



**HAL**  
open science

## Estimating the prevalence of congenital disaccharidase deficiencies using allele frequencies from gnomAD

C. de Leusse, C. Roman, B. Roquelaure, Alexandre Fabre

► **To cite this version:**

C. de Leusse, C. Roman, B. Roquelaure, Alexandre Fabre. Estimating the prevalence of congenital disaccharidase deficiencies using allele frequencies from gnomAD. Archives de Pédiatrie, 2022, 29 (8), pp.599-603. 10.1016/j.arcped.2022.08.005 . hal-04034595

**HAL Id: hal-04034595**

**<https://amu.hal.science/hal-04034595>**

Submitted on 5 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimating the prevalence of congenital disaccharidase deficiencies using allele frequencies from gnomAD

C. de Leusse<sup>1</sup>, C. Roman<sup>1</sup>, B. Roquelaure<sup>1</sup>, A. Fabre<sup>1,2,\*</sup>

<sup>1</sup> AP-HM, Hôpital de la Timone Enfants, Service de pédiatrie multidisciplinaire, Marseille, France

<sup>2</sup> Aix Marseille Univ, INSERM, MMG, Marseille, France

## A B S T R A C T

**Background:** There are currently three known congenital disaccharidase deficiencies: congenital lactase deficiency (CLD), congenital sucrase-isomaltase deficiency (CSD), and congenital trehalase deficiency (CTD). No congenital deficiency has been described for maltase-glucoamylase (MGAM).

**Methods:** A literature search was performed in PubMed for the pathogenic variants CLD, CSD, and CTD and the articles retrieved were analyzed to estimate the prevalence of congenital disaccharidase deficiencies.

**Results:** Based on reported variants, the estimated prevalence was 1.3 per 10<sup>6</sup> births (95% CI: 1.1–1.7) for CLD, and 31.4 per 10<sup>6</sup> births (95% CI: 28.3–34.8) for CSD. Using data on previously reported variants and variants predicted to be loss-of-function in gnomAD, the overall estimated prevalence was 2.3 per 10<sup>6</sup> births (95% CI: 1.9–2.9) for CLD, 57.6 per 10<sup>6</sup> births (95% CI: 52.5–63.2) for CSD, and 9.2 per 10<sup>6</sup> births (95% CI: 2.5–3.7) for CTD.

**Conclusion:** The prevalence of CSD was found to be relatively high, while for other congenital disaccharidase deficiencies, the estimated prevalence was very low.

### Keywords:

congenital disaccharidase deficiencies

LCT

SI

TREH

MGAM

gnomAD

## 1. Introduction

Disaccharidases are hydrolase enzymes found in the brush border of the intestine whose main function is to break down disaccharide into monosaccharide [1]. There are four human disaccharidases: lactase-phlorizin hydrolase encoded by the *LCT* gene (OMIM: 603202), maltase-glucoamylase encoded by the *MGAM* gene (OMIM: 154360), sucrase-isomaltase encoded by the *SI* gene (OMIM: 609845), and trehalase encoded by *TREH* gene (OMIM: 275360). Acquired or congenital disaccharidase alterations are linked to carbohydrate intolerance [2]. The main manifestations of carbohydrate intolerance are bloating, diarrhea, constipation, flatulence, and borborygmus after ingestion of the non-digested disaccharides [1]. Three congenital disaccharidase deficiencies have been described to date: congenital lactase deficiency (CLD, OMIM: 223000), congenital sucrase-isomaltase deficiency (CSD, OMIM: 222900), and congenital trehalase deficiency (CTD, OMIM: 612119). The prevalence of congenital disaccharidase deficiencies has been estimated in some populations (mostly from enzymatic and more recently from genetic data). The prevalence of CLD was estimated to be 1 per 60,000 births in Finland with a frequency of carriers ranging from 1/35 to 1/133 [3]. Reported

prevalences of CSD range from 0.04% in Americans of European descent [4] to 3% in Inuits [5], and the prevalence of CTD has been estimated at 8% in Greenlanders [6]. The expansion of genomic variant databases has made it possible to directly evaluate the prevalence of rare Mendelian diseases (e.g., Wilson disease [7] or lysosomal acid lipase deficiency [8]). The aim of this study was to evaluate the prevalence of congenital disaccharidase deficiencies using the gnomAD variant database [9].

