha GV, Bejjani BA, Bleoo S, et al. A multicenter study of the frequency and distribution of *GJB2* and GJB6 mutations in a large North American cohort. *Genet Med*. 2007;9:413-426.
46. de la Luz Arenas-Sordo M, Menendez I, Hernandez-Zamora E, et al. Unique spectrum of *GJB2* mutations in Mexico. *Int J Pediatr Otorhinolaryngol*. 2012;76:1678-1680.
47. Batissoco AC, Abreu-Silva RS, Braga MC, et al. Prevalence of *GJB2* (connexin-26) and GJB6 (connexin-30) mutations in a cohort of 300 Brazilian hearing-impaired individuals: implications for diagnosis and genetic counseling. *Ear Hear*. 2009;30:1-7.
48. Tamayo ML, Olarte M, Gelvez N, et al. Molecular studies in the *GJB2* gene (Cx26) among a deaf population from Bogota, Colombia: results of a screening program. *Int J Pediatr Otorhinolaryngol*. 2009;73:97-101.
49. Gravina LP, Foncuberta ME, Prieto ME, Garrido J, Barreiro C, Chertkoff L. Prevalence of DFNB1 mutations in Argentinean children with non-syndromic deafness. Report of a novel mutation in *GJB2*. *Int J Pediatr Otorhinolaryngol*. 2010;74:250-254.
50. Cifuentes L, Arancibia M, Torrente M, Acuna M, Farfan C, Rios C. Prevalence of the 35*del*G mutation in the *GJB2* gene in two samples of non-syndromic deaf subjects from Chile. *Biol Res*. 2013;46:239-242.
51. Snoeckx RL, Hassan DM, Kamal NM, Van Den Bogaert K, Van Camp G. Mutation analysis of the *GJB2* (connexin 26) gene in Egypt. *Hum Mutat*. 2005;26:60-61.
52. Ammar-Khodja F, Faugere V, Baux D, et al. Molecular screening of deafness in Algeria: high genetic heterogeneity involving DFNB1 and the Usher loci, DFNB2/USH1B, DFNB12/USH1D and DFNB23/USH1F. *Eur J Med Genet*. 2009;52:174-179.
53. Abidi O, Boulouiz R, Nahili H, et al. *GJB2* (connexin 26) gene mutations in Moroccan patients with autosomal recessive non-syndromic hearing loss and carrier frequency of the common *GJB2*-35*del*G mutation. *Int J Pediatr Otorhinolaryngol*. 2007;71:1239-1245.
54. Gasmelseed NM, Schmidt M, Magzoub MM, et al. Low frequency of deafness-associated *GJB2* variants in Kenya and Sudan and novel *GJB2* variants. *Hum Mutat*. 2004;23:206-207.
55. Trabelsi M, Bahri W, Habibi M, et al. *GJB2* and GJB6 screening in Tunisian patients with autosomal recessive deafness. *Int J Pediatr Otorhinolaryngol*. 2013;77:714-716.
56. Javidnia H, Carson N, Awubwa M, Byaruhanga R, Mack D, Vaccani JP. Connexin gene mutations among Ugandan patients with nonsyndromic sensorineural hearing loss. *Laryngoscope*. 2014;124:E373-E376.
57. Hamelmann C, Amedofu GK, Albrecht K, et al. Pattern of connexin 26 (*GJB2*) mutations causing sensorineural hearing impairment in Ghana. *Hum Mutat*. 2001;18:84-85.
58. Yuan Y, Guo W, Tang J, et al. Molecular epidemiology and functional assessment of novel allelic variants of *SLC26A4* in non-syndromic hearing loss patients with enlarged vestibular aqueduct in China. *PLoS One*. 2012;7:e49984.
59. Wu CC, Lu YC, Chen PJ, et al. Phenotypic analyses and mutation screening of the *SLC26A4* and FOXI1 genes in 101 Taiwanese families with bilateral nonsyndromic enlarged vestibular aqueduct (DFNB4) or Pendred syndrome. *Audiol Neurootol*. 2010;15:57-66.
60. Dai P, Yuan Y, Huang D, et al. Molecular etiology of hearing impairment in Inner Mongolia: mutations in *SLC26A4* gene and relevant phenotype analysis. *J Transl Med*. 2008;6:74.
61. Anwar S, Riazuddin S, Ahmed ZM, et al. *SLC26A4* mutation spectrum associated with DFNB4 deafness and Pendred's syndrome in Pakistanis. *J Hum Genet*. 2009;54:266-270.
62. Kahrizi K, Mohseni M, Nishimura C, et al. Identification of *SLC26A4* gene mutations in Iranian families with hereditary hearing impairment. *Eur J Pediatr*. 2009;168:651-653.
63. Tekin M, Akcayoz D, Comak E, et al. Screening the *SLC26A4* gene in probands with deafness and goiter (Pendred syndrome) ascertained from a large group of students of the schools for the deaf in Turkey. *Clin Genet*. 2003;64:371-374.
64. Albert S, Blons H, Jonard L, et al. *SLC26A4* gene is frequently involved in nonsyndromic hearing impairment with enlarged vestibular aqueduct in Caucasian populations. *Eur J Hum Genet*. 2006;14:773-779.
65. Borck G, Roth C, Martine U, Wildhardt G, Pohlenz J. Mutations in the PDS gene in German families with Pendred's syndrome: V138F is a founder mutation. *J Clin Endocrinol Metab*. 2003;88:2916-2921.
66. Coyle B, Reardon W, Herbrick JA, et al. Molecular analysis of the PDS gene in Pendred syndrome. *Hum Mol Genet*. 1998;7:1105-1112.
67. Pera A, Villamar M, Vinuela A, et al. A mutational analysis of the *SLC26A4* gene in Spanish hearing-impaired families provides new insights into the genetic causes of Pendred syndrome and DFNB4 hearing loss. *Eur J Hum Genet*. 2008;16:888-896.
68. Pourova R, Janousek P, Jurovcik M, et al. Spectrum and frequency of *SLC26A4* mutations among Czech patients with early hearing loss with and without enlarged vestibular aqueduct (EVA). *Ann Hum Genet*. 2010;74:299-307.
69. Rendtorff ND, Schrijver I, Lodahl M, et al. *SLC26A4* mutation frequency and spectrum in 109 Danish Pendred syndrome/DFNB4 probands and a report of nine novel mutations. *Clin Genet*. 2013;84:388-391.
70. Campbell C, Cucci RA, Prasad S, et al. Pendred syndrome, DFNB4, and PDS/*SLC26A4* identification of eight novel mutations and possible genotype-phenotype correlations. *Hum Mutat*. 2001;17:403-411.
71. de Moraes VC, dos Santos NZ, Ramos PZ, Svidnicki MC, Castilho AM, Sartorato EL. Molecular analysis of *SLC26A4* gene in patients with nonsyndromic hearing loss and EVA: identification of two novel mutations in Brazilian patients. *Int J Pediatr Otorhinolaryngol*. 2013;77:410-413.

