Received: 26 April 2023 | Revised: 7 June 2023 | Accepted: 10 July 2023

DOI: 10.1111/hel.13008

ORIGINAL ARTICLE

Helicobacter WILEY

Genome-wide association study of *Helicobacter pylori* serological status in Latin American children

Iasmin Souza Lima[1] | Thiago Magalhães da Silva[2] | Stefan Weiss[3,4] | Georg Homuth[3] | Markus M. Lerch[4] | Camila A. Figueiredo[5] | Neuza Maria Alcantara-Neves[5] | Maurício Lima Barreto[6] | Cintia Rodrigues Marques[1]

[1]Multidisciplinary Institute in Health, Federal University of Bahia, Vitória da Conquista, Brazil

[2]Department of Biological Sciences, State University of Southwest Bahia, Jequié, Brazil

[3]Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany

[4]Department of Medicine A, University Medicine Greifswald, Greifswald, Germany

[5]Institute of Health Sciences, Federal University of Bahia, Salvador, Brazil

[6]Center of Data and Knowledge Integration for Health, Instituto Gonçalo Muniz, Fundação Osvaldo Cruz, Salvador, Brazil

Correspondence
Thiago Magalhães da Silva, Department of Biological Sciences, State University of Southwest Bahia, Jequié, Bahia, Brazil.
Email: thiago@uesb.edu.br

Abstract

Background: Few genome-wide association studies (GWAS) on *Helicobacter pylori* infection susceptibility have been conducted for admixed populations from developing countries. Here, we performed a GWAS to identify genetic factors associated with *H. pylori* serostatus in a cohort of admixed children from a large Latin American urban center.

Methods: A cross-sectional study involving 1161 children from 4 to 11 years old living in poor areas of Salvador, in northeastern Brazil. Logistic regression analysis was performed to detect associations between single-nucleotide variants (SNVs) and *H. pylori* seropositivity, assuming an additive genetic model. Enrichment analyses were conducted using the MAGMA v1.10 software.

Results: We found 22 SNVs to be suggestively associated ($p < 10^{-5}$) with *H. pylori* seropositivity. The most suggestive SNV was the rs77955022 ($p = 4.83e\text{-}07$) located in an intronic region of *EXOC3* at 5p15.33. The second most suggestively associated SNV was rs10914996 ($p = 8.97e\text{-}07$), located in an intergenic region at 1p34.3. Furthermore, we were able to replicate three SNVs ($p < 0.05$) in the Study of Health in Pomerania (SHIP) cohort: the rs2339212 and rs4795970, both located at 17q12 near *TMEM132E*, as well as the rs6595814, an intronic variant of *FBN2* at 5q23.3. The enrichment analysis indicated the participation of genes and metabolic pathways related to the regulation of the digestive system and gastric acid secretion in the risk of seropositivity for *H. pylori*.

Conclusions: Additional studies are required to validate these association findings in larger population samples and to get insight into the underlying physiological mechanisms.

KEYWORDS
childhood, ELISA, genetics, *Helicobacter pylori*, susceptibility

1 | INTRODUCTION

Helicobacter pylori is a spiral-shaped, gram-negative, microaerophilic bacterium that colonizes the human gastric mucosa, causing gastrointestinal diseases such as gastritis,[1] peptic ulcer,[2] and gastric cancer.[3] At least half of the world's population is infected by this pathogen. Most infections occur during childhood, with socioeconomic status and housing conditions as major risk factors.[4] Developing countries have higher prevalence rates, with estimates of 70.1% for the African region, 69.4% for South American countries,

and 66.6% for Western Asia. In Brazil, the estimated prevalence of *H. pylori* infection amounts to 71.2%.[5]

The local inflammation triggered by *H. pylori* alters the gastric mucosal microenvironment, modulating the host's immune response.[6] It is also known that host genetic factors can interfere with the infection by the pathogen.[7] Several studies have shown that variants in the host genes *IL-10*, *IL-1β*, *IL-1R*, *TNFα*, and *TLRs* are associated with *H. pylori* infection phenotypes.[8] These data demonstrate that the individual genetic makeup influences susceptibility to *H. pylori* infection, especially in terms of anti-*H. pylori* IgG production. Studies aiming to identify genetic factors associated with the acquisition of *H. pylori* in populations from developing countries, however, are still scarce. Furthermore, to the best of our knowledge, no GWASs have so far been conducted to specifically investigate genetic variants involved in *H. pylori* susceptibility in children. Thus, we performed a GWAS for *Helicobacter pylori* infection as reflected in significant anti-*H. pylori* IgG levels using a study sample based on children of a Brazilian urban population highly admixed and with a remarkable African ancestry.

2 | MATERIALS AND METHODS

2.1 | Study population and biological samples collection

The discovery study was based on the SCAALA (Social Changes, Asthma, and Allergy in Latin America) cohort, including 1161 of a total of 1445 children who comprised the original cohort, aged 4–11 years and living in outlying areas in Salvador's city, Bahia, Brazil. The selection criteria for the study population are shown in Figure S1. As shown in Table S1, no significant differences for *H. pylori* seroprevalence, sex, age, and Native American ancestry proportion were observed between children included and excluded from the study. Individuals included in the study, however, had a lower proportion of African ancestry (and consequently greater European ancestry) compared with those who were excluded. The collection of data and biological samples of the members of the cohort occurred in 2005.[9] Ethics approval for this study was obtained from the National Council of Ethics of Brazil (CONEP, resolution number 15895, Brasília) and from Research Ethics Committee of the UFBA Institute of Public Health (register 003–05/CEP-ISC). Written informed consent was obtained from the guardians of each child. Furthermore, the research was conducted in accordance with the principles outlined in the Declaration of Helsinki.

2.2 | Serological detection of anti-*Helicobacter pylori* IgG

The presence of IgG antibodies in blood samples was determined by ELISA using a commercially available kit (Diamedix, Miami, FL, USA) following the manufacturer's instructions. The cutoff was determined by an index value obtained by the ratio of sample absorbance to the absorbance of a calibrator (a solution containing human serum or defibrinated plasma, with IgG antibodies weakly reactive with *H. pylori* and 0.1% sodium azide). A ratio >1.1 was considered positive. Borderline subjects were removed from the analysis.[10] A total of 1370 subjects were successfully tested for anti-*H. pylori* IgG. Of these, 919 (67.1%) were negative, 363 (26.5%) were positive, and 88 (6.4%) were borderline. Of the 1282 individuals with valid *H. pylori* serostatus data, 1161 had genome-wide genotypic data available.