## 2. Methods

A literature search was performed in PubMed for pathogenic variants associated with CLD, CSD, or CTD using the terms “LCT diarrhea,” “congenital lactase deficiency,” “SI deficiency,” and “congenital sucrase-isomaltase deficiency.” A similar search was performed using the term “MGAM deficiency” and “maltase-glucoamylase deficiency.” The last search was performed on 6 June 2020. Articles were included if they described variants associated with congenital disaccharidase deficiency, provided information on pathogenic variants, and the number of variants were retrieved. Variant names were checked using Mutalyzer<sup>1</sup> and corrected if necessary (NCBI reference sequences: *LCT*, NG\_008104.2; *SI*, NG\_017043.1; *TREH*, NG\_023321.1; and *MGAM*, NG\_033954.1). Each individual’s status (homozygous for

\* Corresponding author: Professor Alexandre FABRE, MD-PhD, Service de pédiatrie multidisciplinaire, Hôpital de la Timone Enfants, 264 Rue Saint-Pierre, 13005 Marseille, France – Tel: +33 (0)4 91 38 80 60, Fax: +33 (0)4 91 38 6756

E-mail address: alexandre.fabre@ap-hm.fr (A. Fabre).

variant, compound heterozygous) and the type of variant (missense, loss of function) were also recorded. Variants listed in ClinVar<sup>2</sup> as pathogenic or likely pathogenic were also collected.

Allele frequencies were downloaded from gnomAD (gnomAD\_v2.1.1) on 6 May 2020. All loss-of-function variants except those with flags were included in prevalence estimates.

Birth prevalences were estimated as described by Gao et al. [7] using the Hardy–Weinberg equation with 95% Wilson score confidence intervals (CIs).

### 3. Results

A total of 18 articles were included (six on CLD and 12 on CSD), describing 42 different variants (14 of *LCT* and 28 of *SI*), and 171 alleles (84 of *LCT* and 87 of *SI*), among which 40 alleles (46.8%) were in homozygous state (33 of *LCT* and seven of *SI*). Overall, 29 of the variants were missense (six of *LCT* and 23 of *SI*) and 12 were loss of function (eight for *LCT* and four for *SI*), and one *SI* variant had nomenclature error. Three loss of function variants of *SI* not reported in the literature were retrieved from ClinVar. No pathogenic variants of *MGAM* were retrieved either from the literature or from ClinVar. The total number of variants included in the analysis was therefore 44 (14 of *LCT*, 30 of *SI*, Table 1) Table 1.A, Table 1.B.

In total, 26 of the selected variants (six of *LCT*, 20 of *SI*) were described at least once in gnomAD. A further 135 predicted loss-of-function variants not reported in the literature were found in gnomAD (52 of *LCT*, 21 of *SI*, and 62 of *MGAM*), five of which had more than 100 alleles reported:

- Four variants of *SI*: c.2789A>G, p.Gln930Arg with 139 alleles, c.5234T>G p.Phe1745Cys with 270 alleles, c.3218G>A p.Gly1073Asp with 356 alleles and c.1730T>G p.Val577Gly with 438 alleles, in four cases reported in homozygous state (one for p.Val577Gly, one for p.Phe1745Cys, and two for p.Gly1073Asp). These four variants account for 67.6% of the alleles either described in the literature or predicted to be loss-of-function. These alleles are mostly found in non-Finnish Europeans (74.2%).
- One *LCT* variant: c.4170T>A p.Tyr1390Ter with 246 alleles, reported once in homozygous state. This variant accounts for 71.3% of the alleles either described in the literature or predicted to be loss-of-function and is mostly present in Finns (97.2%).

Based on variants reported in the literature, the estimated global prevalence was 1.3 per 10<sup>6</sup> births (95% CI: 1.1–1.7) for CLD, and 31.4 per 10<sup>6</sup> births (95% CI: 28.3–34.8) for CSD. Including reported variants and those predicted to be loss-of-function in gnomAD, the estimated global prevalence was 2.3 per 10<sup>6</sup> births (95% CI: 1.9–2.9) for CLD, 57.6 per 10<sup>6</sup> births (95% CI: 52.5–63.2) for CSD, and 0.8 per 10<sup>6</sup> births (95% CI: 0.6–1.1) for putative *MGAM* deficiency. Estimated birth prevalences in different populations are listed in Table 2. Note that the estimated prevalence of CLD in Finns is 115.56/10<sup>6</sup> birth (95% CI: 90.54–147.48) and the prevalence of CSD in non-Finnish Europeans is 128.5 per 10<sup>6</sup> births (95% CI: 115.15–143.4).