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

72. Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res*. 2002;12:339-348.
73. Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet*. 2003;4:598-612.
74. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res*. 2008;18:830-838.
75. Van Laer L, Coucke P, Mueller RF, et al. A common founder for the 35*del*G *GJB2* gene mutation in connexin 26 hearing impairment. *J Med Genet*. 2001;38:515-518.
76. Abidi O, Boulouiz R, Nahili H, et al. 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*. 2008;377:971-974.
77. Kokotas H, Van Laer L, Grigoriadou M, et al. Strong linkage disequilibrium for the frequent *GJB2* 35delG mutation in the Greek population. *Am J Med Genet A*. 2008;146A:2879-2884.
78. Najmabadi H, Nishimura C, Kahrizi K, et al. *GJB2* mutations: passage through Iran. *Am J Med Genet A*. 2005;133A:132-137.
79. Dzhemileva LU, Posukh OL, Barashkov NA, et al. Haplotype diversity and reconstruction of ancestral haplotype associated with the c.35*del*G mutation in the *GJB2* (Cx26) gene among the Volgo-Ural populations of Russia. *Acta Naturae*. 2011;3:52-63.
80. Yan D, Park HJ, Ouyang XM, et al. Evidence of a founder effect for the 235*del*C mutation of *GJB2* (connexin 26) in east Asians. *Hum Genet*. 2003;114:44-50.
81. Ohtsuka A, Yuge I, Kimura S, et al. *GJB2* deafness gene shows a specific spectrum of mutations in Japan, including a frequent founder mutation. *Hum Genet*. 2003;112:329-333.
82. Dzhemileva LU, Barashkov NA, Posukh OL, et al. Carrier frequency of *GJB2* gene mutations c.35*del*G, c.235*del*C and c.167*del*T among the populations of Eurasia. *J Hum Genet*. 2010;55:749-754.
83. Yao J, Lu Y, Wei Q, Cao X, Xing G. A systematic review and meta-analysis of 235*del*C mutation of *GJB2* gene. *J Transl Med*. 2012;10:136.
84. Barashkov NA, Dzhemileva LU, Fedorova SA, et al. Carrier frequency of *GJB2* gene mutations c.35*del*G, c.235*del*C and c.167*del*T among the populations of Eurasia. *J Hum Genet*. 2010;55:749-754.
85. Seeman P, Sakmaryova I. High prevalence of the IVS 1 + 1 G to A/*GJB2* mutation among Czech hearing impaired patients with monoallelic mutation in the coding region of *GJB2*. *Clin Genet*. 2006;69:410-413.
86. RamShankar M, Girirajan S, Dagan O, et al. Contribution of connexin26 (*GJB2*) mutations and founder effect to non-syndromic hearing loss in India. *J Med Genet*. 2003;40:e68.
87. Joseph AY, Rasool TJ. High frequency of connexin26 (*GJB2*) mutations associated with nonsyndromic hearing loss in the population of Kerala, India. *Int J Pediatr Otorhinolaryngol*. 2009;73:437-443.
88. Han SH, Park HJ, Kang EJ, et al. Carrier frequency of *GJB2* (connexin-26) mutations causing inherited deafness in the Korean population. *J Hum Genet*. 2008;53:1022-1028.
89. Kelley PM, Harris DJ, Comer BC, et al. Novel mutations in the connexin 26 gene (*GJB2*) that cause autosomal recessive (DFNB1) hearing loss. *Am J Hum Genet*. 1998;62:792-799.
90. Bason L, Dudley T, Lewis K, et al. Homozygosity for the V37I connexin 26 mutation in three unrelated children with sensorineural hearing loss. *Clin Genet*. 2002;61:459-464.
91. Wilcox SA, Saunders K, Osborn AH, et al. High frequency hearing loss correlated with mutations in the *GJB2* gene. *Hum Genet*. 2000;106:399-405.
92. Bruzzone R, Veronesi V, Gomes D, et al. Loss-of-function and residual channel activity of connexin26 mutations associated with non-syndromic deafness. *FEBS Lett*. 2003;533:79-88.
93. Hutchin T, Coy NN, Conlon H, et al. Assessment of the genetic causes of recessive childhood non-syndromic deafness in the UK—implications for genetic testing. *Clin Genet*. 2005;68:506-512.
94. Park HJ, Shaukat S, Liu XZ, et al. Origins and frequencies of *SLC26A4* (PDS) mutations in east and south Asians: global implications for the epidemiology of deafness. *J Med Genet*. 2003;40:242-248.
95. Dai P, Stewart AK, Chebib F, et al. Distinct and novel *SLC26A4*/Pendrin mutations in Chinese and U.S. patients with nonsyndromic hearing loss. *Physiol Genomics*. 2009;38:281-290.
96. Guo YF, Liu XW, Guan J, et al. *GJB2, SLC26A4* and mitochondrial DNA A1555G mutations in prelingual deafness in Northern Chinese subjects. *Acta Otolaryngol*. 2008;128:297-303.
97. Wu CC, Yeh TH, Chen PJ, Hsu CJ. Prevalent *SLC26A4* mutations in patients with enlarged vestibular aqueduct and/or Mondini dysplasia: a unique spectrum of mutations in Taiwan, including a frequent founder mutation. *Laryngoscope*. 2005;115:1060-1064.

Downloaded from aor.sagepub.com at SHINSHU UNIV MATSUMOTO LIB on May 26, 2015

Annals of Otology, Rhinology & Laryngology
1–12
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0003489415575059
aor.sagepub.com
SAGE

Deafness Gene Variations in a 1120 Nonsyndromic Hearing Loss Cohort: Molecular Epidemiology and Deafness Mutation Spectrum of Patients in Japan

Shin-ya Nishio, PhD[1,2] and Shin-ichi Usami, MD, PhD[1,2]

Abstract

Objectives: To elucidate the molecular epidemiology of hearing loss in a large number of Japanese patients analyzed using massively parallel DNA sequencing (MPS) of target genes.

Methods: We performed MPS of target genes using the Ion PGM system with the Ion AmpliSeq and HiSeq 2000 systems using SureSelect in 1389 samples (1120 nonsyndromic hearing loss cases and 269 normal hearing controls). We filtered the variants identified using allele frequencies in a large number of controls and 12 predication program scores.

Results: We identified 8376 kinds of variants in the 1389 samples, and 409 835 total variants were detected. After filtering the variants, we selected 2631 kinds of candidate variants. The number of *GJB2* mutations was exceptionally high among these variants, followed by those in *CDH23, SLC26A4, MYO15A, COL11A2, MYO7A*, and *OTOF*.

Conclusions: We performed a large number of MPS analyses and clarified the genetic background of Japanese patients with hearing loss. This data set will be a powerful tool to discover rare causative gene mutations in highly heterogeneous monogenic diseases and reveal the genetic epidemiology of deafness.

Keywords

hearing loss, massively parallel DNA sequencing, next-generation DNA sequencer, molecular epidemiology

Introduction

Congenital hearing loss is one of the most common sensory disorders, occurring in 1 of 700 to 1000 newborns. Approximately 50% to 70% of cases are attributable to genetic causes,[1] and 10% to 25% of cases are attributable to congenital cytomegalovirus infection. More than 80 genes have been identified as a cause of hearing loss and an estimated 100 genes are involved in hearing loss.[2]

Despite such advances in gene identification, clinicians and/or geneticists sometimes encounter difficulties related to molecular diagnosis in a clinical setting; for example, the family size is not large enough to allow linkage analysis, meaning that only limited familial information for predicting the causative gene is available. In such cases, targeted exon sequencing of selected genes using massively parallel DNA sequencing (MPS) technology will potentially enable us to systematically tackle previously intractable monogenic disorders and improve molecular diagnosis.