2.3 | Genotyping and quality control

The DNA extraction and genotyping methodologies were previously described.[11] Single-nucleotide variants were genotyped for the SCAALA population using the commercial Illumina HumanOmni2.5–8 Kit BeadChip panel (www.illumina.com). Closely related individuals ($n = 64$) were excluded from the total of those successfully genotyped ($n = 1310$) using the method implemented in the REAP software (Related Estimation in Admixed Populations).[12] Of the remaining 1246 individuals, 1161 had serological data for *H. pylory* available. Markers with genotyping call rates of less than 0.98 and individuals with missing data for more than 10% of SNVs were excluded. After quality control steps, a total of 1,912,180 SNVs with Minor Allele Frequency (MAF) ≥1% were available for analysis. The proportion of individual African, European, and Native American ancestry was estimated using the ADMIXTURE software.[13] Briefly, a tri-hybrid model was used in the unsupervised analyses performed. As external panels, samples of African and European individuals from the HapMap project and 93 Native American of the Human Genome Diversity (HGDP) project were used. A total of 370.539 SNVs shared by samples from the HapMap, HGDP and the study population were used for genetic ancestry estimate. Genotyping data have been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk.ega/), which is hosted by the EBI, under accession number EGAS00001001245.

2.4 | Statistical analysis

The association with *H. pylori* seropositivity was analyzed using binary logistic regression models in the software Plink 1.9,[14] assuming an additive genetic effect. Analyses were adjusted for the first three main components of genetic variance (PC1, PC2 and PC3), age, and sex. An association was considered suggestive when $5.0 \times 10^{-8} < p\text{-value} < 10^{-5}$ (genome-wide significance: $p\text{-value} < 5.0 \times 10^{-8}$).[15] The genotypes were imputed using the IMPUTE2 package[16] on the public panel from 1000 Genomes Project Phase I data "version 3,"[17] which contained 1092 individuals of various ethnicities. Quality control was carried out once more following imputation and the SNPs, which presented a MAF lower than 1%, a deviance in the Hardy–Weinberg equilibrium ($p < 10–4$) or had a genotyping call rate of under 95% were excluded.

2.5 | Replication strategy

Replication of the suggestive associations for *H. pylori* seropositivity was assessed in the Study of Health in Pomerania (SHIP).[18] The SHIP study consists of two independent population-based cohorts (SHIP-START and SHIP-TREND) from West Pomerania, a region in the north-east of Germany. Participants in the SHIP project are adult individuals aged 20–79 years recruited between 1997 and 2001 (SHIP-START $n = 6265$) and between 2008 and 2012 (SHIP-TREND $n = 8016$). The SHIP cohort comprised the study population of one of the largest *H. pylori* serostatus GWAS ever performed,[19] which justified its choice as a replication set for this study. Anti–*H pylori* IgG antibody titers were measured using commercial enzyme immunoassays (Pyloriset EIA-G III ELISA; Orion) according to the manufacturer's instructions, and results from the SHIP-START and SHIP-TREND cohorts for *H. pylori* seropositivity were then meta-analyzed. A total of 4802 individuals (52.2% women and 56.2% *H. pylori* seropositive) were included in the replication analysis.

2.6 | Enrichment analysis

Enrichment analysis was performed using the MAGMA v1.10 software.[20] Briefly, in a first step, a principal component regression model including sex, age, and individual genetic ancestry (PC1, PC2, and PC3) as covariates is fitted and the *p*-values and gene correlation matrix are then used in the second part to perform the pathway analysis. For gene association analysis, we defined regions 5 kb up/downstream from the initial and final transcription sites for 19,427 genes, according to the GRCh37/hg19 public database of catalogued genes. Empirical *p*-values were obtained through permutational methods, being performed 100,000 simulations. Gene sets used in pathway analysis were obtained from Msigdb v7.0 for "Curated gene sets" and "GO (Gene Ontology) terms," including a total of 17,027 metabolic pathways.

2.7 | In silico functional annotation

To assess the potential role of investigated SNVs as expression quantitative trait locus (eQTL) on stomach and/or whole blood cells, we used data available from the Genotype-Tissue Expression (GTEx) Portal (www.gtexportal.org/home/). The RegulomeDB database was also used to search for SNVs with known and predicted regulatory elements in the human genome, including regions of DNase hypersensitivity, binding sites of transcription factors, and promoter regions characterized previously for transcription regulation. A description of the scoring scheme used by the RegulomeDB platform is available at: https://regulomedb.org/regulome-help/.

3 | RESULTS

The general characteristics of the study population are described in Table S2. The majority of the sample was male (53.49%), with a mean age of 6.3 years (SD = 1.67). The prevalence of seropositivity for *H. pylori* was 27.74%. The averages of global African, European, and Native American ancestry in the overall sample were, respectively, 50.4% (SD = 14.1%), 43.2% (SD = 13.0%) and 6.4% (SD = 3.1%).

We found 22 SNVs to be suggestively associated with *Helicobacter pylori* seropositivity (Table 1 and Figure 1A). Of this total, two variants stood out in our study: the more suggestively associated was rs77955022-T allele (OR = 2.27; 95% CI = 1.65–3.13; $p = 4.83e\text{-}07$), located on chromosome 5p15.33 in an intron of *EXOC3* encoding the exocystic complex component 3 (Figure 2A,B). The second most suggestively associated SNV was rs10914996-G allele (OR = 0.61; 95% CI = 0.50–0.74; $p = 8.97e\text{-}07$), located in in an intergenic region at the 1p34.3 locus (Figure 2C,D). It is in moderate linkage disequilibrium (LD) with two SNVs in the same locus that have also been suggestively associated with *H. pylori* seropositivity in our discovery set: rs7548924 ($r^2 = 0.63$) and rs693227 ($r^2 = 0.76$). The observed and expected *p*-values and the value of the genomic inflation factor (λ), which was 1.01, indicate little chance of the results being influenced by population stratification (Figure 1B**).** The allele frequency distribution of the 22 Top SNVs among African, European, and Native American populations is shown in Table S3. Among the 22 Top SNVs, four were ancestry informative markers (AIMs), showing an allele frequency difference greater than or equal to 30.0% comparing these parental populations. The predict probability of carrying the effect allele of these AIMs according to the proportions of individual genetic ancestry is shown in Figure S2. As seen in the Figure, the most pronounced effect of African and European ancestry was observed for the predict probability of carrying the risk allele for *H. pylori* seropositivity of the rs6595814 (G allele), located in the *FBN2* gene.

