



**HAL**  
open science

## Human adaptation to arsenic in Bolivians living in the Andes

Jessica de Loma, Mário Vicente, Noemi Tirado, Franz Ascui, Marie Vahter, Jacques Gardon, Carina M Schlebusch, Karin Broberg

► **To cite this version:**

Jessica de Loma, Mário Vicente, Noemi Tirado, Franz Ascui, Marie Vahter, et al.. Human adaptation to arsenic in Bolivians living in the Andes. *Chemosphere*, 2022, 301, pp.134764. 10.1016/j.chemosphere.2022.134764 . hal-04543475

**HAL Id: hal-04543475**

**<https://hal.science/hal-04543475v1>**

Submitted on 12 Apr 2024

**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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



## Human adaptation to arsenic in Bolivians living in the Andes

Jessica De Loma<sup>a</sup>, Mário Vicente<sup>b</sup>, Noemi Tirado<sup>c</sup>, Franz Ascui<sup>d</sup>, Marie Vahter<sup>a</sup>, Jacques Gardon<sup>e</sup>, Carina M. Schlebusch<sup>b,f,g</sup>, Karin Broberg<sup>a,\*</sup>

<sup>a</sup> Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

<sup>b</sup> Human Evolution, Department of Organismal Biology, Uppsala University, Uppsala, Sweden

<sup>c</sup> Genetics Institute, Universidad Mayor de San Andrés, La Paz, Bolivia

<sup>d</sup> Programa de Salud Familiar Comunitaria e Intercultural, Ministerio de Salud Bolivia, La Paz, Bolivia

<sup>e</sup> Hydrosciences Montpellier, Université de Montpellier, Institut de Recherche pour le Développement, Centre National de la Recherche Scientifique, Montpellier, France

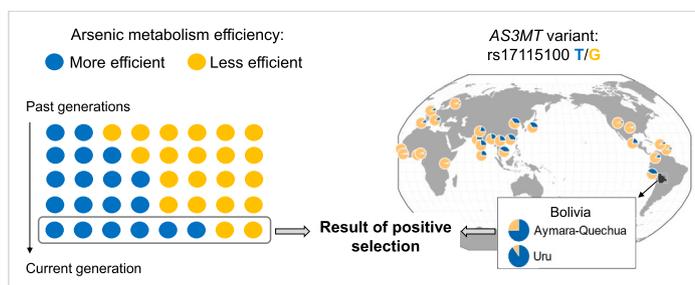
<sup>f</sup> Palaeo-Research Institute, University of Johannesburg, P.O. Box 524, Auckland Park, 2006, South Africa

<sup>g</sup> SciLifeLab Uppsala, Sweden

### HIGHLIGHTS

- Adaptation to arsenic has occurred in Andean indigenous populations.
- Bolivians had a high frequency of genetic variants linked to efficient arsenic metabolism.
- They had strong signals of positive selection near *AS3MT* (arsenic methylating enzyme).
- Candidate genetic variants linked to differential expression of *AS3MT* were identified.

### GRAPHICAL ABSTRACT



### ARTICLE INFO

Handling editor: A. Gies

#### Keywords:

Selection  
Toxic  
AS3MT  
Tolerance  
Drinking water  
Evolution

### ABSTRACT

Humans living in the Andes Mountains have been historically exposed to arsenic from natural sources, including drinking water. Enzymatic methylation of arsenic allows it to be excreted more efficiently by the human body. Adaptation to high-arsenic environments via enhanced methylation and excretion of arsenic was first reported in indigenous women in the Argentinean Andes, but whether adaptation to arsenic is a general phenomenon across native populations from the Andes Mountains remains unclear. Therefore, we evaluated whether adaptation to arsenic has occurred in the Bolivian Andes by studying indigenous groups who belong to the Aymara-Quechua and Uru ethnicities and have lived in the Bolivian Andes for generations. Our population genetics methods, including genome-wide selection scans based on linkage disequilibrium patterns and allele frequency differences, in combination with targeted and whole-genome sequencing and genotype–phenotype association analyses, detected signatures of positive selection near the gene encoding arsenite methyltransferase (*AS3MT*), the main

**Abbreviations:** AS3MT, arsenite methyltransferase; DMA, dimethylarsinic acid; hg19, human reference genome build version 37; HPLC, high-performance liquid chromatography; HPLC-HG-ICP-MS, high-performance liquid chromatography online with hydride generation and inductively coupled plasma-mass spectrometry; iAs, inorganic arsenic [As(III) plus As(V)]; iHS, integrated haplotype score; IQR, interquartile range; LD, linkage disequilibrium; LSBL, locus-specific branch length; MAF, minor allele frequency; MMA, monomethylarsonic acid; SAC, San Antonio de los Cobres; SNP, single-nucleotide polymorphism; U–As, urinary arsenic (sum of iAs, MMA, and DMA concentrations); XP-EHH, cross-population extended haplotype homozygosity.

\* Corresponding author. Institute of Environmental Medicine, Unit of Metals and Health, Karolinska Institutet, Box 210, 17177, Stockholm, Sweden.

E-mail address: [karin.broberg@ki.se](mailto:karin.broberg@ki.se) (K. Broberg).

<https://doi.org/10.1016/j.chemosphere.2022.134764>

Received 11 February 2022; Received in revised form 22 April 2022; Accepted 25 April 2022

Available online 28 April 2022

0045-6535/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

arsenic methylating enzyme. This was among the strongest selection signals (top 0.5% signals via locus-specific branch length and extended haplotype homozygosity tests) at a genome-wide level in the Bolivian study groups. We found a large haplotype block of 676 kb in the *AS3MT* region and identified candidate functional variants for further analysis. Moreover, our analyses revealed associations between *AS3MT* variants and the fraction of mono-methylated arsenic in urine and showed that the Bolivian study groups had the highest frequency of alleles associated with more efficient arsenic metabolism reported so far. Our data support the idea that arsenic exposure has been a driver for human adaptation to tolerate arsenic through more efficient arsenic detoxification in different Andean populations.

## 1. Introduction

Arsenic is a ubiquitous element in the volcanic bedrock of the Andes Mountains. This toxic element naturally leaches from the rocks into the sources of drinking water for inhabitants of these regions. Inorganic arsenic (iAs) is a potent toxicant classified as a group 1 carcinogen associated with multiple types of cancer (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012), as well as with non-cancerous outcomes such as cardiovascular diseases (Chen et al., 2013; Khan et al., 2020) and diabetes (Bräuner et al., 2014; Grau-Perez et al., 2017). Prenatal or early-life iAs exposure has been associated with increased child morbidity (Rahman et al., 2011; Smith et al., 2013) and mortality (Rahman et al., 2010; Shih et al., 2017; Smith et al., 2012).