An increasing number of articles regarding gene discovery and successful clinical application for the identification of genes responsible for deafness using MPS have recently been published.[3-16] We applied MPS technology to (1) discover causative mutations in relatively rare causative genes[12,13] and (2) clarify the molecular epidemiology.[12] Our results demonstrated that MPS-based screening is powerful in terms of identifying mutations in rare causative genes, and from an epidemiological view point, *GJB2* mutations are involved in 30% to 40% cases of deafness, while the remaining cases of hearing loss arise from various rare genes/mutations that were not easy to identify using the conventional one-by-one screening approach.

For clinical application to genetic heterogeneous diseases, systemic screening of known genes in a cost-effective manner is required. Hybridization-based capture is commonly used for genomic target enrichment, but for clinical application,

[1]Department of Otorhinolaryngology, Shinshu University School of Medicine, Matsumoto, Japan
[2]Department of Hearing Implant Sciences, Shinshu University School of Medicine, Matsumoto, Japan

Corresponding Author:
Shin-ichi Usami, Department of Otorhinolaryngology, Shinshu University School of Medicine, 3-1-1 Asahi, Matsumoto 390-8621, Japan.
Email: usami@shinshu-u.ac.jp

polymerase chain reaction (PCR)–based technologies in combination with MPS have also been proposed.[11,13,16,17]

In the current study, on the basis of our PCR-based technologies in combination with MPS,[13,17] we increased the number of patients (1120 cases of nonsyndromic hearing loss) to establish a database for clinical molecular diagnosis and to confirm the molecular epidemiology of deafness. Data analysis concerning diagnostic sensitivity and specificity, which is important for clinical application, was also performed.

Subjects and Methods

Subjects

A total of 1120 Japanese patients (266 autosomal dominant or mitochondrial inheritance cases, 600 autosomal recessive inheritance or sporadic cases, and 253 unknown family history cases) with bilateral nonsyndromic sensorineural hearing loss from 53 ear, nose, and throat departments nationwide participated in this study. In addition, 269 normal hearing controls, confirmed by pure-tone audiometry, were also enrolled. Informed written consent was obtained from all subjects, their next of kin, caretakers, or guardians (in the case of minors) prior to participation. This study was approved by the Shinshu University Ethical Committee and the ethics committees of all other participating institutions listed in the Acknowledgments.

Genetic Analysis

We performed the MPS analysis using an Ion PGM with Ion AmpliSeq for 1174 samples (905 hearing loss cases and 269 normal hearing controls) and using HiSeq 2000 with SureSelect in 215 cases.

Amplicon Library Preparation and Ion PGM Platform Sequencing

Amplicon libraries of the target gene exons from 63 genes reported to cause nonsyndromic hearing loss[2] were prepared with an Ion AmpliSeq Custom Panel (Life Technologies, Foster City, California, USA). These libraries were designed with an Ion AmpliSeq Designer (Life Techologies), and amplicon libraries were prepared using an Ion AmpliSeq Library Kit 2.0 and an Ion Xpress Barcode Adapter 1-96 Kit (Life Technologies) according to the manufacturer's instructions. After the amplicon libraries were prepared, they were diluted to 20 pM, and the same amount of libraries from the 6 libraries of 6 patients were pooled for 1 sequence reaction. The emulsion polymerase chain reaction and sequencing were performed with an Ion Torrent Personal Genome Machine (PGM) system using an Ion PGM 200 Sequencing Kit and an Ion 318 Chip (Life Technologies) according to the manufacturer's instructions. The detailed protocol has been described elsewhere.[13,17] The sequence data were mapped to the human genome sequence (build GRCh37/hg19) with the Torrent Mapping Alignment Program. After sequence mapping, the DNA variant regions were piled up with the Torrent Variant Caller plug-in software version 4.0 (Life Technologies).

Targeted Enrichment and HiSeq Platform Sequencing

The SureSelect target enrichment kit, designed for the 112 potentially deaf-causing genes, including the 63 genes reported to cause nonsyndromic hearing loss, the 22 genes reported to cause syndromic hearing loss, and the 36 genes highly expressed in the adult human inner ear by microarray analysis, was used in this study.[18] The detailed gene list is described in our previous report.[12] A 3-μg DNA aliquot was fragmented using the Covaris S2 System (Covaris, Woburn, Massachusetts, USA) to a fragment length of about 200 bp. Furthermore, the target regions were enriched using the SureSelect Target DNA Enrichment kit with a barcode adapter (Agilent Technologies, Santa Clara, California, USA) according to the manufacturer's protocol. The same amount of libraries from each of 12 patients was pooled into 1 tube and analyzed in 1 lane of the Illumina HiSeq 2000 sequencer (Illumina, San Diego, California, USA) according to the manufacturer's protocol. The sequence data were processed by filtering the read quality to QV = 30 as a cutoff and duplicate reads removed. After the filtering process, sequence reads were mapped to the human genome sequence (build GRCh37/hg19) using BWA software.[19] After sequence mapping, the DNA variant regions were piled up with GATK software.[20]

Filtering Detected Variants

After detecting the variants, the effects of the variants were analyzed using ANNOVAR software.[21,22] The missense, nonsense, insertion/deletion, and splicing variants were selected among the identified variants. Variants were further selected as <1% of: (1) the 1000 genome database,[23] (2) the 6500 exome variants,[24] (3) the human genetic variation database (data set for 1208 Japanese exome variants),[25] (4) the 269 in-house Japanese normal hearing controls, and (5) 1000 control data in the deafness variation database.[26] The filtering process is shown in Figure 1.

Results and Discussion

DNA Sequencing Metrics and Accuracy of Each Sequencing System

MPS metrics used in this study are summarized in Supplemental Table 1 (available in the online journal). The

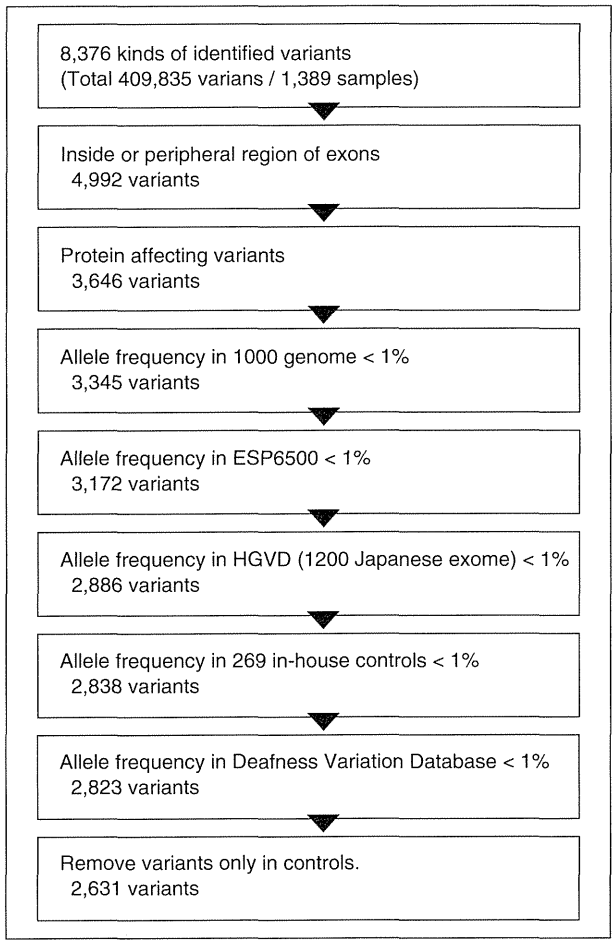

Figure 1. Algorithm applied in this study. The nonsense, splice-site, insertion-deletion, and missense variants were chosen according to this algorithm.