An expression heat map of the protein-coding genes to which the top SNVs were mapped was constructed based on the GTEx v8 database, which includes 54 different tissue types (Figure S3). A higher relative expression in gastric tissue is observed mainly for *EXOC3* and *SLC9A3-OT1* (*CTD-2228K2.5*) genes. Interestingly, an enrichment of the mapped genes to the differentially expressed genes (DEG) sets in stomach compared with all other tissue was also observed, although this trend was not statistically significant (Figure S4).

Table 1 also presents the replication results in the Study of Health in Pomerania (SHIP) cohort. Twenty SNVs suggestively associated with *H. pylori* seropositivity in the discovery set were tested for association in SHIP. Three of these were replicated ($p < 0.05$) with the same effect direction as in the discovery sample: the rs2339212-A allele (OR = 1.14; 95% CI = 1.048–1.241; $p = 0.002$) and the rs4795970-A allele (OR = 0.879; 95% CI = 0.808–0.957; $p = 0.003$) are in high LD ($r^2 = 0.96$). Both are located near *TMEM132E* encoding the transmembrane protein 132E at the 17q12 locus (Figure S5A,B). Furthermore, the rs6595814-G allele (OR = 1.39; 95% CI = 1.01–1.29; $p = 0.046$) located in *FBN2* encoding fibrillin at the 5q23.3 locus (Figure S5C,D).

The results of the gene-based analysis are shown in Table 2 and Figure S6. Considering the Bonferroni correction, two genes located on chromosome 20 (*RTFDC1* and *GCNT7*) were significantly

TABLE 1 The 22 SNVs suggestively associated with seropositivity for *Helicobacter pylori* in the discovery and replication sets.

SNV information				SCAALA (Discovery)			SHIP (Replication)		
SNV	CHR	Gene[a]	EA	EAF	OR (95%CI)	*p*	EAF	OR (95%CI)	*p*
rs10914996	1	*MIR552*	G	0.42	0.61 (0.50–0.74)	8.97E-07	0.41	1.03 (0.95–1.13)	0.480
rs7548924	1	*MIR552*	T	0.31	0.59 (0.48–0.73)	1.56E-06	0.23	0.97 (0.88–1.01)	0.601
rs693227	1	*MIR552*	A	0.44	0.63 (0.52–0.77)	4.02E-06	0.46	1.00 (0.92–1.09)	0.963
rs10208804	2	MIR4262	T	0.12	1.89 (1.43–2.50)	7.07E-06	0.14	0.99 (0.88–1.12)	0.918
rs116318062	4	*RPF2P2*	T	0.03	3.2 (1.91–5.36)	9.78E-06	NA	NA	NA
rs77955022	5	*EXOC3*	T	0.08	2.27 (1.65–3.13)	4.83E-07	0.21	1.04 (0.94–1.15)	0.458
rs6595814	5	*FBN2*	G	0.38	**1.58 (1.30–1.92)**	**4.83E-06**	0.13	**1.14 (1.01–1.30)**	**0.046**
rs9295348	6	*RPS6KA2*	C	0.33	1.61 (1.32–1.95)	1.75E-06	0.30	1.04 (0.95–1.14)	0.390
rs2371541	7	*PCLO*	G	0.45	1.54 (1.27–1.86)	9.64E-06	0.62	0.99 (0.91–1.08)	0.883
rs2812531	10	*C10orf35*	T	0.24	0.57 (0.45–0.73)	7.09E-06	0.44	0.94 (0.87–1.03)	0.194
rs1872753	10	*LINC01443*	G	0.31	0.63 (0.51–0.77)	7.18E-06	0.15	1.04 (0.93–1.07)	0.456
rs7098808	10	*FAM107B*	T	0.08	2.02 (1.48–2.75)	7.59E-06	0.06	1.06 (0.89–1.28)	0.508
rs7105675	11	*OPCML*	T	0.35	1.62 (1.33–1.98)	1.85E-06	0.34	1.01 (0.92–1.10)	0.893
rs75137136	11	*LOC105376634*	A	0.02	4.20 (2.30–7.69)	3.18E-06	NA	NA	NA
rs9515577	13	*LOC107984621*	T	0.07	2.11 (1.52–2.94)	8.65E-06	0.06	0.95 (0.79–1.14)	0.579
rs12050717	15	MYO5A	T	0.11	0.44 (0.31–0.62)	3.20E-06	0.03	1.10 (0.87–1.40)	0.423
rs12149021	16	*CDH13*	T	0.21	1.70 (1.36–2.12)	2.78E-06	0.27	1.03 (0.94–1.14)	0.473
rs8055256	16	*CDH13*	G	0.21	1.70 (1.36–2.13)	4.33E-06	0.18	1.08 (0.97–1.20)	0.168
rs17176567	16	*CDH13*	C	0.21	1.68 (1.35–2.10)	4.92E-06	0.27	1.03 (0.94–1.13)	0.502
rs2339212	17	*TMEM132E*	A	0.5	**1.55 (1.28–1.87)**	**4.99E-06**	0.40	**1.14 (1.05–1.24)**	**0.002**
rs4795970	17	*TMEM132E*	A	0.49	**0.65 (0.54–0.78)**	**6.13E-06**	0.60	**0.88 (0.81–0.96)**	**0.003**
rs455980	21	*USP25*	T	0.07	2.14 (1.53–2.99)	9.35E-06	0.08	1.05 (0.90–1.22)	0.562

Note: Odds Ratio (OR) obtained by logistic regression model adjusted for sex, age and three principal components (PC1, PC2 and PC3) of genetic variance (SCAALA cohort) and age and sex (SHIP cohort).

All bold values show $p<0.05$ in the replication set (SHIP).

Abbreviations: EA, Effect Allele; EAF, Effect Allele Frequency.

[a]For intergenic SNVs, the nearest mapped gene (based on UCSC database) is indicated.

associated ($p<2.64e-06$) with seropositivity for *H. pylori* in our discovery population. Suggestive associations ($p<5.29e-05$) were observed for *ATP4B* on chromosome 13 and *ITPKA* on chromosome 15. When correcting for multiple tests using the permutational method, however, several genes were associated with P-perm <0.05, with the 20 most significant shown in Table 2. Enrichment analysis based on gene sets involved in known biological pathways were also conducted from the gene association results (Table 3). When correcting for multiple tests, only the GO: perisynaptic extracellular matrix pathway was significantly associated ($p<0.05$) with seropositivity for *H. pylori*. The top 10 gene sets arranged according to increasing order of nominal *p*-value are shown in Table 3.