We found no reports of *TREH* variants associated with CTD; however, Saleheen et al. [10] have described a loss-of-function variant of *TREH* (c.90-9\_106del) in homozygous state in six patients whose effect was to lower concentrations of apoB-containing lipoprotein subfractions. This variant is present in gnomAD with 281 alleles, mostly found in South Asians (265 alleles) (Table 1C). The total number of alleles for other loss-of-function variants of *TREH* in gnomAD is 98. The c.90-9\_106del variant therefore has a big effect on estimated prevalences. If it is included, the estimated prevalence of CTD is 9.2 per 10<sup>6</sup> births (95% CI: 2.5–3.7) worldwide and 98.47 per 10<sup>6</sup> births

(95% CI: 78.07–124.12) in South Asians. If it is not included, the estimated worldwide prevalence of CTD is 0.2 per 10<sup>6</sup> births (95% CI: 0.14–0.3).

Based only on predicted loss-of-function variants in gnomAD, the estimated worldwide prevalence was 1.44 per 10<sup>6</sup> births (95% CI: 1–1.69) for CLD, 2.56 per 10<sup>6</sup> births (95% CI: 2.25–3.24) for CSD, 2.89 per 10<sup>6</sup> births (95% CI: 2.56–3.61) for CTD, and 0.81 per 10<sup>6</sup> births (95% CI: 0.64–1.21) for putative *MGAM* deficiency.

### 4. Discussion

GnomAD (and its predecessor ExAC) has proven to be an invaluable clinical genetics resource, improving the identification of variants in clinical exomes and our understanding of human genetic variation [11]. GnomAD data have recently been used to estimate the prevalence of rare diseases [7,8]. We used the same method to estimate overall and population-specific prevalences of disaccharidase deficiencies. The estimated prevalence of CLD was very low (< 1.3 per 10<sup>6</sup> births) in all populations other than Finns (115.56 per 10<sup>6</sup> births), with a carrier frequency of pathogenic variants (1/93) in the range of previous estimates (between 1/35 and 1/133 according to Kuokkanen et al. [3]). The estimated prevalence of CSD was higher (57.59 per 10<sup>6</sup> births) with huge differences between populations, ranging from 247.4 and 128.5 per 10<sup>6</sup> births in Ashkenazi Jews and non-Finnish Europeans, respectively, to 0.36 per 10<sup>6</sup> births in East Asians. The estimate for non-Finnish Europeans is about 3 times lower than Welsh and colleagues' [4] estimate for European Americans (400 per 10<sup>6</sup> births), possibly because the latter was based on measurements of decreased enzyme activity, for which there are also non-genetic causes. Urich et al. [12] estimated that four variants (p.Val577Gly p.Gly1073Asp, p.Arg1124Ter, p.Phe1745Cys) account for 59% of alleles responsible for CSD in European populations, which is similar to our estimate (68%). *MGAM* deficiency has not been described to date either in enzymatic or genetic studies; however, the estimated prevalence suggests that although it is quite rare, it should be as frequent as CLD (in populations of non-Finnish ancestry). However, since *MGAM* only accounts for 20% of mucosal maltase activity, the phenotype is probably mild [13].

One of the limitations of this study is that the prevalence estimates were based on loss-of-function and published variants. In the absence of information on missense variants therefore, the estimates, based only on loss-of-function variants, are low. It has been suggested, for example, that CSD is underdiagnosed in Chinese children [14]; however, since no associated variant has ever been described, the estimated prevalence of CSD in East Asians was only 0.36 per 10<sup>6</sup> births (95% CI: 0.11–1.23). For understudied populations therefore, the estimated birth prevalence should be considered lower bounds. Conversely, despite the very rare prevalence, over time the number of patients could become significant, given the birth rate of European countries. For example, France (approximately 800,000 births per year) should have one case of CLD every few years, but to date there is no French case published, possibly because once the diagnosis is made the treatment is relatively easy or because of the absence of an incentive to publish case reports. The other limitation is the use of the gnomAD database that, despite its size, is still a relatively small subset of the human variations; nonetheless, our results are coherent with the frequency of CLD in the Finnish population or the fact that four variants account for most of alleles responsible for CSD.