In humans, iAs is mainly metabolized in the liver by a series of reduction and methylation steps. The methylated metabolites, monomethylarsonic acid (MMA) and dimethylarsinic acid (DMA), are eliminated in urine in combination with some remaining non-methylated iAs (Vahter, 2002). The relative concentrations of the arsenic metabolites in urine are often used as an index of arsenic methylation efficiency (Vahter, 2002). Most human populations described in the literature present on average 10–30% iAs, 10–20% MMA, and 60–80% DMA in urine (Vahter and Concha, 2001). This variability in how humans methylate arsenic influences the susceptibility to arsenic toxicity: individuals with higher %MMA in urine and lower %DMA are at higher risk of arsenic toxicity (Lindberg et al., 2008b; Pu et al., 2007; Steinmaus et al., 2010; Tseng et al., 2005).

The main enzyme that metabolizes arsenic is arsenite methyltransferase (*AS3MT*), which uses methyl groups from one-carbon metabolism (Lin et al., 2002; Pierce et al., 2012; Schlebusch et al., 2015). Variants in *AS3MT* have been shown to modify arsenic methylation (Antonelli et al., 2014; Balakrishnan et al., 2017; Delgado et al., 2021; Engström et al., 2011; Pierce et al., 2012). Apart from *AS3MT* variants, other factors known to influence arsenic methylation in adults include sex (Lindberg et al., 2008a), smoking (Hopenhayn-Rich et al., 1996), pregnancy (Gardner et al., 2011), and nutritional factors directly related to one-carbon metabolism, such as folate and cobalamin (Howe et al., 2014).

Previously, a unique and efficient arsenic metabolism (i.e., low % MMA and high %DMA in urine) was described in a group of indigenous women living in the northern Argentinean Andean town of San Antonio de los Cobres (SAC) and surrounding villages, which have elevated arsenic concentrations in drinking water (Vahter et al., 1995). A later study in women from the same area found that the efficient arsenic metabolism was associated with a high frequency of a “protective” haplotype, with signatures of positive selection near *AS3MT* (Schlebusch et al., 2013, 2015). This was the first description of genetic adaptation to iAs exposure in humans. Since then, other studies from Chile and Argentina (Apata et al., 2017; Eichstaedt et al., 2015) have identified study groups with high frequencies of presumed protective haplotypes of *AS3MT*. However, neither of these studies evaluated whether the *AS3MT* haplotypes or genetic variants were associated with arsenic metabolism phenotypes. Furthermore, the genetic variant(s) of the protective haplotype that drive efficient arsenic methylation have not been identified.

Our recent paper described that women of the Aymara-Quechua and

Uru ethnicities living in communities around Lake Poopó in the Bolivian highlands have been chronically exposed to arsenic (current median of 65 µg/L arsenic in urine) mainly through drinking water (De Loma et al., 2019). We also found that these Bolivian women had efficient arsenic methylation (De Loma et al., 2019), similar to the women from the Argentinean Andes (Vahter et al., 1995). This Bolivian study group, therefore, offers a good opportunity for us to assess if the selection for efficient arsenic metabolism is a general phenomenon across native populations from the Andes Mountains. This is the most comprehensive study so far regarding human adaptation to arsenic and the genetic variants driving this. The Bolivian study populations have not been evaluated in relation arsenic adaptation before and it should be noted that the Uru have been hardly described in the scientific literature.

## 2. Materials and methods

### 2.1. Bolivian study groups

The Aymara and Quechua are the major indigenous groups in the Bolivian highlands in the Andes, whereas many fewer Uru individuals live in this region. For this study, we included 187 individuals: 155 Aymara-Quechua and 27 Uru women, and 5 Uru men to increase the sample size of this study group (flowchart for the study in Supplementary Figure 1). Recruitment, exposure to arsenic, water consumption, and sampling of urine and blood from the women has been described elsewhere (De Loma et al., 2019); the men were recruited and sampled as described for the other study participants. Recruitment took place between 2015 and 2017 from ten villages around Lake Poopó, located in the southern part of the Bolivian Altiplano (3700 m above sea level). Most study participants were women, as the men were often away from home to work. We initially excluded first-degree relatives based on questionnaire information. The ethnic group of the participants was based on their birthplace, complemented with the residency of their parents and grandparents, as reported during the interviews. We recruited participants from all Uru communities around Lake Poopó, and from several of the Aymara and Quechua communities. The Aymara and Quechua ethnic groups are fairly similar from a genetic point of view (Batai and Williams, 2014; Lindo et al., 2018) and the two groups cohabit, especially around Lake Poopó. Therefore, study participants from these two ethnicities were considered as one study group. In contrast, the Uru communities around Lake Poopó, specifically of Uru-Murato descent, have been historically isolated from other communities in that region (de la Barra Saavedra et al., 2011), and very little is known about their genetic background (Sandoval et al., 2013).

This study was approved by the Comité Nacional de Bioética (Bolivia) and the Regional Ethic Committee of Karolinska Institutet (Sweden). Oral and written informed consent was obtained from all participants.

### 2.2. Argentinean study group

To compare the different study groups of indigenous populations in the Andes, we included a subset of women (those for whom we had enough DNA available) from the Argentinean Andes ( $n = 53$ , see Supplementary Figure 1) previously analyzed in the study by Schlebusch

et al. (2015). These women were recruited during 2008 in San Antonio de los Cobres (SAC) located at 3800 m above sea level in the northern Argentinean Andes. During interviews, the participants were asked whether they were originally from the area around SAC and where their parents and grandparents had lived. Directly asking about ethnicity was avoided, since it was a sensitive topic in this region at the time of recruitment. Based on later research in the same region, women identified themselves as belonging mainly to the Kolla community (Harari et al., 2015). Since the participants did not identify their ethnographic background at the time of recruitment, we refer to this study group based on their geographic location (i.e., SAC). Oral and written informed consent was obtained from all participants. This study was approved by the Comisión Provincial de Ética del Ministerio de Salud de Salta (Argentina) and the Regional Ethic Committee of Karolinska Institutet (Sweden).