In addition, we also performed a replication analysis of 23 variants at the 4p14 *TLR* locus, which was previously demonstrated to be associated with *H. pylori* seroprevalence in a GWAS based on a sample of European ancestry[19] in our discovery sample (Table S4). Seven SNVs replicated with $p<0.05$, including the nonsynonymous SNV rs4833095 in *TLR1*. Furthermore, the replicated SNVs are AIMs, with remarkable differences of allelic frequencies between African and European populations.

4 | DISCUSSION

In the present study, we performed the first GWAS for *H. pylori* seropositivity conducted in a pediatric and highly admixed population from a large Latin American urban center. We report just over 20 SNVs that were suggestively associated with seropositivity for *H. pylori*, three of which were replicated in an independent cohort with age structure and genetic ancestry remarkably different from that of the discovery sample.

In our discovery population, the SNV rs77955022 was the top variant suggestively associated with *H. pylori* seropositivity. It is located in an intronic region of *EXOC3* encoding the exocyst complex component 3 protein. EXOC3 is part of a highly conserved exocystic complex that comprises the eight protein subunits EXOC1-EXOC8.[21] The exocystic complex directs secretory vesicles toward the plasma membrane in the cellular exocytosis process.[22] Although EXOC3 has been implicated in infectious processes involving different microbial pathogens such as *Legionella pneumophila*,[23] *Porphyromonas gingivalis*,[24] and *Escherichia coli*,[25] our study is the first to report an association between variants in its encoding gene and *H. pylori* seropositivity.

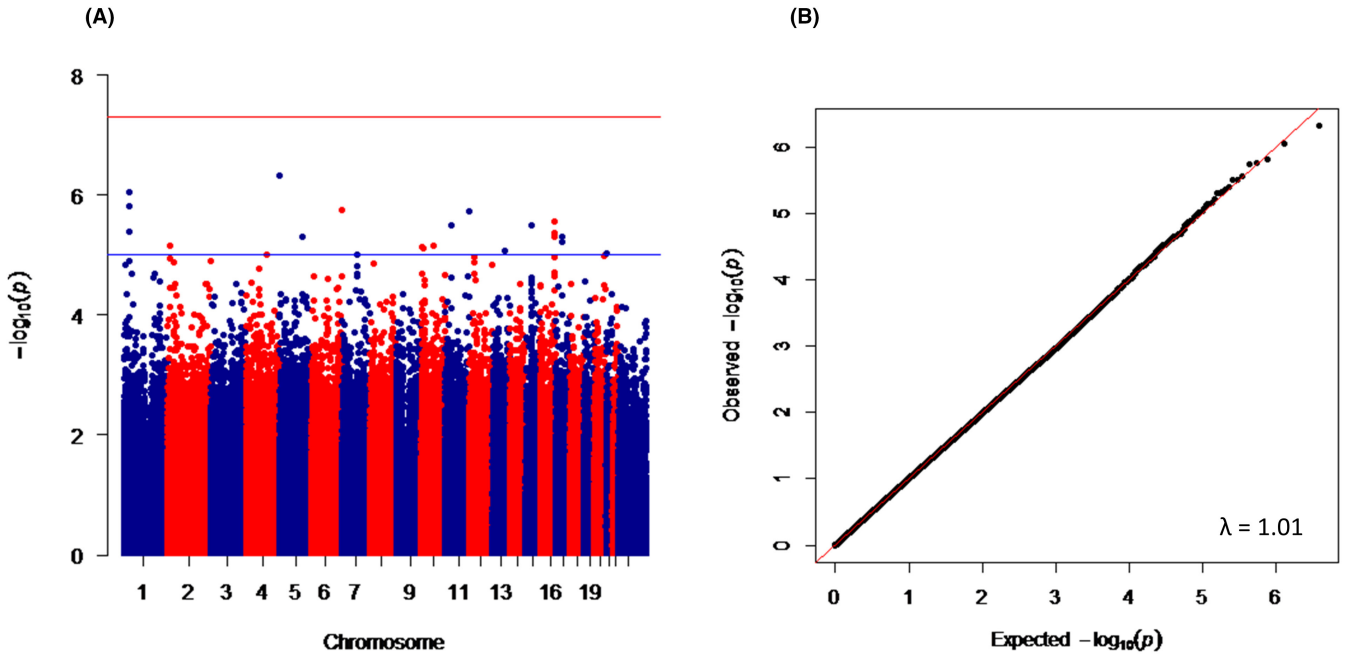

FIGURE 1 Manhattan and QQ plots for *H. pylori* seropositivity in children of SCAALA cohort. (A) Chromosome position is plotted on the *x*-axis and $-\log_{10}$ *p*-values on the *y*-axis. The blue (bottom) and red (upper) lines mark the suggestive (1.0×10^{-5}) and significant (5.0×10^{-8}) association thresholds at the genome-wide level, respectively. (B) Single-nucleotide variants (SNVs) are displayed as black dots and the red line corresponding to the null hypothesis of no true association.