mean depth of coverage of the target region for 1174 samples analyzed by the Ion PGM sequencer was 284.3 ± 94.5× (range, 690.0-96.6×). The percentage of each region with more than 20× coverage (indicating the percentage of each region sequenced 20 times or more by MPS) was 97.6% ± 0.9% (range, 93.1%-99.2%). To reduce the risk of incorrect genotyping and missed *single nucleotide polymorphisms* (SNPs) in poor-coverage regions, we employed a minimum mean depth of coverage of 100 and a minimum percentage of 96% for regions with more than 20× coverage. The mean depth of coverage for 215 samples analyzed by the HiSeq 2000 sequencer was 1536.1 ± 538.4× (range, 206-5925×). The percentage of each region with more than 20× coverage (indicating the percentage of each region sequenced 20 times or more by MPS) was 98.8% ± 0.7% (range, 93.5%-99.4%).

To investigate the accuracy of the MPS used in this study, we compared the results of the Invader assay-based mutation screening[27] and MPS by blinded samples (384 samples were analyzed by both methods). As a result, 99.98% of results were identical in the Ion PGM system. Detailed information regarding this comparison was given in our recent report.[17]

DNA Variants Identified in the Large Japanese Nonsyndromic Hearing Loss Cohort

From the 1389 samples, including 1120 nonsyndromic sensorineural hearing loss cases and 269 controls, we identified 8376 kinds of variants, and 409 835 (average, 295.1 variants/sample; Figure 1) total variants were detected. Among the 8376 variants, 4992 were located in the exon region, 2 were located in exonic regions of micro-RNA MIR96, which is a causative micro RNA associated with DFNA50, and 92 were located in splicing junctions. The others were located in the 3'- untranslated region (UTR), 5'-UTR, intron, and intergenic regions. Among the exon region variants, 3646 affected proteins (2955 missense variants, 76 nonsense variants, 161 frame shift deletions, 71 frame shift insertions, 136 frame shift multibase substitutions, 89 non–frame shift deletions, 4 non–frame shift insertions, 149 non–frame shift multibase substitutions, 2 exonic splice junction substitutions, and 3 stop loss mutations). Together with the splicing junction and noncoding RNA mutations, 3742 variants remained for further analysis.

We filtered these variants using allele frequency <1% of (1) the 1000 genome project; (2) the exome variant server; (3) the human genetic variation database, which contains 1200 Japanese exome data; (4) the 269 in-house Japanese normal hearing controls; and the (5) 1000 controls in the deafness variation database. For this filtering step, we employed <1% frequency as a cutoff line because the most frequent pathogenic variants observed in the Japanese population were *GJB2*: c.235delC and *GJB2*: c.109C>G, and the allele frequencies in the Japanese control population were estimated as 0.4% and 0.6%.[28] However, there were some possibilities to filter out the frequent autosomal recessive (AR)-pathogenic variants; therefore, we did not filter out the pathogenic or likely pathogenic variants previously reported in the deafness variation database[26] and ClinVar.[29]

After filtering the many ethnic controls, 2823 variants remained (Figure 1). Among them, we removed variants only found in controls as probable polymorphisms. Finally, 2631 variants were selected as candidates (Figure 1; 2017 missense mutations, 72 nonsense mutations, 2 stop loss mutations, 131 frame shift deletions, 30 frame shift insertions, 129 frame shift multibase substitutions, 21 non–frame shift deletions, 4 non–frame shift insertions, 138 non–frame shift multibase substitutions, 2 exon split junction substitutions, 77 splicing junction regions, and 2 micro-RNA MIR96 exonic regions).

Table 1. Previously Reported Pathogenic Variants Detected in This Study.