### 2.3. Genome-wide genotyping

For the present study, genome-wide genotyping was performed on DNA from individuals from the Bolivian Andes. Venous blood samples were obtained from the arm with BD Vacutainer Eclipse blood collection needles (Becton Dickinson, USA) in EDTA tubes (Vacuette, Greiner Bio, Austria). DNA was extracted from whole blood samples with the E. Z.N. A. Blood DNA Mini kit (Omega Bio-tek, USA). Genome-wide genotyping was performed at the SNP&SEQ Technology Platform in Uppsala, Sweden on the Illumina Infinium Omni5Exome SNP array. This BeadChip array covers >4.3 million whole-genome variants down to 1% minor allele frequency (MAF), plus novel functional exonic variants. The data were aligned to the human reference genome build, version 37 (hg19). By using PLINK v1.90 (Chang et al., 2015), we filtered for a 5% genotype missingness threshold for single-nucleotide polymorphisms (SNPs), 15% genotype missingness threshold for individuals, indels, and duplicates. Furthermore, A/T and C/G SNPs were removed to avoid strand issues when merging to compare datasets. Only autosomal SNPs were included.

### 2.4. Population structure analyses

For population structure analyses, including principal components and clustering analyses, data from the Bolivian study groups ( $n = 187$ ) were merged with data from (i) South American populations reported by Barbieri et al. (2019) ( $n = 174$ ); (ii) populations of Native American ancestry from North, Central, and South America, as well as Iberian (IBS) and Yoruba (YRI) from the 1000 Genomes Project ( $n = 312$ ) from a previously published dataset assembled by Lazaridis et al. (2014); and (iii) the SAC population ( $n = 120$ ) from Schlebusch et al. (2015). Due to mainly African and European immigration over time in South America, we expected possible admixture with African (represented by YRI) and European (represented by IBS) population groups and they were therefore included in the principal component analyses (PCA) for estimation of admixture. We used a Hardy-Weinberg equilibrium filter ( $p$ -value < 0.001) to account for genotyping errors in the Bolivian (4763 SNPs excluded) and SAC (4017 SNPs excluded) datasets. First-degree relatives ( $n = 13$  from SAC, 7 Aymara-Quechua, and 8 Uru), as determined with the software KING v2.1.4 (Manichaikul et al., 2010), were excluded. In total, 236,127 autosomal SNPs from 765 individuals were analyzed. We further pruned SNPs in high linkage disequilibrium (LD,  $plink -indep-pairwise 200 25 0.4$ ), resulting in 120,384 autosomal SNPs.

Principal component analysis was done using the smartpca program included in EIGENSOFT (Patterson et al., 2006; Price et al., 2006). Admixture fractions were inferred using ADMIXTURE (Alexander et al., 2009) and visually inspected with PONG (Behr et al., 2016). Phylogenetic relationships were inferred using maximum-likelihood-based software TreeMix v.1.12 (Pickrell and Prichard, 2012). See the Supplementary Material file for further method description.

## 2.5. Selection scans

### 2.5.1. Merging with comparative datasets

We merged the data from the Bolivian study groups ( $n = 187$ ) with data of populations from the 1000 Genomes Project (Northern Europeans, CEU,  $n = 80$ ; Han Chinese, CHB,  $n = 80$ ; Colombians, CLM,  $n = 69$ ; Mexicans, MXL,  $n = 52$ ; Peruvians, PEL,  $n = 55$ ; Puerto Ricans, PUR,  $n = 76$ ; and Yoruba, YRI,  $n = 80$ ). We also merged them with genetic data from SAC ( $n = 120$ ), genotyped with Illumina HumanOmni 5 M BeadChip that covers 4.3 million variants, as described in Schlebusch et al. (2015) to aid during the haplotype phasing step. The phasing of haplotypes refers to assigning each allele to the maternal or paternal chromosome. To do that, relatives within the Andean study groups were kept to aid the statistical estimation of the haplotypes. We used a Hardy-Weinberg equilibrium filter ( $p$ -value < 0.001) for each subpopulation to identify SNPs with possible genotyping errors. However, to avoid removing potential selection candidates, which are expected to be out of equilibrium, we excluded only those SNPs that were out of Hardy-Weinberg equilibrium across the four comparative 1000 Genomes Project populations from South America, those out of equilibrium common to Aymara-Quechua and Uru, and those out of equilibrium common to Aymara-Quechua and SAC. In total, 1,451,199 autosomal SNPs from 799 individuals (including first- and second-degree relatives) were kept at this stage.

### 2.5.2. Haplotype phasing and ancestral state

Haplotypes were phased using fastPHASE (v1.4.0) with one random start for 25 iterations and 25 clusters. Once haplotypes were phased, to avoid introducing bias due to over-representation of family genotypes during selection scans, we excluded first- and second-degree relatives ( $n = 30$  from SAC, 12 Aymara-Quechua and 12 Uru) identified with KING, resulting in a total of 90 individuals from SAC, 143 Aymara-Quechua and 20 Uru. To compute the ancestral state of each allele, we used as outgroups the hg19-aligned versions of the reference genomes of the chimpanzee, gorilla, and orangutan as described in Schlebusch et al. (2015). For each SNP, an allele was defined as ancestral if it was present in three of the outgroup genomes. The dataset was then filtered based on this information, with only those SNPs that were the same in the three outgroups retained as their ancestral state was certain.

### 2.5.3. $iHS$ and $XP-EHH$ computation

The integrated haplotype score ( $iHS$ ; Voight et al., 2006) and cross-population extended haplotype homozygosity ( $XP-EHH$ ; Sabeti et al., 2007) are two statistics used to detect positive selection in the genome. These tests are both based on the fact that extended haplotype homozygosity arises when allele frequencies increase faster than the rate at which recombination can break them down (Sabeti et al., 2002).  $XP-EHH$  detects alleles that are near fixation in one population compared to another population. We compared the Bolivian study groups (Aymara-Quechua and Uru separately) to Peruvians from Lima. The same tests were also done comparing the Bolivian study groups to Han Chinese (data not shown). The  $iHS$  and  $XP-EHH$  computations were done using the R package *rehh* (Gautier and Vitalis, 2012), with a maximum distance between two SNPs of 200,000 bp to avoid spurious signals in genomic regions with low SNP density.

### 2.5.4. Locus-specific branch length computation

The locus-specific branch length (LSBL) test is based on fixation index ( $F_{ST}$ ) values that identify SNPs that differ in allele frequencies among three populations (Shriver et al., 2004). Using Weir and Cockerham's equation,  $F_{ST}$  distances measure the differentiation between populations based on the heterozygosity of each SNP position. Pairwise  $F_{ST}$  distances between Aymara-Quechua or Uru, Peruvians from Lima, and Han Chinese were calculated with PLINK.

### 2.5.5. Selection scan signals

Selection scan signals were identified by averaging the  $-\log_{10}$  (p-value) of iHS, as well as the XP-EHH and LSBL values, on a sliding window of ten SNPs to remove the effect of individual outliers (Nelson et al., 2013; Vicente et al., 2019). To assess if selection signals in *AS3MT* were relevant at a whole-genome scale for the Aymara-Quechua and Uru study groups independently, we evaluated the top 0.5% selected SNPs obtained with each positive selection test and study group, and we annotated them to genes with the Ensembl Variant Effect Predictor (VEP) toolset (McLaren et al., 2016). We searched for genes that were within 10 kb on either side of the input SNP to investigate potential distant regulatory effects of SNPs, and we checked if *AS3MT* was in the top list of selected genes.