Interestingly, based on data from the GTEx portal (https://gtexportal.org/home/snp/rs77955022 on February 22, 2023), rs77955022 represents an eQTL for *SLC9A3* (*NHE-3*) as well as for the gene that encodes its antisense RNA (*SLC9A3-AS1*) in gastric tissue. Specifically, the rs77955022-T effect allele associated with increased risk to *H. pylori* seropositivity in our population was associated with reduced *NHE-3* transcript abundance (and *SLC9A3-AS1* overexpression) in gastric cells (Figure 3A,B). In rats, the NHE-3 protein has been localized to the apical membrane of gastric parietal cells, where it colocalizes with the Beta-subunit of the gastric H+-K+-ATPase.[26] In humans, in turn, NHE-3 exhibits localization in the basolateral membrane of gastric surface mucous cells.[27] Based on its location in the gastric tissue, it is assumed that NHE-3, which is a Na/H exchanger, may be involved in alternative proton secretion pathways as well as in gastric mucosa cell protection processes through the diffusion of these protons into the mucous cells.[26,27] The role of genetic variants negatively affecting *NHE-3* expression in gastric cells on the risk of *H. pylori* infection and seroprevalence, however, needs further investigation. On the contrary, considering the ubiquitous *EXOC3* expression (as seen in Figure S1), we cannot rule out the association of rs77955022 with serostatus for *H. pylori* as a result of LD with other putative risk variants at 5p15.33. In this sense, variants with high regulatory potential (RegulomeDB score = 1b and 1f) of the aryl hydrocarbon receptor repressor (*AHRR*) stand out (Table S5). Aryl hydrocarbon receptor (AHR) is a ligand-dependent transcription factor that belongs to the superfamily of basic Helix–Loop–Helix/Per-ARNT-Sim (bHLH/PAS) proteins and is activated by environmental toxins and other low-molecular-weight compounds.[28] In turn, the AHRR has been identified as a negative regulator of *AHR*, competing for binding to the aryl hydrocarbon nuclear receptor translocator (ARNT) and consequently repressing AHR-induced downstream gene expression.[29] Previous studies have demonstrated that AHR plays a pivotal role in response to microorganisms, binding, and sensing distinct pathogen associated molecular patterns (PAMPs) and regulating immune pathways in response to these PAMPs.[30] Furthermore, AHRR is highly expressed in immune cells of barrier organs, such as the skin and intestine, significantly influencing the regulation of the inflammatory response in these sites.[31] A recent study demonstrated that AHR and AHRR expression was negatively correlated with exposure of gastric cells to *H. pylori* both in vivo and in vitro. In addition, AHR and AHRR silencing by siRNAs resulted in an increased synthesis of proinflammatory cytokines such as IL-8, IL-1B and TNF-a, suggesting the involvement of these factors in the antibacterial response to *H. pylori*.[32] If confirmed in other independent study populations, our findings may corroborate the role of AHRR locus in susceptibility to *H. pylori* infection, especially in childhood.

The second more suggestively associated variant found in our discovery population, rs10914996, is an intergenic SNV in the 1p34.3 region. Using imputed data for chromosome 1, several SNVs in LD with rs10914996 were suggestively associated with seropositivity for *H. pylori*, including some with predicted regulatory function according to the ENCODE data, such as rs1170689 ($r^2 = 0.79$, RegulomeDB score = 2b) and rs10915000 ($r^2 = 0.92$, RegulomeDB score = 1f) (Table S6). Cytogenetic analysis has shown that the loss of this locus is involved in colorectal cancer.[33,34] There are still no studies in the literature relating the 1p35.1 locus to *H. pylori* or other infectious agents. More studies are required to better understand the role this region on *H. pylori* acquisition.

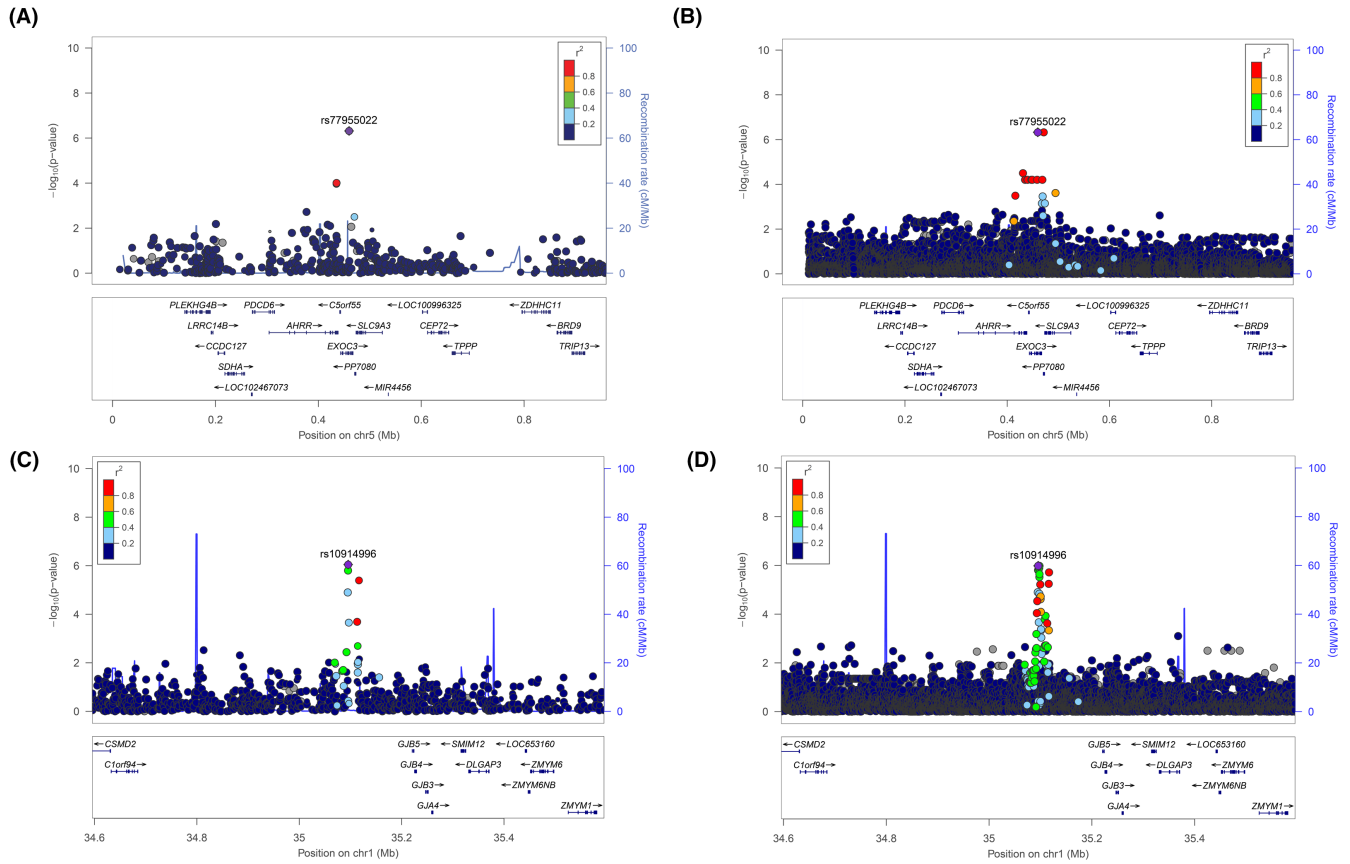

FIGURE 2 Regional plot of the two loci most suggestively associated with *H. pylori* seropositivity in the SCAALA cohort. (A, C) represent the association signals for the SNVs genotyped at loci 5p15.33 and 1p35.1, respectively. (B, D) refer to the association signals for the imputed SNVs at loci 5p15.33 and 1p35.1, respectively. The top SNV in each genomic region is represented by a purple diamond. Pairwise r^2 values are from hg19/1000 Genomes American data (November 2014 release).