Brush-border proteins are subjected to selection pressure related to carbohydrate diets [15] and related congenital diseases are indeed relatively rare except in certain populations such as Finns for CLD and Inuits for CSD and CTD. A puzzling feature is that three main pathogenic variants are carried by nearly 0.7% of the European population; whether this is related to genetic drift remains to be studied.

For CTD, despite having been described nearly half a century ago [16] and reported in 8% of Greenland natives [6], no associated

**Table 1**

Reported variants of congenital disaccharidase deficiencies on PubMed

1.A **LCT**: Lactase-phlorizin hydrolase encoded by the gene *LCT*

References	Gene	Variant nomenclature	Amino acid sequence change	Types	Allele count	Number of homozygotes
Torniainen et al. 2009	LCT	c.1692_1696del	p.(Val565Leufs*3)	Missense	1	0
Torniainen et al. 2009	LCT	c.2062T>C	p.(Ser688Pro)	Missense	1	0
Coromina et al. 2005	LCT	c.2232_2253dup	p.(Leu752Lysfs*19)	Loss of function	2	1
Wanes et al. 2019	LCT	c.3362C>T	p.(Ser1121Leu)	Missense	2	1
Fazeli et al. 2015	LCT	c.3448del	p.(Ser1150Profs*19)	Loss of function	2	1
Kuokkanen et al. 2006, Torniainen et al. 2009, Wanes et al. 2019	LCT	c.4087G>A	p.(Gly1363Ser)	Missense	7	3
Kuokkanen et al. 2006, Torniainen et al. 2009	LCT	c.4170T>A	p.(Tyr1390Ter)	Loss of function	61	27
Uchida et al. 2012	LCT	c.4419C>G	p.(Tyr1473Ter)	Loss of function	1	0
Torniainen et al. 2009	LCT	c.4760G>A	p.(Arg1587His)	Missense	1	0
Torniainen et al. 2009	LCT	c.4834G>T	p.(Glu1612Ter)	Loss of function	1	0
Kuokkanen et al. 2006	LCT	c.4998_5001del	p.(Ser1666Argfs*56)	Loss of function	2	0
Uchida et al. 2012	LCT	c.5387del	p.(Asp1796Alafs*18)	Loss of function	1	0
Kuokkanen et al. 2006	LCT	c.653_654del	p.(Ser218Cysfs*6)	Loss of function	1	0
Kuokkanen et al. 2006	LCT	c.804G>C	p.(Gln268His)	Missense	1	0

**Table 1.B**Sucrase-isomaltase encoded by the gene *SI*

References	Gene	Variant nomenclature	Amino acid sequence change	Types	Allele count	Number of homozygotes
Jacob et al. 2000	SI	c.1019A>C	p.(Gln340Pro)	Missense	2	1
Gericke et al. 2017	SI	c.1607A>T	p.(Asp536Val)	Missense	1	0
Sander et al. 2005	SI	c.1648del	p.(Gln550Argfs*19)	Loss of function	1	0
Alfalah et al. 2009, Haberman et al. 2017, Sander et al. 2005, Uhrich et al. 2012, Gericke et al. 2017	SI	c.1730T>G	p.(Val577Gly)	Missense	13	0
Sander et al. 2005	SI	c.1780T>C	p.(Ser594Pro)	Missense	2	0
Ritz et al. 2003	SI	c.1859T>C	p.(Leu620Pro)	Missense	2	1
Keiser et al. 2006	SI	c.1903T>C	p.(Cys635Arg)	Missense	2	1
Sander et al. 2005	SI	c.2080A>C	p.(Thr694Pro)	Missense	1	0
Clinvar	SI	c.2159+2T>G		Loss of function	1	0
Gericke et al. 2017	SI	c.2222T>C	p.(Leu741Pro)	Missense	1	0
Husein et al. 2019	SI	c.2624T>C	p.(Phe875Ser)	Missense	1	0
Sander et al. 2005	SI	c.26887+1G>C nomenclature problem			2	0
Marcadier et al. 2015	SI	c.273_274del	p.(Gly92Leufs*8)	Loss of function	2	1
Gericke et al. 2017	SI	c.2789A>G	p.(Gln930Arg)	Missense	1	0
Gericke et al. 2017	SI	c.2791T>A/C	p.(Trp931Arg)	Missense	1	0
Gericke et al. 2017	SI	c.2792G>A	p.(Trp931Ter)	Loss of function	1	0
Gericke et al. 2017	SI	c.315G>C/T	p.(Trp105Cys)	Missense	1	0
Gericke et al. 2017, Sander et al. 2005, Uhrich et al. 2012, Alfalah et al. 2009, Gericke et al. 2017	SI	c.3218G>A	p.(Gly1073Asp)	Missense	23	0
Ouwendijk et al. 1996	SI	c.3293A>C	p.(Gln1098Pro)	Missense	2	1
Gericke et al. 2017, Uhrich et al. 2012	SI	c.3370C>T	p.(Arg1124Ter)	Loss of function	4	0
Spodsberg et al. 2001	SI	c.350A>G	p.(Gln117Arg)	Missense	2	1
Clinvar	SI	c.3586_3587del	p.(Met1196Valfs*15)	Loss of function	1	0
Alfalah et al. 2009, Sander et al. 2005	SI	c.3686G>A	p.(Cys1229Tyr)	Missense	5	1
Sander et al. 2005	SI	c.4099A>G	p.(Arg1367Gly)	Missense	1	0
Naim et al. 2005	SI	c.4427G>C	p.(Gly1476Ala)	Missense	1	0
Gericke et al. 2017	SI	c.4592G>A	p.(Cys1531Tyr)	Missense	1	0
Haberman et al. 2017	SI	c.4593T>G	p.(Cys1531Trp)	Missense	1	0
Gericke et al. 2017	SI	c.4630C>T	p.(Arg1544Cys)	Missense	1	0
Gericke et al. 2017	SI	c.4817C>T	p.(Thr1606Ile)	Missense	1	0
Sander et al. 2005, Sander et al. 2005, Uhrich et al. 2012, Alfalah et al. 2009, Gericke et al. 2017	SI	c.5234T>G	p.(Phe1745Cys)	Missense	11	0
Clinvar	SI	c.853G>T	(p.Glu285Ter)	Loss of function	1	0