### 2.6. Protective haplotype associated with efficient arsenic metabolism

Higher %MMA in urine, reflecting less efficient methylation of absorbed iAs, is associated with an increased risk of arsenic-related health effects (Lindberg et al., 2008b; Steinmaus et al., 2006). Therefore, lower %MMA and higher %DMA in urine indicate efficient methylation of iAs to DMA. A haplotype of 52 SNPs surrounding *AS3MT* that were significantly associated with low %MMA in urine (“protective haplotype”) was previously identified in SAC (Schlebusch et al., 2015). In the current study, some of the 52 SNPs were filtered out of the dataset because we used different filtering settings in the present study compared with Schlebusch et al. (2015), and due to merging datasets genotyped with different arrays. Out of the 52 SNPs, 35 were included in the Bolivian SNP array dataset (located in the interval chr10:104315215–10:104852121, hg19; 537 kb length). These 35 SNPs were distributed along the whole protective haplotype. Based on these 35 SNPs, we calculated the frequency of haplotypes matching the putative protective haplotype in the Bolivian Aymara-Quechua ( $n = 143$ ) and Uru ( $n = 20$ ) study groups, and in the SAC study group ( $n = 90$ ), after excluding first- and second-degree relatives (see section 2.5). Finally, the haplotypes were plotted in R for the three Andean study groups separately using the protective alleles identified in SAC as a reference, i.e., alleles associated with lower %MMA in urine.

### 2.7. Genotype–phenotype associations

Concentrations of iAs metabolites in urine in the Bolivian study groups were measured by high-performance liquid chromatography online with hydride generation and inductively coupled plasma-mass spectrometry (HPLC-HG-ICP-MS) as previously described (De Loma et al., 2019). To estimate exposure to iAs, we used the sum of urinary arsenic metabolite concentrations. Urinary arsenic was adjusted to the average osmolality to correct for variations in urine dilution. The efficiency of arsenic metabolism was assessed by the fractions of iAs metabolites in urine (i.e., phenotypes in this context), presented as a percentage of the total sum of metabolites (%iAs, %MMA, and %DMA). The concentration of iAs was calculated as the sum of As(III) and As(V) concentrations.

Associations between the 35 SNPs constituting the protective haplotype and arsenic metabolism efficiency phenotypes were evaluated using multivariable linear regression models for the Bolivian study groups combined ( $n = 163$ , out of which 143 were Aymara-Quechua and 20 were Uru; relatives excluded based on KING). These models were adjusted for age, sex, and urinary arsenic concentration to facilitate comparison with previous studies (Balakrishnan et al., 2017; Pierce et al., 2012). We also adjusted the models for ethnicity, since we combined both Bolivian study groups to increase the sample size. Furthermore, genome-wide associations were performed with the R package GenABEL (version 1.8), as described in the Supplementary Material file.

### 2.8. PacBio targeted sequencing of a region within the protective haplotype

To identify which SNPs within the protective haplotype are potentially driving the adaptation to arsenic in the Andean populations from Bolivia and Argentina, we sequenced  $\pm 2.5$  kb from the top SNP (rs486955 located at chr10:104546284, hg19) associated with lower % MMA in SAC (Schlebusch et al., 2015). A subset of 96 individuals from the Argentinean and Bolivian study groups was included in the targeted sequencing: 53 from SAC (samples for which we had enough DNA), 24 Aymara-Quechua (12 with the lowest %MMA and 12 with the highest), and 19 Uru (samples for which we had enough DNA). We first performed long-range PCR (4.8-kb amplicon) with individually barcoded fusion HPLC primers as described in the Supplementary Material file. The amplicon for each individual had two unique barcodes that allowed all samples to be pooled in an equimolar manner. Using the Pacific Biosciences Sequel I system powered by Single-Molecule, Real-Time (SMRT) technology, long reads were obtained at the National Genomic Infrastructure/Uppsala Genome Center, Sweden. Data analysis of these obtained HiFi reads is described in the Supplementary Material file.

### 2.9. Whole-genome sequencing

To further explore whether there were significant genetic variants not included in the pre-defined SNP array and not covered by the targeted sequencing, we performed whole-genome sequencing. Since adaptation to arsenic was first identified in SAC, we selected 12 women from the SAC group (out of which 11 were also selected for PacBio sequencing) for whole-genome sequencing. The 12 women were selected based on their protective haplotypes, as defined with the SNP array data: 4 women who were homozygous for the protective haplotype, 4 with one protective haplotype and one mixed haplotype, and 4 with one non-protective haplotype and one mixed haplotype or both mixed. A mixed haplotype here refers to the protective haplotype formed by alleles associated with lower %MMA and higher %MMA (see Section 2.6). Sequencing libraries, cluster generation, and 150 cycle paired-end sequencing with S4 flowcell and NovaSeq 6000 system (Illumina, USA) were performed at the National Genomic Infrastructure/Uppsala Genome Center. Further description of the processing of the results is found in the Supplementary Material file.

## 3. Results

### 3.1. Selection signals near *AS3MT* in Bolivian populations

General characteristics of the Bolivian study groups are described in Table 1. The Aymara-Quechua and Uru study groups are currently exposed to moderate levels of arsenic (median urinary arsenic concentrations of 72 and 61  $\mu\text{g/L}$ , respectively). We genotyped 187 individuals of Aymara-Quechua and Uru descent for approximately 4.5 million genome-wide SNPs from a pre-defined SNP array. Based on principal component analysis, the Bolivian study groups clustered with other Andean populations previously evaluated by Lazaridis et al. (2014) and Schlebusch et al. (2015) (Fig. 1A). After the African and European admixture are taken into consideration in principal component (PC) 1 and PC2, respectively (Supplementary Figure 3), PC3 separates Central American from South American groups while PC4 may indicate a differentiation of Andean groups from Amazonian groups with the Uru and Karitiana/Surui defining the extremes (Fig. 1A). Further, the TreeMix results suggest that Aymara-Quechua and Uru belong to a clade representing coastal and Andean populations, distinct from groups of the Amazonia. These results corroborate our PCA results (Supplementary Figure 4). Little European ancestry was identified in the Bolivian study groups, and their African ancestry was negligible (Fig. 1B). Other Aymara and Quechua populations previously assembled in Lazaridis et al. (2014) had similar ancestry patterns to the Aymara-Quechua