We were able to replicate three SNVs suggestively associated with *H. pylori* seropositivity in our discovery set in the SHIP cohort. Two of these, rs2339212 and rs4795970, are in high linkage disequilibrium and are located at locus 17q12, close to *TMEM132E* and upstream of a cluster of cytokine-coding genes. Interestingly, variants at the 17q12 locus including rs12601721 that is also located close to *TMEM132E* were found to be associated with the levels of circulating Macrophage Inflammatory Protein 1 Beta (MIP-1B or CCL4) in a GWAS conducted elsewhere.[35] In addition, rs12601721 obtained from the imputed dataset for chromosome 17 was nominally associated with *H. pylori* seropositivity in our discovery population ($p = 0.034$), showing low LD with rs2339212 ($r^2 = 0.11$). This result seems biologically plausible considering that MPI-1B produced by monocytes, macrophages, lymphocytes, and other cell types has the capacity to recruit a variety of immune cells to sites of microbial infection.[36] While cytokine production represents a known critical component influencing infection/disease outcomes associated with *H. pylori*,[8] this is the first study reporting a putative involvement of the 17q12 locus in the susceptibility for infection by this bacterium. Further studies, therefore, are needed to better understand the impact of variants in this genomic region on the risk of *H. pylori* infection and seroprevalence. The third SNV replicated in the SHIP cohort was rs6595814, which is an intronic variant located in *FBN2* at the 5q23.3 locus. This gene encodes a component of connective tissue microfibrils that might be involved in elastic fiber assembly.[37] The rs6595814 represents an eQTL for *FBN2* in whole blood cells (https://gtexportal.org/home/snp/rs6891484 on 17 April 2023) and is also an AIM, with the G allele showing a 55.0% frequency difference between African and European populations of 1000 Genomes Project (Table S3). Other AIMs with potential regulatory function according to ENCODE data and in LD with rs6595814 were also identified in the analysis using imputed dataset for chromosome 5 (Table S7). Although associations of *FBN2* variants with *H. pylori* infection have not been previously reported, it has been associated with gut microbiome diversity in healthy individuals.[38] Specifically, certain *FBN2* SNVs were negatively associated with the relative abundance of *Ruminococcus flavefaciens* in individual gut microbiota. Interestingly, a lower abundance of the genus *Ruminococcus* in fecal samples has been reported in *H. pylori*–infected patients.[39] This suggests a possible involvement of genes that influence the diversity of the gut microbiota with *H. pylori* infection, which deserves further investigation.

Among the genes most significantly associated in the enrichment analysis, *ATP4B* draws special attention because it is directly involved in gastric acid secretion.[40] In fact, this gene encodes the beta subunit of the gastric H+, K+-ATPase, showing altered expression in the

presence of atrophic gastritis and other outcomes related to *H. pylori*, such as gastric cancer.[41] Furthermore, although not statistically significant, a predominance of metabolic pathways related to the regulation of the digestive system and gastric acid secretion was also observed in the analysis of gene sets. Also noteworthy is the presence of metabolic pathways related to the transduction of cellular signals, such as those involved in the activity of G protein-coupled glutamate receptors. Interestingly, the role of glutamate receptors in inducing tolerogenic reprogramming of dendritic cells and in favoring Treg cell differentiation in the context of *H. pylori* infection has been previously reported.[42] These mechanisms may promote pathogen resistance in the gastric mucosa and persistent infection.[43] Finally, it is also noteworthy the presence of *AHRR* and *EXOC3* among the genes most associated with seropositivity for *H. pylori*, corroborating the results of the single-allele association strategy.

Here, we also were able to replicate association of 4p14 *TLR* locus with seropositivity for *H. pylori*. Importantly, all the SNVs replicated are AIMs, with considerable allele frequency differences between Africans and Europeans. This reinforces the potential involvement of this locus in the acquisition of the infection in childhood and/or in the host immune response to *H. pylori*. These results also suggest that at least part of the ethnic differences observed for *H. pylori* prevalence can be due to differentially distributed genetic variants between populations of African and European ancestry.[44]

The present study has some limitations. First, the serological diagnosis by ELISA does not allow distinguishing between past and active infection. This, however, is somewhat mitigated when we consider that *H. pylori* infection is usually stable over time and persists unless eradication treatment is administered.[45] Second, the performance of the ELISA in younger children is usually poorer than that observed in older ones, with a significant reduction in the sensitivity and specificity of assay.[46] In light of this, we analyzed the SNVs suggestively associated with *H. pylori* seropositivity in the subgroup of older children only (>6 years). While the smaller sample size ($n = 713$) weakened the statistical power and thus the statistical significance of the signals, the effect magnitudes for all SNVs were consistent and comparable to that observed for the overall dataset (Table S8). Third, the replication set (SHIP cohort) has a very different age structure and genetic architecture from our discovery set

TABLE 2 Results of gene-based association analysis conducted in the MAGMA software.

Gene	Chromosome	*p*	P-perm[a]
RTF2	20	1.17E-07	1.00E-05
GCNT7	20	6.38E-07	1.00E-05
ATP4B	13	2.54E-05	3.00E-05
ITPKA	15	4.97E-05	3.00E-05
PDCD5	19	9.22E-05	4.00E-05
DCP1A	3	0.000151	0.00014
CD99L2	X	0.00022	0.00016
UGT1A7	2	0.000238	0.00024
UGT1A9	2	0.000277	0.00025
MAP6	11	0.000282	0.00026
UNC45B	17	0.000283	0.00026
WNK3	X	0.000284	0.00026
AHRR	5	0.000302	0.00028
TSPAN11	12	0.000319	0.0003
TNFSF14	19	0.000328	0.0003
FAM209B	20	0.000341	0.00034
EXOC3	5	0.000345	0.00035
SLC35G3	17	0.000387	0.00036
HES7	17	0.000412	0.00039
RSPO4	20	0.000419	0.00042

[a]*p*-value corrected by permutation test (total of 100,000 fixed permutations to each tested gene).

TABLE 3 Results of pathway analysis for *H. pylori* seropositivity using MAGMA software.