**Table 1.C**TREH: Trehalase encoded by the gene *TREH*

Reference	Gene	Variant nomenclature	Amino acid sequence change	Types	Allele count	Number of homozygotes
Saleheen D et al. 2017	TREH	c.90-9_106del	?	Loss of function	12	6

**Table 2**

Estimated prevalence of heterozygous carriers of deleterious variants and of congenital sucrase-isomaltase deficiency, congenital lactase deficiency, putative maltase-glucoamylase deficiency, and congenital trehalase deficiency, calculated from pooled allele frequencies of previously reported variants and predicted loss-of-function variants in eight populations.

	Global	African	Latino	Ashkenazi Jewish	East Asian	Finnish	Non-Finnish European	Other	South Asian
Congenital sucrase-isomaltase deficiency Heterozygous carrier	1/132 (95% CI: 1/126–1/138)	1/267 (95% CI: 1/210–1/339)	1/333 (95% CI: 1/275–1/404)	1/64 (95% CI: 1/54–1/75)	1/1656 (95% CI: 1/900–1/3049)	1/445 (95% CI: 1/583–1/339)	1/88 (1/84–1/93)	1/125 (95% CI: 1/95–1/165)	1/331 (95% CI: 1/269–1/406)
Congenital sucrase-isomaltase deficiency prevalence	57.59/10 <sup>6</sup> birth (95% CI: 52.5–63.19)	14.06/10 <sup>6</sup> birth (95% CI: 8.69–22.72)	9.01/10 <sup>6</sup> birth (95% CI: 6.14–13.21)	247.4/10 <sup>6</sup> birth (95% CI: 178.52–342.58)	0.36/10 <sup>6</sup> birth (95% CI: 0.11–1.23)	5.05/10 <sup>6</sup> birth (95% CI: 2.94–8.68)	128.5/10 <sup>6</sup> birth (95% CI: 115.15–143.4)	63.71/10 <sup>6</sup> birth (95% CI: 36.72–110.4)	9.15/10 <sup>6</sup> birth (95% CI: 6.08–13.79)
Congenital lactase deficiency Heterozygous carrier	1/656 (95% CI: 1/591–1/729)	1/876 (95% CI: 1/561–1/876)	1/8696 ((95% CI: 1/3378–1/22222)	1/3390 (95% CI: 1/1155–1/10000)	1/6289 (95% CI: 2146–18519)	1/93 (95% CI: 1/82–1/105)	1/2193 (95% CI: 1/1653–1/2907)	1/1031 (95% CI: 1/500–2128)	1/5102 (1/2336–1/11111)
Congenital lactase deficiency Prevalence	2.32/10 <sup>6</sup> birth (95% CI: 1.88–2.87)	1.3/10 <sup>6</sup> birth (95% CI: 0.53–3.18)	0.01/10 <sup>6</sup> birth (95% CI: 0.00–0.09)	0.09/10 <sup>6</sup> birth (95% CI: 0.01–0.75)	0.03/10 <sup>6</sup> birth (95% CI: 0.00–0.22)	115.56/10 <sup>6</sup> birth (95% CI: 90.54–147.48)	0.21/10 <sup>6</sup> birth (95% CI: 0.12–0.37)	0.94/10 <sup>6</sup> birth (95% CI: 0.22–4)	0.04/10 <sup>6</sup> birth (95% CI: 0.01–0.18)
Putative congenital maltase-gluco-amylase deficiency carrier	1/1091 (95% CI: 950–1252)	1/294 (95% CI: 1/226–1/382)	1/1429 (95% CI: 1/952–2146)	0	1/1016 (95% CI: 1/635–1/1626)	1/12048 (95% CI: 1/3311–1/43478)	1/1479 (95% CI: 1/1174–1/1866)	1/2012 (95% CI: 1/685–1/5917)	1/995 (95% CI: 1/697–1/420)