**Table 1**  
Characteristics of the Bolivian individuals included in the study ( $n = 163$ ).

	Median (IQR) <sup>a</sup>		p-value <sup>b</sup>
	Aymara-Quechua	Uru	
Number of participants	143	20	NA
Sex (female)	100%	75%	NA
Age (years)	36 (29–47)	34 (27–47)	0.266
Body weight (kg)	60 (53–68)	69 (58–77)	0.018
iAs (%)	13 (8.6–17)	13 (8.8–18)	0.718
MMA (%)	7.8 (6.3–9.7)	6.6 (5.7–8.1)	0.046
DMA (%)	79 (74–83)	79 (74–85)	0.677
U–As (µg/L) <sup>c</sup>	72 (48–132)	61 (43–88)	0.090
Tobacco smoking (yes)	3.5%	5.5% <sup>d</sup>	NA

Abbreviations: DMA, dimethylarsinic acid; iAs, inorganic arsenic; IQR, interquartile range; MMA, monomethylarsonic acid; U–As, urinary arsenic (sum of iAs, MMA, and DMA concentrations).

<sup>a</sup> Interquartile range (IQR) presented as 25th and 75th percentile.

<sup>b</sup> p-value for Wilcoxon rank-sum (Mann-Whitney) test.

<sup>c</sup> Sum of urinary iAs metabolite concentrations adjusted for average urinary osmolality (727 mOsm/kg) of the total study group as described in De Loma et al. (2019).

<sup>d</sup> Two participants did not report their smoking habits. Out of those who did report their smoking habits ( $n = 18$ ), only one was a smoker.

participants from this study (Fig. 1B).

To study whether positive selection influenced the genomic traits related to arsenic metabolism in the two Bolivian study groups, we used three selection scan methods: two based on haplotype length (iHS and XP-EHH) and one based on pairwise population differentiation at each given locus (LSBL). To minimize the detection of spurious single-SNP outliers, we calculated the average p-values for iHS, and the average XP-EHH and LSBL values across the genome by sliding-window averaging. For each selection scan method and study group, we focused on the top 0.5% candidate SNPs. Subsequently, we identified genes within 10 kb on either side of each top SNP. We then evaluated whether *AS3MT* was included in the top selected genes. Indeed, *AS3MT* was among the top 0.5% genome-wide signals for the LSBL scans in both Bolivian study groups, and in the XP-EHH scan of the Aymara-Quechua group. No other known genes related to arsenic metabolism was found among the top 0.5% genome-wide signals. The LSBL scans for the Bolivian study groups, with Peruvians from Lima and Han Chinese as comparative populations, showed strong peaks near *AS3MT* (Fig. 2A). The results for the XP-EHH tests, when using Peruvians from Lima as a comparative group, showed a clear elevation of XP-EHH values above the genome-wide average around *AS3MT*, particularly for the Aymara-Quechua group (Fig. 2B). The iHS results did not reveal clear signals near *AS3MT* for either Bolivian study group (Fig. 2C).

### 3.2. SNPs associated with arsenic metabolism efficiency in Bolivia

The frequency of haplotypes that exactly matched the protective haplotype (formed by alleles associated with low %MMA in urine) was similar between SAC (63%) and Aymara-Quechua (66%), while Uru had a higher frequency (75%) (Supplementary Figure 5). In Schlebusch et al. (2015), the frequency of exactly matched haplotypes for SAC was 58%, but it should be noted that the Schlebusch et al. study included a longer haplotype (see section 2.6).

We then investigated whether SNPs constituting the protective haplotype also associated with efficient arsenic metabolism in the Bolivian study groups. We performed linear regression analysis between genetic variants constituting the protective haplotype and each urinary arsenic metabolite fraction (%iAs, %MMA, and %DMA, as continuous variables) adjusting for age, sex, urinary arsenic concentration, and ethnicity (Supplementary Tables 1–3, the minor (effect) allele and MAF are described in Supplementary Table 4). Out of the 35 SNPs constituting the protective haplotype, ten SNPs were significantly associated (p-value < 0.05) with %MMA in urine (Table 2, Fig. 3). The SNP most

strongly associated with %MMA was rs17115100 (chr10:104591393, hg19), where the mean %MMA was 7.7% for TT carriers, 8.9% for TG, and 8.9% for GG (B coefficient = 0.89, p-value = 0.02). No SNPs within the protective haplotype were significantly associated with %iAs or %DMA, but the genotypes associated with significantly lower %MMA were in general associated with lower %iAs and with higher %DMA (Supplementary Tables 1 and 3).

The ten SNPs associated with %MMA were spread within the protective haplotype. The top SNP rs17115100 (38 kb upstream of *AS3MT*) is located 45 kb downstream from the top SNP identified in SAC (rs486955) in Schlebusch et al. (2015) (Supplementary Figure 6). Minor allele frequencies (MAF; minor allele designated as the least frequent allele in the study groups) of the SNPs constituting the protective haplotype are reported in Supplementary Table 4. The T-allele of rs17115100, associated with lower %MMA in urine, was present at a much higher frequency in the Bolivian study groups (75% for Aymara-Quechua and 90% for Uru) compared to other populations in the 1000 Genomes Project (Peruvians: 41%, Han Chinese: 35%, Yoruba: 4%; Supplementary Figure 7). In addition, the C-allele of rs486955, the top SNP associated with lower %MMA in SAC, showed a high frequency in the Bolivian study groups (Aymara-Quechua: 73%, Uru: 82%, Peruvians: 41%, Han Chinese: 47%, Yoruba: 26%; Supplementary Figure 7). Based on the gene expression GTEx portal database, rs17115100 is associated with differential expression of *AS3MT* in several tissues, including pancreas, thyroid, skin, pituitary, and cerebellum, but not in liver, which is the main site of iAs methylation.

To explore if other genomic regions outside of the protective haplotype were also related to the efficiency of arsenic metabolism GWAS were performed for each arsenic metabolite fraction (%iAs, %MMA, and %DMA). No significant peaks in the Manhattan plots were found for any arsenic metabolite fraction in the Bolivians in unadjusted or adjusted analyses or when accounting for population structure and relatedness (data not shown).