Identified Variants	ESP6500	1000g	HGVD	dbSNP138	SIFT	PP2 HDVI	PP2 HVAR_	ClinVar	DVD	Disease	PMID	Pat NUM	CNT Num
ACTG1:NM_001614:c.353A>T:p.K118M				rs104894544	D	B	P	Pathogenic (DFNA20)	Pathogenic	NSHL-Dominant	13680526	3	0
ACTG1:NM_001614:c.721G>A:p.E241K				rs267606631	D	D	D	Pathogenic (DFNA20)	Pathogenic	NSHL-Dominant	19477959	1	0
CDH23:NM_022124:c.2407G>A:p.V803I			0.00141		T	B	B		Pathogenic	NSHL-Recessive	22899989	0	1
CDH23:NM_022124:c.2866G>A:p.E956K					D	D	D		Pathogenic	NSHL-Recessive	22899989	4	0
CDH23:NM_022124:c.4249C>T:p.R1417W			0.001255		D	D	P		Pathogenic	NSHL-Recessive	22899989	6	2
CDH23:NM_022124:c.5131G>A:p.V1711I	0.000079	0.000399361	0.001247	rs181611778	T	D	D		Pathogenic	NSHL-Recessive	22899989	2	0
CDH23:NM_022124:c.5147A>C:p.Q1716P					T	D	D		Pathogenic	NSHL-Recessive	17850630	4	0
CDH23:NM_022124:c.5627G>A:p.S1876N			0.003179		T	P	P		Pathogenic	NSHL-Recessive	22899989	6	0
CDH23:NM_022124:c.6085C>T:p.R2029W			0.002271		D	D	D		Pathogenic	NSHL-Recessive	17850630	19	1
CDH23:NM_022124:c.6319C>T:p.R2107X					T	.	.		Pathogenic	Usher syndrome	11090341	1	0
CDH23:NM_022124:c.6389C>T:p.A2130V					T	B	B		Pathogenic	NSHL-Recessive	22899989	2	0
CDH23:NM_022124:c.6861T>G:p.N2287K					D	D	D		Pathogenic	NSHL-Recessive	22899989	1	0
CDH23:NM_022124:c.719C>T:p.P240L		0.000199681	0.002725	rs121908354	T	D	D	Pathogenic (DFNB12)	Pathogenic	NSHL-Recessive	17850630	45	2
CDH23:NM_022124:c.902G>A:p.R301Q	0.000081			rs121908355	T	D	D	Pathogenic (Alport syndrome)	Pathogenic	NSHL-Recessive	17850630	2	0
CDH23:NM_022124:c.9127C>T:p.R3043W	0.00008				D	D	P		Pathogenic	Usher syndrome	21569298	1	0
COCH:NM_004086:c.263G>A:p.G88E				rs121908928	T	D	D	Pathogenic (DFNA9)	Pathogenic	NSHL-Dominant	9806553	1	0
COL11A2:NM_080680:c.2492C>T:p.S831L	0.000118			rs121912949	T	D	P		Pathogenic	Otospondylomegaepiphyseal dysplasia, AD	NULL	1	0
COL4A5:NM_000495:c.2215C>G:p.P739A		0.00344371	0.059811	rs104886164	T	B	B	Pathogenic (Alport syndrome)				19	0
COL4A5:NM_000495:c.2858G>T:p.G953V	0.000189	0.00794702	0.01005	rs78972735	.	.	.	Pathogenic (Alport syndrome)				2	0
CRYM:NM_001888:c.941A>C:p.K314T				rs104894512	D	P	B	Pathogenic (AD-NSHL)	Pathogenic	NSHL-Dominant	12471561	2	0
EYA1:NM_000503:c.1276G>A:p.G426S			0.00134	rs121909199	D	D	D	Pathogenic (BOR syndrome)	Pathogenic	BOR syndrome, AD	10655545	2	0
EYA1:NM_000503:c.1319G>A:p.R440Q				rs121909196	D	D	D	Pathogenic (Melnick-Fraser syndrome)	Pathogenic	BOR syndrome, AD	10464653	1	0
EYA1:NM_000503:c.724A>G:p.S242G		0.000199681	0.01083	rs191838840	T	B	B		Pathogenic	BOR syndrome, AD	12701758	4	0
GJB2:NM_004004:c.109G>A:p.V37I	0.001307	0.0153754	0.006806	rs72474224	T	D	D	Pathogenic (DFNB1A)	Pathogenic	NSHL-Recessive	10633133	47	4
GJB2:NM_004004:c.134G>A:p.G45E			0.00349	rs72561723	D	D	D	Pathogenic (DFNB1A)	Pathogenic	KID syndrome, DFNB1A	10501520	46	1
GJB2:NM_004004:c.146C>T:p.A49V			0.002494		D	P	B		Pathogenic	NSHL-Recessive	12560944	2	0
GJB2:NM_004004:c.212T>C:p.I71T			0.001166		D	D	P		Pathogenic	NSHL-Recessive	12560944	1	1
GJB2:NM_004004:c.223C>T:p.R75W				rs104894402	D	D	D	Pathogenic (DFNA3A)	Pathogenic	NSHL-Dominant	9856479	1	0
GJB2:NM_004004:c.235delC:p.L79fs		0.00159744		rs80338943				Pathogenic (DFNB1A)	Pathogenic	NSHL-Recessive	10501520	166	3
GJB2:NM_004004:c.257C>G:p.T86R					D	D	D		Pathogenic	NSHL-Recessive	12560944	11	0
GJB2:NM_004004:c.257C>T:p.T86M					D	D	D		Pathogenic	NSHL-Recessive	17041943	2	0
GJB2:NM_004004:c.29T>C:p.L10P					D	D	D		Pathogenic	NSHL-Recessive	12865758	1	0
GJB2:NM_004004:c.334_335del:p.K112fs									Pathogenic	NSHL-Recessive	9529365	1	0
GJB2:NM_004004:c.368C>A:p.T123N	0.000154	0.00179712	0.006146	rs111033188	T	B	B	Probable nonpathogenic	Pathogenic	NSHL-Recessive	10983956	4	2
GJB2:NM_004004:c.379C>T:p.R127C					D	D	B		Pathogenic	NSHL-Recessive	11587277	1	0
GJB2:NM_004004:c.389G>C:p.G130A					D	D	D		Pathogenic	NSHL-Recessive	12792423	1	0
GJB2:NM_004004:c.408C>A:p.Y136X			0.00349		T	.	.		Pathogenic	NSHL-Recessive	10501520	46	1
GJB2:NM_004004:c.427C>T:p.R143W	0.000231	0.000199681	0.002331	rs80338948	D	D	D	Pathogenic (DFNB1A)	Pathogenic	NSHL-Recessive	9471561	29	0
GJB2:NM_004004:c.511G>A:p.A171T	0.000154	0.00039936 1	0.001166	rs201004645	T	P	B	Probable nonpathogenic	Pathogenic	NSHL-Recessive	11438992	1	0

(continued)

Table 1. (continued)