Pathway	No. genes	*p*	P-cor
GO: perisynaptic extracelular matrix	5	2.81E-06	0.047
GO: negative regulation of digestive system process	17	1.73E-05	0.294
GO: adenylate cyclase inhibiting g protein coupled glutamate receptor signaling pathway	9	1.88E-05	0.320
GO: G protein coupled glutamate receptor activity	9	1.88E-05	0.320
GO: synapse associated extracellular matrix	6	2.56E-05	0.434
GO: dendrite terminus	13	5.99E-05	1.0
GO: dendritic growth cone	9	0.0001	1.0
GO: regulation of digestive system process	43	0.0001	1.0
GO: regulation of gastric acid secretion	11	0.0003	1.0
GO: inorganic diphosphate phosphatase activity	5	0.0004	1.0

Abbreviations: GO, Gene Ontology; P-cor, Bonferroni corrected *p*-value.

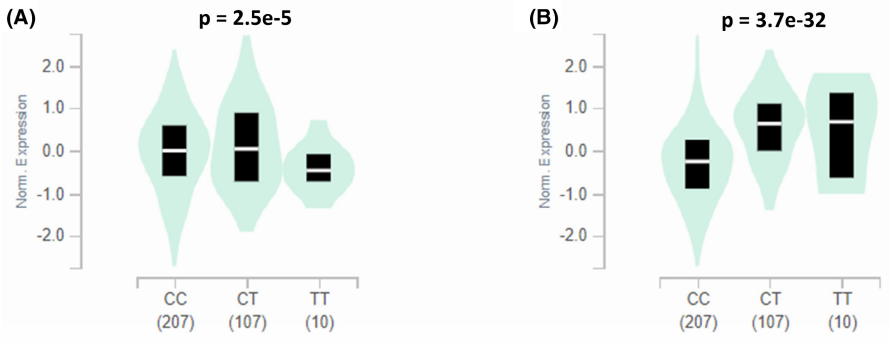

FIGURE 3 Transcript levels of *SLC9A3* (A) and *SLC9A3-AS1* (B) in gastric tissue in relation to *EXOC3* rs77955022 genotypes.

(SCAALA cohort). Indeed, some of the Top SNVs suggestively associated in our discovery set, such as rs116318062 and rs75137136, could not be tested in the SHIP cohort due to the fact that they are monomorphic or extremely rare in European populations (Table S3). Finally, the relatively small sample size precludes the identification of genetic variants with a more modest effect on the risk of *H. pylori* infection and might cause the absence of genome-wide significant associations (p-value $<5.0\times10^{-8}$). Conducting joint analyses including large study samples that are genetically more similar to our discovery set may help to overcome this limitation in the future by increased statistical power.

5 | CONCLUSION

To our best knowledge, we conducted the first GWAS for *Helicobacter pylori* serological status in a cohort of Latin American children. In this study, we detected SNVs suggestively associated with *H. pylori* seroprevalence not previously described in the literature and we replicated three of these in an independent European population-based cohort. Thus, these novel associations should be validated in larger study samples ethnically similar and age-adjusted to our discovery sample. If validated, functional studies to understand the physiological impact of these variants on *Helicobacter pylori* infection would be desirable.

AUTHOR CONTRIBUTIONS

CRM and TMS conceived the study. MLB is the cohort coordinator, providing samples and data. TMS coordinated the genetic analyses. ISL, CRM, and TMS wrote the manuscript. All the authors contributed with discussion on the results and on the manuscript. The authors SW, GH, MML, and NMA-N contributed with data, laboratorial analyses or statistical analyses.

FUNDING INFORMATION

This work was supported by the Department of Science and Technology (DECIT, Ministry of Health) and National Fund for Scientific and Technological Development (FNDCT, Ministry of Science and Technology), Funding of Studies and Projects (FINEP, Ministry of Science and Technology, Brazil), Coordination of Improvement of Higher Education Personnel (CAPES, Ministry of Education, Brazil). ISL was supported by the Foundation for Research Support of the State of Bahia (FAPESB).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Thiago Magalhães da Silva https://orcid.org/0000-0001-5171-0123
Stefan Weiss https://orcid.org/0000-0002-3553-4315

REFERENCES

1. Warren JR, Marshall B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet*. 1983;1(8336):1273-1275.
2. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet*. 1984;1(8390):1311-1315.
3. Schistosomes, liver flukes and *Helicobacter pylori*. *IARC Monogr Eval Carcinog Risks Hum*. 1994;61:1-241.
4. Yuan C, Adeloye D, Luk TT, et al. The global prevalence of and factors associated with *Helicobacter pylori* infection in children: a systematic review and meta-analysis. *Lancet Child Adolesc Health*. 2022;6(3):185-194.
5. Hooi JKY, Lai WY, Ng WK, et al. Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis. *Gastroenterology*. 2017;153(2):420-429.
6. Nejati S, Karkhah A, Darvish H, Validi M, Ebrahimpour S, Nouri HR. Influence of *Helicobacter pylori* virulence factors CagA and VacA on pathogenesis of gastrointestinal disorders. *Microb Pathog*. 2018;117:43-48.
7. Malaty HM, Engstrand L, Pedersen NL, Graham DY. *Helicobacter pylori* infection: genetic and environmental influences. A study of twins. *Ann Intern Med*. 1994;120(12):982-986.
8. Figueiredo CA, Marques CR, Costa Rdos S, da Silva HB, Alcantara-Neves NM. Cytokines, cytokine gene polymorphisms and *Helicobacter pylori* infection: friend or foe? *World J Gastroenterol*. 2014;20(18):5235-5243.
9. Barreto ML, Cunha SS, Alcantara-Neves N, et al. Risk factors and immunological pathways for asthma and other allergic diseases in children: background and methodology of a longitudinal study in a large urban center in northeastern Brazil (Salvador-SCAALA study). *BMC Pulm Med*. 2006;6:15.
10. Dattoli VC, Veiga RV, da Cunha SS, Pontes-de-Carvalho LC, Barreto ML, Alcantara-Neves NM. Seroprevalence and potential risk factors for *Helicobacter pylori* infection in Brazilian children. *Helicobacter*. 2010;15(4):273-278.