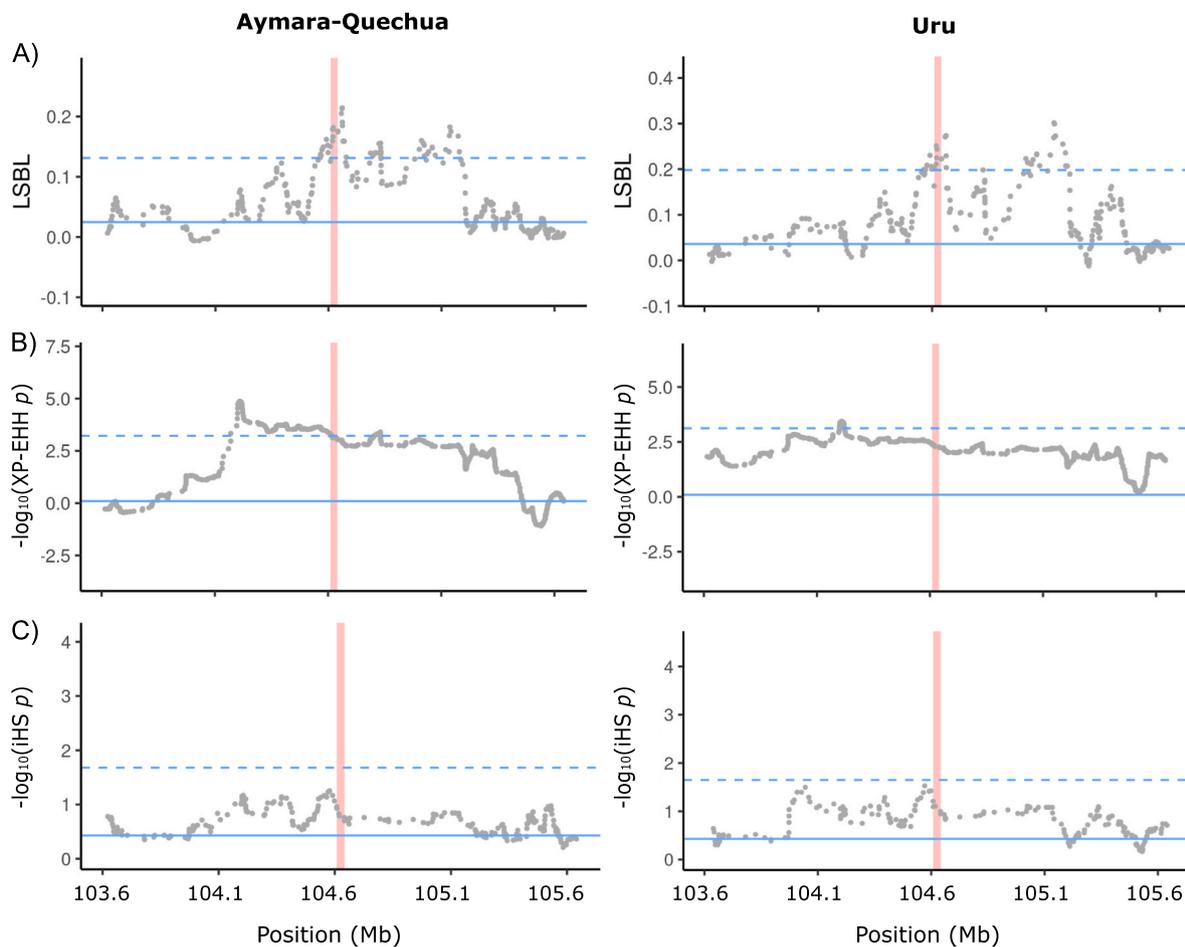
### 3.3. Sequencing to identify all SNPs in the Andean populations

We next evaluated whether there were SNPs that were not included in the pre-defined SNP arrays potentially driving the low %MMA in the Bolivian and Argentinean study groups. We first sequenced a 4.8-kb region upstream of *AS3MT* around the top SNP (rs486955) previously found to be associated with urinary %MMA in SAC (Schlebusch et al., 2015) for 96 individuals from the Bolivian and Argentinean Andes study groups. Characteristics of the study participants from SAC are presented in Supplementary Table 5. The variant calling identified 12 SNPs (1014 reads per sample on average; Supplementary Figure 8 of MAFs calculated with the ancestral allele as reference). All 12 SNPs have been previously described (NCBI dbSNP at <https://ncbi.nlm.nih.gov/snp>). Using these 12 bi-allelic SNPs, a total of five haplotypes were found in the 96 individuals, with two being shared across the three Andean study groups (Supplementary Table 6).

Next, we performed whole-genome sequencing for 12 individuals from SAC. We focused on a 1.5-Mb region around the protective haplotype, in which 2985 polymorphisms were identified. Pairwise correlations ( $r^2$ ) to estimate LD were calculated between the 2985 polymorphisms and showed a larger than previously reported LD block between positions chr10:104544753 and chr10:105221154 (676 kb, Supplementary Figure 9). In this block, 623 polymorphisms had an  $r^2 > 0.8$  with rs486955, and thus were in high LD with the top SNP from the genotype–phenotype associations from SAC. All of these 623 polymorphisms were present in other populations from the 1000 Genomes Project; therefore, the Andean study groups do not have unique polymorphisms in this genomic region.

From all the genetic variants within the 1.5-Mb region, we excluded polymorphisms that i) were present in populations of African descent (YRI), where no protective haplotype has been found (Schlebusch et al., 2015); ii) had a MAF below 0.4 in Peruvians from Lima, to select alleles





**Fig. 2.** Selection scan plots zoomed in on *AS3MT* (marked with red vertical lines). Results, genome-wide averages (blue solid lines), and top 0.5% SNPs (blue dashed lines) were based on averaged values using ten-SNP windows across the genome. A) Well-defined peaks near *AS3MT* in both Bolivian study groups were identified using the locus-specific branch length (LSBL) method with PEL (Peruvians) and CHB (Han Chinese) populations as comparative groups. B) SNPs near *AS3MT* were among the top 0.5% selected for Aymara-Quechua when using the cross population extended haplotype homozygosity (XP-EHH) test compared to PEL. C) There were no clear signals based on the integrated haplotype scores (iHS). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

**Table 2**

Top SNPs within the protective haplotype associated with %MMA in urine of the Bolivian study group ( $n = 163$ ) based on linear regression models adjusted for age, sex, urinary arsenic concentration, and ethnicity.