Identified Variants	ESP6500	1000g	HGVD	dbSNP138	SIFT	PP2 HDVI	PP2 HVAR_	ClinVar	DVD	Disease	PMID	Pat NUM	CNT Num
GJB2:NM_004004:c.571T>C:p.F191L		0.000199681	0.004115		D	D	D	Probable nonpathogenic	Pathogenic	NSHL-Recessive	12772454	0	1
GJB2:NM_004004:c.583A>G:p.M195V			0.001166		D	D	D		Pathogenic	NSHL-Recessive	20497192	4	0
GJB2:NM_004004:c.95G>A:p.R32H				rs111033190	D	D	D	Pathogenic (DFNB1A)	Pathogenic	NSHL-Recessive	11493200	3	0
GJB2:NM_004004c.299_300del:p.H100fs				rs111033204				Pathogenic (DFNB1A)	Pathogenic	NSHL-Recessive	10633133	14	0
GJB3:NM_024009:c.538C>T:p.R180X		0.000199681		rs74315319	T	.	.	Pathogenic (DFNA2B)	Benign*	NULL	NULL	1	0
GJB3:NM_024009:c.547G>A:p.E183K	0.000077	0.000998403	0.001361	rs74315318	D	D	D	Pathogenic (DFNA2B)	Benign*	NULL	NULL	2	0
GJB3:NM_024009:c.580G>A:p.A194T		0.00139776	0.01179	rs117385606	T	B	B	Pathogenic	Benign*	NULL	NULL	16	2
GJB6:NM_006783:c.689dupA:p.N230fs	0.000639							Pathogenic				9	0
KCNQ4:NM_004700:c.546C>G:p.F182L		0.000599042	0.006579	rs80358273	T	B	B	Pathogenic (DFNA2)	Pathogenic	NSHL-Dominant	17033161	8	1
LOXHD1:NM_144612:c.4480C>T:p.R1494X	0.001314	0.000199681		rs201587138	T	.	.		Pathogenic	NSHL-Recessive	23226338	2	0
LOXHD1:NM_144612:c.469C>T:p.R157C		0.000399361			.	.	.		Pathogenic	Fuchs corneal dystrophy		1	0
LOXHD1:NM_144612:c.4714C>A:p.R1572R	0.000657	0.0181709	0.051502	rs75949023				Pathogenic (DFNB77)	Benign*	NULL	NULL	83	18
MARVELD2:NM_001244734:c.1295+1G>A					.	.	.		Pathogenic	NSHL-Recessive	18084694	1	0
MYH9:NM_002473:c.2104C>T:p.R702C				rs80338826	D	D	D	Pathogenic (Fechtner syndrome)	Pathogenic	Epstein syndrome	10973259	1	0
MYH9:NM_002473:c.2114G>A:p.R705H				rs80338828	D	D	D	Pathogenic (DFNB17)	Pathogenic	NSHL/MYH9 related diseases, AD	11023810	1	0
MYO15A:NM_016239:c.6731G>A:p.G2244E					D	D	D		Pathogenic	NSHL-Recessive	17546645	2	0
MYO15A:NM_016239:c.8467G>A:p.D2823N					D	D	D		Pathogenic	NSHL-Recessive	22736430	1	0
MYO6:NM_004999:c.3496C>T:p.R1166X				rs121912558	T	.	.	Pathogenic (DFNB37)	Pathogenic	NSHL-Recessive	12687499	1	0
MYO7A:NM_000260:c.2005C>T:p.R669X	0.000081			rs111033201	T	.	.	Pathogenic (USH1B)	Pathogenic	Usher syndrome	9718356	1	0
MYO7A:NM_000260:c.2311G>T:p.A771S			0.003129		D	P	P		Pathogenic	Usher syndrome	20844544	4	1
MYO7A:NM_000260:c.3508G>A:p.E1170K				rs111033214	D	D	D	Pathogenic (USH1B)	Pathogenic	Usher syndrome	10425080	1	0
MYO7A:NM_000260:c.3602G>C:p.C1201S		0.000798722	0.002287	rs117966637	D	D	D	Unknown	Pathogenic	Usher syndrome	23237960	3	1
MYO7A:NM_000260:c.3718C>T:p.R1240W	0.000079	0.000199681		rs371374104	D	D	D		Pathogenic	Usher syndrome	16963483	1	0
MYO7A:NM_000260:c.3979G>A:p.E1327K	0.000079			rs373169422	D	D	D		Pathogenic	Usher syndrome	12112664	1	0
MYO7A:NM_000260:c.635G>A:p.R212H				rs28934610	D	D	D	Pathogenic (USH1B)	Pathogenic	Usher syndrome	7870171	2	0
MYO7A:NM_000260:c.652G>A:p.D218N	0.00008			rs201539845	D	D	D	Pathogenic (DFNA11)	Pathogenic	NSHL-Recessive	21150918	1	0
OTOF:NM_194248:c.1236delC:p.P412fs								Pathogenic (DFNB9)				2	0
OTOF:NM_194248:c.1273C>T:p.R425X					T	.	.	Pathogenic (DFNB9)				1	0
OTOF:NM_194248c.4023+1G>A		0.00179712	0.002269	rs186810296	.	.	.	Pathogenic				5	2
PCDH15:NM_033056:c.733C>T:p.R245X	0.000384			rs111033260	T	.	.	Pathogenic (USH1F)	Benign*	NULL	NULL	2	0
SIX1:NM_005982:c.386A>G:p.Y129C				rs104894478	D	D	D	Pathogenic (BOR syndrome 3)	Pathogenic	BOR syndrome, AD	15141091	1	0
SLC26A4:NM_000441:c.1001+1G>A	0.000461			rs80338849	.	.	.	Pathogenic (DFNB4)	Pathogenic	Pendred syndrome-Recessive	9618167	1	0
SLC26A4:NM_000441:c.1115C>T:p.A372V				rs121908364	D	D	D	Pathogenic (DFNB4)	Pathogenic	NSHL; NSHL with EVA, AR	10190331	1	0
SLC26A4:NM_000441:c.1174A>T:p.N392Y		0.000199681		rs201562855	D	D	D		Pathogenic	NSHL-Recessive	12676893	1	0
SLC26A4:NM_000441:c.1229C>T:p.T410M	0.000231	0.000199681	0.00134	rs111033220	D	D	D	Pathogenic (DFNB4)	Pathogenic	NSHL with EVA/Pendred syndrome	9618167	16	1
SLC26A4:NM_000441:c.1315G>A:p.G439R					D	D	D		Pathogenic	NSHL-Recessive	17851929	2	0
SLC26A4:NM_000441:c.1489G>A:p.G497S				rs111033308	D	D	D	Pathogenic (DFNB4)	Pathogenic	NSHL with EVA/Pendred syndrome	9500541	1	0
SLC26A4:NM_000441:c.1579A>C:p.T527P			0.00134		D	D	D		Pathogenic	NSHL-Recessive	17851929	3	0
SLC26A4:NM_000441:c.165-13T>G									Pathogenic	NSHL with EVA, AR	19645628	0	1
SLC26A4:NM_000441:c.1804-6G>A		0.000599042		rs377713770					Pathogenic	NSHL-Recessive	15574297	2	0

(continued)

Table 1. (continued)