Putative congenital maltase-gluco-amylase deficiency prevalence	0.84/10 <sup>6</sup> birth (95% CI: 0.64–1.11)	11.6/10 <sup>6</sup> birth (95% CI: 6.85–19.62)	0.49/10 <sup>6</sup> birth (95% CI: 0.22–1.10)	No loss of function variants in gnomAD	0.97/10 <sup>6</sup> birth (95% CI: 0.38–2.48)	0.01/10 <sup>6</sup> birth (95% CI: 0.00–0.09)	0.46/10 <sup>6</sup> birth (95% CI: 0.29–0.73)	0.25/10 <sup>6</sup> birth (95% CI: 0.03–2.13)	1.01/10 <sup>6</sup> birth (95% CI: 0.5–2.06)
Congenital trehalase deficiency Heterozygous carrier	1/575 (95% CI: 1/520–1/636)	1/1751 (95% CI: 1/887–1/3460)	1/3058 (95% CI: 1/1706–1/5464)	No loss of function variants in gnomAD	1/4386 (1/1704–1/11236)	1/2188 (95% CI: 1/1221–1/3922)	1/1961 (95% CI: 1/1488–1/2584)	1/536 (95% CI: 1/937–1/307)	1/101 (95% CI: 1/90–1/113)
Congenital trehalase deficiency Heterozygous carrier without c.90-9_106del		1/1751 (95% CI: 1/887–1/3460)			1/9009 (95% CI: 2463–32258)	1/2188 (95% CI: 1/1221–1/3922)	1/1949 (95% CI: 1/1475–1/2577)	No loss of function variants in gnomAD	1/1597 (95% CI: 1/1011–1/2525)
Congenital trehalase deficiency Prevalence	9.02/10 <sup>6</sup> birth (95% CI: 2.47–3.7)	0.33/10 <sup>6</sup> birth (95% CI: 0.08–1.27)	0.11/10 <sup>6</sup> birth (95% CI: 0.03–0.34)	No loss of function variants in gnomAD	0.05/10 <sup>6</sup> birth (95% CI: 0.01–0.34)	0.21/10 <sup>6</sup> birth (95% CI: 0.07–0.67)	0.26/10 <sup>6</sup> birth (95% CI: 0.15–0.45)	3.48/10 <sup>6</sup> birth (95% CI: 1.14–10.6)	98.47/10 <sup>6</sup> birth (95% CI: 78.07–124.12)
Congenital trehalase deficiency Prevalence without c.90-9_106del	0.2/10 <sup>6</sup> birth (95% CI: 0.14–0.3)	0.33/10 <sup>6</sup> birth (95% CI: 0.08–1.27)	0.09/10 <sup>6</sup> birth (95% CI: 0.03–0.29)		0.01/10 <sup>6</sup> birth (95% CI: 0–0.16)	0.21/10 <sup>6</sup> birth (95% CI: 0.07–0.67)	0.26/10 <sup>6</sup> birth (95% CI: 0.15–0.46)	No loss of function variants in gnomAD	0.39/10 <sup>6</sup> birth (95% CI: 0.16–0.98)

CI: confidence interval

genetic variant has ever been described. Saleheen et al. [10] did not report specific phenotypic information for the homozygous individual in their study or any functional analysis of *TREH*: c.90-9\_106del. Since this variant is relatively frequent (1.06% in gnomAD), it is unclear whether it is truly a loss-of-function variant. The estimated prevalence of CTD in other populations is very low (around 0.2 per 10<sup>6</sup> births), with the caveat that it is based on predicted loss-of-function variants only; however, this estimate is compatible with the results reported by Murray et al., who found that only one of their 847 UK patients had trehalase activities below the normal range [17].