SNP ID	Chr	Position <sup>a</sup>	Alleles <sup>b</sup>	MAF AQ	MAF U	Gene	Location	B (95% CI)	p-value
rs17115100	10	104591393	G/T	0.25	0.10	<i>CYP17A1</i>	intron	0.89 (0.13, 1.65)	0.021
rs7085104	10	104628873	A/G	0.22	0.10	<i>AS3MT</i>	intron	0.88 (0.10, 1.65)	0.027
rs17884001	10	104661245	C/T	0.25	0.10	<i>AS3MT</i>	3' UTR	0.80 (0.04, 1.56)	0.039
rs12221064	10	104677126	C/T	0.25	0.10	<i>CNNM2</i>	intron	0.80 (0.04, 1.56)	0.039
rs17115213	10	104681143	A/G	0.25	0.10	<i>CNNM2</i>	intron	0.80 (0.04, 1.56)	0.039
rs78821730	10	104684544	G/A	0.25	0.10	<i>CNNM2</i>	intron	0.80 (0.04, 1.56)	0.039
rs11191535	10	104815876	G/A	0.25	0.13	<i>CNNM2</i>	intron	0.78 (0.02, 1.54)	0.044
rs943037	10	104835919	C/T	0.25	0.13	<i>CNNM2</i>	coding <sup>c</sup>	0.78 (0.02, 1.54)	0.044
rs12217501	10	104851889	T/C	0.25	0.13	<i>NT5C2</i>	intron	0.78 (0.02, 1.54)	0.044
rs10786719	10	104637992	A/G	0.22	0.10	<i>AS3MT</i>	intron	0.78 (0.01, 1.56)	0.049

Abbreviations: AQ, Aymara-Quechua; B, coefficient estimate; Chr, chromosome; CI, 95% confidence interval; MAF, minor allele frequency (in the study group); MMA, monomethylarsonic acid; SNP, single-nucleotide polymorphism; U, Uru.

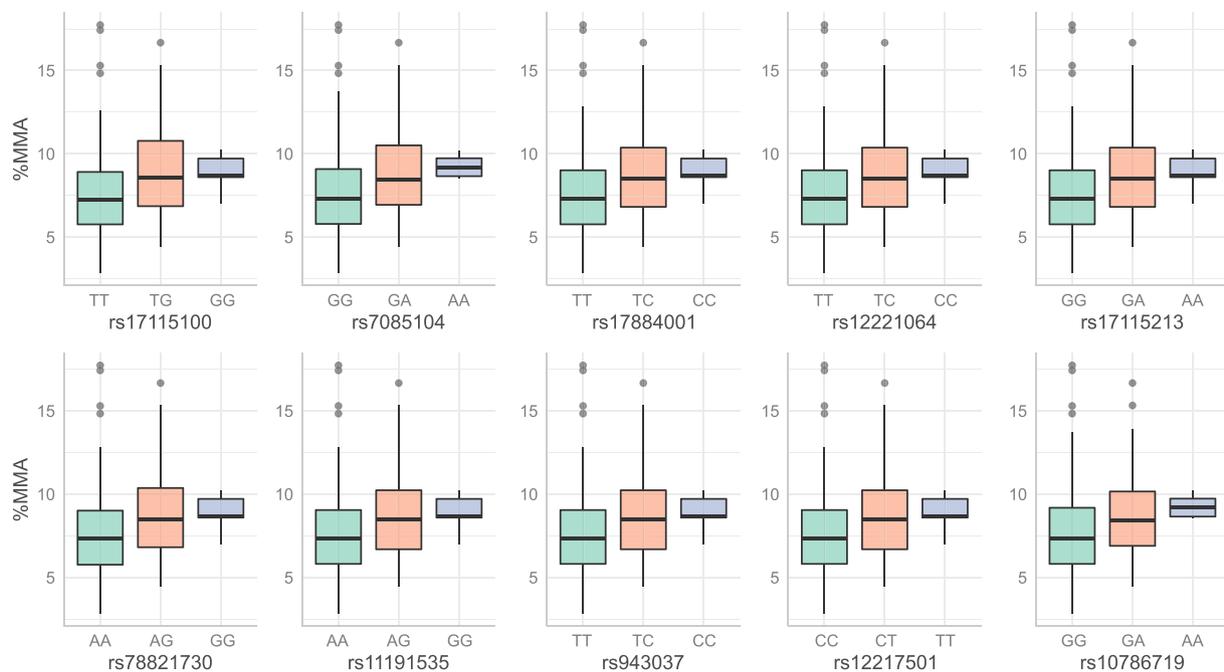
<sup>a</sup> Based on hg19.

<sup>b</sup> First allele stated is the minor allele in the Bolivian study group used to calculate the MAF and associated with higher %MMA.

<sup>c</sup> Synonymous variant.

database, two of these SNPs (rs56060736 and rs72846190) are associated with differential expression of *AS3MT* in several tissues including esophagus, lung, hypothalamus, and whole blood. Furthermore, the open-source database JASPAR CORE collection included in Genome Browser, a track representing predicted binding sites for transcription

factors, indicated that all 7 polymorphisms were located at predicted transcription factor binding sites.



**Fig. 3.** Boxplots of the SNPs within the protective haplotype significantly associated with %MMA in urine ( $p$ -value < 0.05) in the Bolivian study groups (Aymara-Quechua and Uru combined).

#### 4. Discussion

We here present the most comprehensive study so far regarding human adaptation to arsenic and the genetic variants driving this. For this, we combined molecular epidemiology (genotype-phenotype analyses) with population genetics (population structure and selection scans) to understand gene-environment interactions in several native populations in the Andes. We found that native populations in the Bolivian Andes, previously described as having an efficient arsenic metabolism phenotype, had signs of positive selection near the gene *AS3MT*. In fact, the selection signals near *AS3MT* were among the strongest at a genome-wide level in the Bolivian study groups compared with groups from Peru and China. Moreover, we identified the protective haplotype, previously defined based on SNPs associated with low urinary %MMA in SAC (Schlebusch et al., 2015), in the Bolivians, and we associated SNPs in this haplotype with %MMA in urine. By using several genotyping techniques, including whole genome sequencing, we were able to identify candidates of functional genetic variants that may be driving this more efficient metabolism of arsenic. We can, based on our results, draw general conclusions regarding adaptation to arsenic, namely selection for efficient arsenic metabolism is a general phenomenon across native populations from the Andes Mountains. This provides new knowledge regarding how humans have evolved to handle harsh toxicants in the environments, as well as new insights for biomarkers of susceptibility of this element.