11. Kehdy FS, Gouveia MH, Machado M, et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc Natl Acad Sci U S A*. 2015;112(28):8696-8701.
12. Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N. Estimating kinship in admixed populations. *Am J Hum Genet*. 2012;91(1):122-138.
13. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655-1664.
14. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575.
15. Reed E, Nunez S, Kulp D, Qian J, Reilly MP, Foulkes AS. A guide to genome-wide association analysis and post-analytic interrogation. *Stat Med*. 2015;34(28):3769-3792.
16. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529.
17. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-1073.
18. Volzke H. Study of health in Pomerania (SHIP). Concept, design and selected results. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2012;55(6–7):790-794.
19. Mayerle J, den Hoed CM, Schurmann C, et al. Identification of genetic loci associated with *Helicobacter pylori* serologic status. *JAMA*. 2013;309(18):1912-1920.
20. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219.
21. Mei K, Li Y, Wang S, et al. Cryo-EM structure of the exocyst complex. *Nat Struct Mol Biol*. 2018;25(2):139-146.
22. Nishida-Fukuda H. The exocyst: dynamic machine or static tethering complex? *Bioessays*. 2019;41(8):e1900056.
23. Arasaki K, Kimura H, Tagaya M, Roy CR. Legionella remodels the plasma membrane-derived vacuole by utilizing exocyst components as tethers. *J Cell Biol*. 2018;217(11):3863-3872.
24. Takeuchi H, Takada A, Kuboniwa M, Amano A. Intracellular periodontal pathogen exploits recycling pathway to exit from infected cells. *Cell Microbiol*. 2016;18(7):928-948.
25. Miao Y, Bist P, Wu J, et al. Collaboration between distinct Rab small GTPase trafficking circuits mediates bacterial clearance from the bladder epithelium. *Cell Host Microbe*. 2017;22(3):330-342.e334.
26. Kirchhoff P, Wagner CA, Gaetzschmann F, Radebold K, Geibel JP. Demonstration of a functional apical sodium hydrogen exchanger in isolated rat gastric glands. *Am J Physiol Gastrointest Liver Physiol*. 2003;285(6):G1242-G1248.
27. Kulaksiz H, Bektas H, Cetin Y. Expression and cell-specific and membrane-specific localization of NHE-3 in the human and Guinea pig upper gastrointestinal tract. *Cell Tissue Res*. 2001;303(3):337-343.
28. Vogel CFA, Haarmann-Stemmann T. The aryl hydrocarbon receptor repressor – more than a simple feedback inhibitor of AhR signaling: clues for its role in inflammation and cancer. *Curr Opin Toxicol*. 2017;2:109-119.
29. Pirzadeh M, Khalili N, Rezaei N. The interplay between aryl hydrocarbon receptor, H. pylori, tryptophan, and arginine in the pathogenesis of gastric cancer. *Int Rev Immunol*. 2022;41(3):299-312.
30. Moura-Alves P, Fae K, Houthuys E, et al. AhR sensing of bacterial pigments regulates antibacterial defence. *Nature*. 2014;512(7515):387-392.
31. Brandstatter O, Schanz O, Vorac J, et al. Balancing intestinal and systemic inflammation through cell type-specific expression of the aryl hydrocarbon receptor repressor. *Sci Rep*. 2016;6:26091.
32. Zhu R, Gao C, Wang L, et al. Involvement of aryl hydrocarbon receptor and aryl hydrocarbon receptor repressor in *Helicobacter pylori*-related gastric pathogenesis. *J Cancer*. 2018;9(15):2757-2764.
33. Thorstensen L, Qvist H, Heim S, et al. Evaluation of 1p losses in primary carcinomas, local recurrences and peripheral metastases from colorectal cancer patients. *Neoplasia*. 2000;2(6):514-522.
34. Matsuzaki M, Nagase S, Abe T, et al. Detailed deletion mapping on chromosome 1p32-p36 in human colorectal cancer: identification of three distinct regions of common allelic loss. *Int J Oncol*. 1998;13(6):1229-1233.
35. Ahola-Olli AV, Wurtz P, Havulinna AS, et al. Genome-wide association study identifies 27 loci influencing concentrations of circulating cytokines and growth factors. *Am J Hum Genet*. 2017;100(1):40-50.
36. Menten P, Wuyts A, Van Damme J. Macrophage inflammatory protein-1. *Cytokine Growth Factor Rev*. 2002;13(6):455-481.
37. Hill CH, Mecham R, Starcher B. Fibrillin-2 defects impair elastic fiber assembly in a homocysteinemic chick model. *J Nutr*. 2002;132(8):2143-2150.
38. Scepanovic P, Hodel F, Mondot S, et al. A comprehensive assessment of demographic, environmental, and host genetic associations with gut microbiome diversity in healthy individuals. *Microbiome*. 2019;7(1):130.
39. Chen CC, Liou JM, Lee YC, Hong TC, El-Omar EM, Wu MS. The interplay between *Helicobacter pylori* and gastrointestinal microbiota. *Gut Microbes*. 2021;13(1):1-22.
40. Yao X, Smolka AJ. Gastric parietal cell physiology and *Helicobacter pylori*-induced disease. *Gastroenterology*. 2019;156(8):2158-2173.
41. Thorell K, Bengtsson-Palme J, Liu OH, et al. In vivo analysis of the viable microbiota and *Helicobacter pylori* transcriptome in gastric infection and early stages of carcinogenesis. *Infect Immun*. 2017;85(10):e00031-17.
42. Kabisch R, Semper RP, Wustner S, Gerhard M, Mejias-Luque R. *Helicobacter pylori* gamma-Glutamyltranspeptidase induces Tolerogenic human dendritic cells by activation of glutamate receptors. *J Immunol*. 2016;196(10):4246-4252.
43. Oertli M, Noben M, Engler DB, et al. *Helicobacter pylori* gamma-glutamyl transpeptidase and vacuolating cytotoxin promote gastric persistence and immune tolerance. *Proc Natl Acad Sci U S A*. 2013;110(8):3047-3052.
44. Brown H, Cantrell S, Tang H, Epplein M, Garman KS. Racial differences in *Helicobacter pylori* prevalence in the US: a systematic review. *Gastro Hep Adv*. 2022;1(5):857-868.
45. Deltenre M, de Koster E. How come I've got it? (a review of *Helicobacter pylori* transmission). *Eur J Gastroenterol Hepatol*. 2000;12(5):479-482.
46. Kindermann A, Konstantopoulos N, Lehn N, Demmelmair H, Koletzko S. Evaluation of two commercial enzyme immunoassays, testing immunoglobulin G (IgG) and IgA responses, for diagnosis of *Helicobacter pylori* infection in children. *J Clin Microbiol*. 2001;39(10):3591-3596.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lima IS, da Silva TM, Weiss S, et al. Genome-wide association study of *Helicobacter pylori* serological status in Latin American children. *Helicobacter*. 2023;28:e13008. doi:10.1111/hel.13008