## 5. Conclusion

Using information from a database of genetic variants, we found that the birth prevalence of CSD is relatively high, notably in European populations, and that other congenital disaccharidase deficiencies have a very low prevalence. For well-studied populations such as Europeans, our results are in keeping with published data and are probably a reasonably good approximation of true birth prevalences. For other populations, the prevalence estimates should be considered lower bounds and more studies describing variants are needed.

## Financial disclosure

None

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We thank Paul Guerry (Green Grow Scientific) for language editing and Dr. Jake Mann for sharing details about the method used to compute prevalence using gnomAD data.

## References

- [1] Burke M. Carbohydrate Intolerance and Disaccharidase Measurement - a Mini-Review. *Clin Biochem Rev* 2019;40:167–74.
- [2] Lentze MJ. The History of Maltose-active Disaccharidases. *J Pediatr Gastroenterol Nutr* 2018;66(Suppl 3):S4–6.
- [3] Kuokkanen M, Kokkonen J, Enattah NS, et al. Mutations in the translated region of the lactase gene (*LCT*) underlie congenital lactase deficiency. *Am J Hum Genet* 2006;78:339–44.
- [4] Welsh JD, Poley JR, Bhatia M, et al. Intestinal disaccharidase activities in relation to age, race, and mucosal damage. *Gastroenterology* 1978;75:847–55.
- [5] Marcadier JL, Boland M, Scott CR, et al. Congenital sucrose-isomaltase deficiency: identification of a common Inuit founder mutation. *CMAJ* 2015;187:102–7.
- [6] Gudmand-Høyer E, Fenger HJ, Skovbjerg H, et al. Trehalase deficiency in Greenland. *Scand J Gastroenterol* 1988;23:775–8.
- [7] Gao J, Brackley S, Mann JP. The global prevalence of Wilson disease from next-generation sequencing data. *Genet Med* 2019;21:1155–63.
- [8] Carter A, Brackley SM, Gao J, et al. The global prevalence and genetic spectrum of lysosomal acid lipase deficiency: A rare condition that mimics NAFLD. *J Hepatol* 2019;70:142–50.
- [9] Karczewski KJ, Francioli LC, Tiao G, et al. Genome Aggregation Database Consortium, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43.
- [10] Saleheen D, Natarajan P, Armean IM, et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* 2017;544:235–9.
- [11] Koch L. Exploring human genomic diversity with gnomAD. *Nat Rev Genet* 2020;21:448.
- [12] Uhrich S, Wu Z, Huang JY, et al. Four mutations in the *SI* gene are responsible for the majority of clinical symptoms of CSID. *J Pediatr Gastroenterol Nutr* 2012;55(Suppl 2):S34–5.
- [13] Robayo-Torres CC, Diaz-Sotomayor M, Hamaker BR, et al. 13C-Labeled-Starch Breath Test in Congenital Sucrase- isomaltase Deficiency. *J Pediatr Gastroenterol Nutr* 2018;66(Suppl 3):S61–4.
- [14] Geng L, Li DY, Ou W, et al. Congenital sucrose-isomaltase deficiency: an under-diagnosed disease in Chinese children. *BMC Pediatr* 2014;14:11.
- [15] Pontremoli C, Mozzi A, Forni D, et al. Natural Selection at the Brush-Border: Adaptations to Carbohydrate Diets in Humans and Other Mammals. *Genome Biol Evol* 2015;7:2569–84.
- [16] Madzarovová-Nohejlova J. Trehalase deficiency in a family. *Gastroenterology* 1973;65:130–3.
- [17] Murray IA, Coupland K, Smith JA, et al. Intestinal trehalase activity in a UK population: establishing a normal range and the effect of disease. *Br J Nutr* 2000;83:241–5.