Identified Variants	ESP6500	1000g	HGVD	dbSNP138	SIFT	PP2 HDVI	PP2 HVAR_	ClinVar	DVD	Disease	PMID	Pat NUM	CNT Num
SLC26A4:NM_000441:c.2162C>T:p.T721M				rs121908363	D	D	D	Pathogenic (DFNB4)	Pathogenic	NSHL with EVA/Pendred syndrome	10190331	4	0
SLC26A4:NM_000441:c.2168A>G:p.H723R		0.000399361	0.002264	rs121908362	D	D	D	Pathogenic (DFNB4)	Pathogenic	NSHL with EVA/Pendred syndrome	9618166	53	2
SLC26A4:NM_000441:c.2219G>T:p.G740V	0.000154			rs111033310	T	B	B	Unknown	Pathogenic	NSHL with EVA, AR	16570074	1	0
SLC26A4:NM_000441:c.2228T>A:p.L743X					T	.	.		Pathogenic	NSHL with EVA, AR	19954013	2	0
SLC26A4:NM_000441:c.225C>G:p.L75L	0.000231	0.000399361	0.002141	rs187447337					Pathogenic	NSHL-Recessive	23185506	1	1
SLC26A4:NM_000441:c.2283A>G:p.T761T		0.000399361	0.038462	rs202033028					Pathogenic	NSHL-Recessive	23185506	1	0
SLC26A4:NM_000441:c.367C>T:p.P123S			0.001166		T	D	D		Pathogenic	NSHL-Recessive	14508505	2	0
SLC26A4:NM_000441:c.439A>G:p.M147V			0.001667		D	D	D		Pathogenic	NSHL-Recessive	14508505	3	0
SLC26A4:NM_000441:c.601-1G>A			0.001166		.	.	.		Pathogenic	Pendred syndrome-Recessive	14508505	5	0
SLC26A4:NM_000441:c.678T>C:p.A226A									Pathogenic	NSHL-Recessive	23185506	1	0
SLC26A4:NM_000441:c.697G>C:p.V233L					T	D	D	Unknown	Pathogenic	NSHL-Recessive	17443271	1	0
SLC26A4:NM_000441:c.757A>G:p.I253V			0.001166		T	P	P		Pathogenic	NSHL-Recessive	23185506	2	0
SLC26A4:NM_000441:c.918+1G>A					.	.	.		Pathogenic	Pendred syndrome-Recessive	9618166	1	0
SLC26A4:NM_000441:c.919-18T>G									Pathogenic	NSHL-Recessive	20137612	3	0
SLC26A4:NM_000441:c.919-2A>G			0.00134	rs111033313	.	.	.	Pathogenic (DFNB4)	Pathogenic	NSHL with EVA/Pendred syndrome	10874637	8	0
SLC26A4:NM_000441:c.920C>T:p.T307M	0.000077	0.000199681	0.001166	rs144691257	D	D	D		Pathogenic	NSHL with EVA/Mondini, AR	16570074	2	0
SLC26A4:NM_000441:c.G1975G>C:p.V659L		0.000199681		rs200455203	D	P	B		Pathogenic	NSHL-Recessive	17443271	2	0
TECTA:NM_005422:c.1685C>T:p.T562M			0.00187		T	D	P		Pathogenic	NSHL-Dominant	21520338	0	1
TECTA:NM_005422:c.4198C>T:p.H1400Y		0.000199681	0.00271		T	D	P		Pathogenic	NSHL-Dominant	22718023	2	2
TECTA:NM_005422:c.5372C>G:p.P1791R					T	B	B		Pathogenic	NSHL-Dominant	21520338	1	0
TECTA:NM_005422:c.5597C>T:p.T1866M	0.000077			rs140236996	D	D	D		Pathogenic	NSHL-Dominant	20947814	1	0
TMC1:NM_138691:c.1165C>T:p.R389X	0.000077	0.000199681		rs151001642	T	.	.	Pathogenic	Pathogenic	NSHL-Recessive	15605408	1	0
TMIE:NM_147196:c.257G>A:p.R86Q					D	D	P		Pathogenic	NSHL-Recessive	20206386	1	0
TMPRSS3:NM_024022:c.916G>A:p.A306T		0.000199681	0.002058	rs181949335	.	D	D	Probable-pathogenic	Pathogenic	NSHL-Recessive	17551081	1	1
USH1C:NM_005709:c.1016G>A:p.R339Q					D	P	B		Pathogenic	Usher syndrome	22135276	0	1
USH2A:NM_206933:c.1876C>T:p.R626X					T	.	.		Pathogenic	Usher syndrome	10729113	1	0
USH2A:NM_206933:c.2802T>G:p.C934W		0.000798722	0.003333	rs201527662	D	D	D	Pathogenic (USH2A)	Benign*	NULL	NULL	3	1
USH2A:NM_206933:c.802G>A:p.G268R				rs111033280	D	D	D	Unknown	Pathogenic	Usher syndrome	18273898	1	0
USH2A:NM_206933:c.8254G>A:p.G2752R		0.000399361		rs201863550	D	D	D		Pathogenic	Usher syndrome	19737284	1	0
USH2A:NM_206933:c.8559-2A>G		0.000199681			.	.	.	Pathogenic (USH2A)	Pathogenic	Usher syndrome	19023448	6	0
WFS1:NM_006005:c.1846G>T:p.A616S		0.000199681	0.003411		T	B	B		Pathogenic	NSHL-Dominant	16408729	1	0
WFS1:NM_006005:c.1957C>T:p.R653C	0.000231	0.000199681	0.1	rs201064551	D	D	D		Pathogenic	Diabetes, AD		1	0
WFS1:NM_006005:c.2051C>T:p.A684V					D	D	D	Pathogenic (Wolfram-like syndrome, AD)	Pathogenic	Wolfram-like syndrome (deafness with optic atrophy), AD		1	0
WFS1:NM_006005:c.2146G>A:p.A716T				rs28937893	T	D	P	Pathogenic (Wolfram-like syndrome, AD)	Pathogenic	Wolfram-like syndrome (deafness with optic atrophy), AD	11709537	1	0
WFS1:NM_006005:c.2171C>T:p.P724L				rs28937890	D	D	D	Pathogenic (Wolfram-like syndrome, AD)	Pathogenic	Wolfram-like syndrome (deafness with optic atrophy), AD	9771706	0	1
WFS1:NM_006005:c.2507A>C:p.K836T					T	D	D		Pathogenic	NSHL-Dominant	19877185	1	0
WFS1:NM_006005:c.2590G>A:p.E864K				rs74315205	T	D	D	Pathogenic (Wolfram-like syndrome, AD)	Pathogenic	Wolfram-like syndrome (deafness with optic atrophy), AD		2	0
WFS1:NM_006005c.2185G>A:p.D729N		0.000399361			T	B	B		Pathogenic	Wolfram syndrome, AR	12107816	1	0

Abbreviations: 1000g, 1,000 genome database[23]; ClinVar, Clinical variation database29; DVD: Deafness variation database[26]; ESP6500, 6500 exome variants[24]; HGVD, Human Genetic Variation Database[25]; NUM, identified allele number in 269 normal hearing controls; Pat NUM, identified allele number in 1120 hearing loss cases; CNT NUM, identified allele number in 269 controls; Benign*, recently re-categorized variants (from pathogenic to benign) using a large number of many ethnic controls allele frequencies[26]; AD, autosomal dominant; AR autosomal recessive; NSHL, non-syndromic hearing loss; BOR, Branchio-oto-renal; EVA, enlarged vestibular aqueduct; PP2, PolyPhen2; PMID, PubMed ID.

Of the 2631 variants, 1694 (64.4%) were found in 1 patient (Figure 2). A total of 392 variants (14.9%) were found in 2 patients, 139 (5.3%) were found in 3, 92 (3.5%) were found in 4, 47 (1.8%) were found in 5, and 267 (10.1%) were found in 6 or more.

Previously Reported and Identified Pathogenic Variants in the Large Japanese Nonsyndromic Hearing Loss Cohort

Of the 2631 candidate variants, 105 were categorized as pathogenic variants in the deafness variation database, and 49 were categorized as pathogenic variants in ClinVar (Table 1). Thirty-seven variants were categorized as pathogenic in both databases, and 6 variants (*GJB3*:NM_001005752: c.538C>T:p.R180X, *GJB3*:NM_024009: c.547G>A:p.E183K, *GJB3*:NM_024009:c.580G>A:p.A194T, *LOXHD1*:NM_144612:c.4714C>A:p.R1572R, *USH2A*:NM_007123: c.2802T>G:p.C934W, and *PCDH15*:NM_001142767:c.622C>T:p.R208X) were categorized as pathogenic variants in ClinVar, but categorized as nonpathogenic in the deafness variation database. In contrast, 3 variants (*GJB2*:NM_004004: c.368C>A:p.T123N, *GJB2*:NM_004004: c.511G>A:p.A171T, and *GJB2*:NM_004004: c.571T>C:p.F191L) were categorized as pathogenic variants in the deafness variation database; however, they were categorized as nonpathogenic variants in ClinVar.

Among these previously reported pathogenic variants, 26 were autosomal dominant mutations in *ACTG1, COCH, COL11A2, CRYM, EYA1, GJB2, GJB3, KCNQ4, MYH9, SIX1, TECTA*, and *WFS1*; 88 were autosomal recessive mutations in *CDH23, GJB2, GJB3, GJB6, LOXHD1, MARVELD2, MYO15A, MYO6, MYO7A, OTOF, SLC26A4, OTOF, TMC1, TMIE, TMPRSS3, USH1C, USH2A*, and *WFS1*; and 2 were X-linked mutations in *COL4A5*.

The most frequent mutation was *GJB2*:c.235delC, which was found in 166 alleles from 1120 patients with hearing loss and 3 alleles in the 269 normal hearing controls. *SLC26A4*:c.2168A>G (p.H723R) were the second most frequent; 53 alleles were found in 1120 hearing loss cases, and 2 alleles were found in the 269 controls.

Comparison of Previously Reported Pathogenic Mutations and Newly Identified Variants

To determine whether the missense mutations affect prediction cutoffs of the computer programs for protein function, we compared the prediction scores of the pathogenic variants previously reported to those of the newly identified variants using 12 computer programs including ANNOVAR.[21,22] As a result, the previously reported pathogenic variants were predicted to cause more severe effects (or damage) to protein function than those of the newly identified variants. The average SIFT* prediction score for the previously reported pathogenic variants for autosomal dominant nonsyndromic hearing loss (AD-NSHL: 25 variants) was 0.86 ± 0.24, that for autosomal recessive inheritance nonsyndromic hearing loss (AR-NSHL: 59 variants) was 0.88 ± 0.23, and that for the newly identified missense variants (1926 variants) was 0.74 ± 0.32 (Figure 2, *SIFT scores from the ANNOVAR software were converted to 1-SIFT scores; therefore, a higher score indicated a more damaging variant). The PolyPhen2 results were similar to the SIFT results (Polyphen2 HVID: the AD-NSHL variant score was 0.83 ± 0.32, the AR-NSHL variant score was 0.87 ± 0.29, and the newly identified variant score was 0.60 ± 0.43; Polyphen2 HVAR: the AD-NSHL variant score was 0.77 ± 0.35, the AR-NSHL variant score was 0.79 ± 0.34, and the newly identified variant score was 0.50 ± 0.43). The LRT, Mutation Taster, Mutation Assessor, FATHMM, Radial SVM, LR, GERP++, PhyloP, and SiPhy 29-way log odds scores were similar (Figure 2). All prediction programs are based on some similar strategies and are not completely independent of each other. However, each prediction program estimates the effect of amino acid changes from different viewpoints to some extent (some programs estimate the homology among many species, while others estimate the properties of amino acids). Therefore, it is conceivable that combining the results of multiple prediction programs might be better than using the results of each individual prediction program.

To maximize prediction appropriateness, we converted the results of each prediction program to a z-score (using all missense variant results: AD-NSHL variants + AR-NSHL variants + novel variants = 2010 variants) and calculated the average z-score of the 12 prediction programs (Figure 2, Table 2). As a result, the z-score of the AD-NSHL variants was 0.65 ± 0.45, that of the AR-NSHL variants was 0.60 ± 0.55, and that of the newly identified variants was -0.27 ± 0.65. These results clearly reveal differences between the previously reported pathogenic variants and the newly identified variants, including both the pathogenic variants and rare polymorphisms. As a result of the statistical analysis, the average z-score of 12 prediction programs indicated a P value lower than that of each of 12 prediction programs (AD-pathogenic vs novel: $P = 4.2 \times 10^{-7}$, AR-pathogenic vs novel: $P = 5.1 \times 10^{-9}$, Tukey's HSD test). As a notable result, the *GJB2*:c.368C>A (p.T123N) variant revealed the lowest score of -1.14 in the previously reported pathogenic variant group. This variant was recategorized as a rare polymorphism in our previous report.[28] Of course, in silico analysis has a limitation in the prediction of the pathogenicity and segregation analysis for family samples, and in vitro or in vivo studies are required to make conclusions about the pathogenicity of each variant.

From these results, we further selected the missense variants with average z-scores >0.05 (average -1 standard deviation of previously reported AR-NSHL variants) as candidates and analyzed the molecular epidemiology and mutation spectrum in Japanese patients with hearing loss.

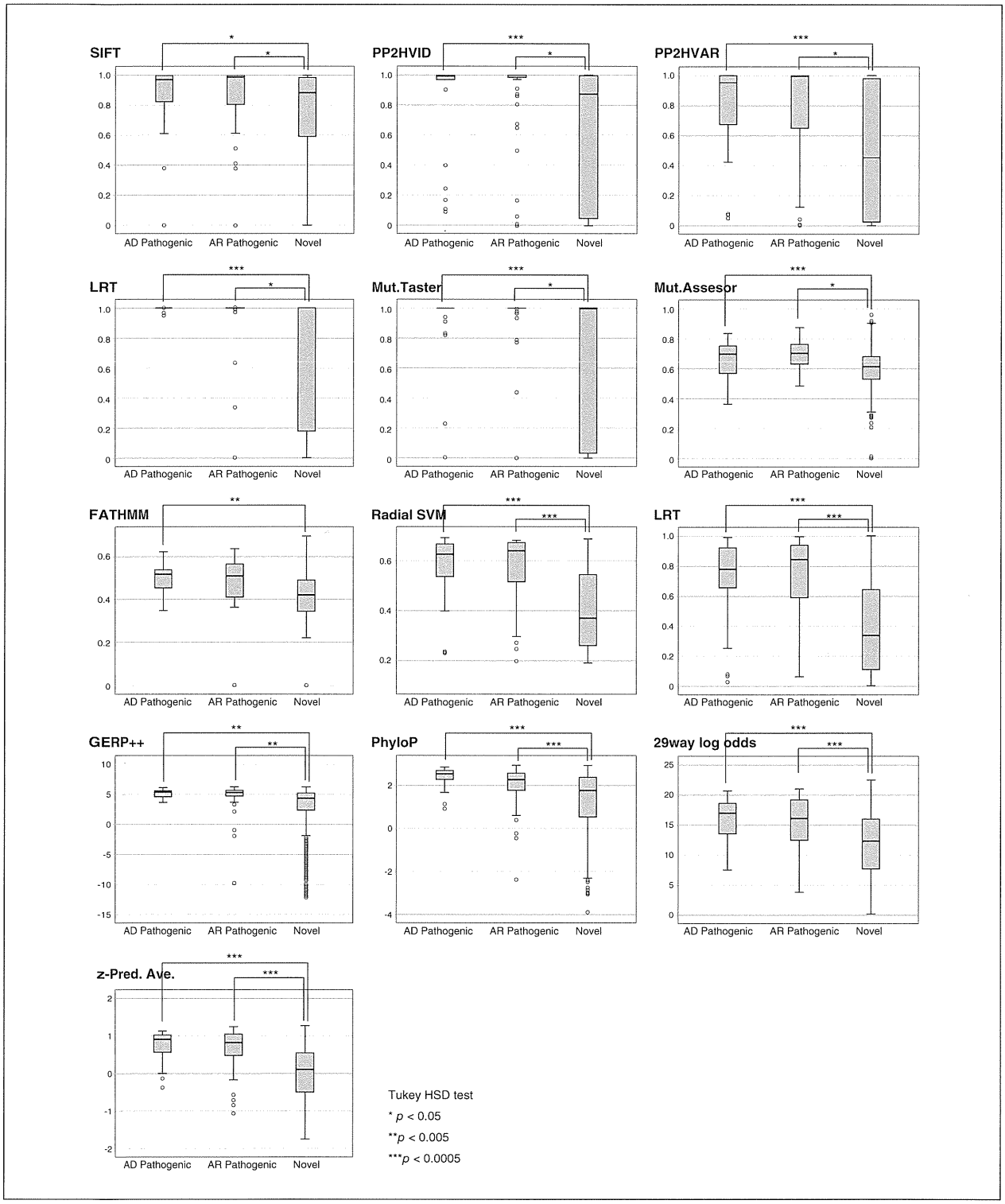

Figure 2. Distributions of the 12 computer prediction software programs (SIFT, Polyphen2 HVID, Polyphen2 HVAR, LRT, Mutation Taster, Mutation Assessor, FATHMM, Radial SVM, LR, GERP++, PhyloP, and SiPhy 29-way log odds) for identifying the missense mutations and the average z-score of the 12 prediction programs described in the text. The statistical analysis was performed using SPSS version 18 (SPSS Inc, Chicago, Illinois, USA).