Since the first description of arsenic adaptation in humans (Schlebusch et al., 2015), selection for *AS3MT* has been identified in study groups from Argentina, Chile, Peru, and Bolivia (Apata et al., 2017; Eichstaedt et al., 2015; Jacovas et al., 2018), but compared to our study, these studies had less dense genetic data, fewer selection scan methods, and no characterization of participants' arsenic exposure or metabolic efficiency. Apata et al. (2017) found that the frequency of four *AS3MT* alleles previously associated with lower %MMA was higher in a population historically exposed to high concentrations of arsenic through drinking water in the Atacama Desert of northern Chile than in populations exposed to lower arsenic concentrations. The authors suggested that this difference in haplotype frequency resulted from adaptation to arsenic exposure; however, natural selection was never tested. A study

exploring signs of positive selection in relation to living at high altitude in indigenous populations from South America, including Aymara individuals from Bolivia and Quechua individuals from Peru, identified *AS3MT* among the top selected genes (Jacovas et al., 2018). The study by Eichstaedt et al. (2015) included individuals with indigenous ancestry from northwestern Argentina, including 16 individuals from SAC, supporting the previous findings by Schlebusch et al. (2015). Interestingly, the study by Eichstaedt et al. (2015) did not reveal any selective signatures when using haplotype-based tests (iHS and XP-EHH), but it did when using allele frequency-based tests (PBS, or population branch statistic, which is comparable to LSBL in our study). The fact that more robust findings are obtained with the test based on allelic frequencies (i.e. LSBL) indicates that the selection might have acted on standing genetic variation already present before the populations were exposed to elevated arsenic concentrations, rather than resulting from new mutations (which would preferably be detected by iHS and XP-EHH). This is in line with our previous findings that the protective haplotype was also present in East Asian populations, although to a much lesser extent than in the Andean populations (Schlebusch et al., 2015), and it is further supported in this study by the lack of unique SNPs near *AS3MT* in the Andean populations when using next-generation sequencing.

Although we were able to measure current arsenic exposure, we do not know which arsenic concentrations in drinking water are needed for adaptation, and it should be noted that the historical levels of arsenic exposure are unknown. Nevertheless, the current levels of arsenic are moderate to high, and there are water sources that presumably have been used for drinking water with much higher arsenic levels. For example, in the river of SAC, we found concentrations of arsenic around 1000  $\mu\text{g/L}$  (Concha et al., 2010), while hot springs in Argentina (Vahter et al., 1995) and Bolivia (Morteani et al., 2014) have been found to contain several milligrams of arsenic per liter of water.

The selection for efficient arsenic metabolism may have occurred by different mechanisms, such as by increased early-life morbidity (Rahman et al., 2011; Smith et al., 2013) and mortality (Rahman et al., 2010; Shih et al., 2017; Smith et al., 2012); or via adverse effects of arsenic later in life, such as hepatotoxicity, cardiovascular disease, and impaired lung function (Frediani et al., 2018; Kuo et al., 2017; Sobel et al., 2021;

Zhao et al., 2021), which may result in reduced reproduction. As arsenic causes multiple adverse health effects in both children and adults, individuals who carry the more efficient arsenic-methylating haplotype and thus can metabolize arsenic faster and with lower levels of reactive metabolites may have a selective advantage in high-arsenic environments.

In our previous work characterizing the efficiency of arsenic metabolism in these Bolivian women, we found that ethnicity predicted the arsenic metabolism phenotype: Uru women had a more efficient arsenic methylation than Aymara-Quechua women, as defined by low %MMA in urine (De Loma et al., 2019). The fact that the Uru study group had an even higher frequency of the protective haplotype than the Aymara-Quechua group could explain why women with Uru ancestry appeared more efficient at methylating arsenic. The settlement of human populations in this region dates back approximately 10,000 years, as evidenced by the results of carbon 14 dating of charcoal found on campsites of mobile groups of hunters (Núñez et al., 2002). In the oral tradition, Uru communities claim to have been the first to settle down in the Andes, and hypothetically may have been in contact to higher arsenic levels for a longer period, resulting in higher frequency of the protective haplotype. Urus have also been more isolated than Aymara and Quechua (de la Barra Saavedra et al., 2011) which could result in longer haplotypes. We therefore evaluated the level of homozygosity, however, we found only a low number of long homozygous regions (>8 Mb) in the Andean groups (including Uru), implying limited recent endogamy. Nevertheless, it is difficult to determine with absolute certainty what drives the changes in haplotype frequencies. We did not find significant GWAS signals for the arsenic metabolite fractions in the Bolivian study groups, which likely is due to low genetic diversity of the *AS3MT* locus in the Bolivians.

The protective haplotype, formed of SNPs associated with lower % MMA in urine and in strong LD, covers a 500-kb region when using the SNP array data. This LD block includes several genes such as *WBP1L*, *CYP17A1*, *AS3MT*, *CNNM2*, and *NT5C2*. A long haplotype block in the same region around *AS3MT* (374 kb) has been associated with efficient arsenic methylation in a study group from Mexico (Gomez-Rubio et al., 2010). By using whole-genome sequencing, we identified a 676-kb LD block in the Argentinean study group; this block is larger than that identified using pre-defined SNP arrays. We highlighted a list of potential functional genetic variants in the Andean population from Argentina. All seven variants identified are noncoding, but they may affect the expression of *AS3MT* and surrounding genes, as one earlier study have shown for *AS3MT* haplotypes but for which the causal variants not yet have been identified (Engström et al., 2013). The same is true for the leading SNP (rs17115100) in the genotype-phenotype association analyses in the Bolivian study group. Hypothetically, an allele resulting in higher expression of *AS3MT* in the liver could lead to higher protein levels of *AS3MT* and thus, higher arsenic methylation capacity. The candidate functional variants have not been linked to arsenic metabolism in prior studies, and further molecular analysis for example of expression of *AS3MT* in liver tissue in situ needs to be performed to clarify which of these genetic variants are causal for more efficient arsenic methylation, and further, how these SNPs influence the expression of other genes in the haplotype and phenotypes linked to these genes.

Finally, although these indigenous individuals in the Andes generally have efficient arsenic methylation partly due to their genetic makeup, this does not exempt them from being affected by arsenic-related toxicity. Although no signs of hyperkeratosis or changes in skin pigmentation have been reported for the women from SAC (Concha et al., 1998; Vahter et al., 1995), associations between arsenic exposure and several effect biomarkers related to genotoxicity and oxidative stress have been found for those individuals with lower arsenic methylation efficiency (Ameer et al., 2016; Li et al., 2012). Furthermore, arsenic exposure was associated with changes in urinary protein levels in the Bolivian study group (De Loma et al., 2020). For this reason,

further efforts to lower arsenic exposure in these areas are required, such as finding alternative water wells with very low or no arsenic concentrations.

## 5. Conclusions

Overall, our study supports the idea that indigenous populations from the Andes have adapted to arsenic via efficient arsenic methylation produced by positive selection of standing genetic variation. Additional studies are required to confirm which SNPs are the causal genetic variant(s) driving the markedly efficient arsenic methylation of the members of these Andean populations, and to further characterize to what extent other genetic and environmental factors contribute.

## Funding

This work was supported by HydroSciences laboratory (IRD-University of Montpellier, France), the Eric Philip Sörensens Stiftelse (Sweden), the Swedish Research Council (2016–05799 and 2017–01239), Karolinska Institutet (Sweden), the Swedish International Development Cooperation Agency, and the Genetics Institute at Universidad Mayor de San Andrés (Bolivia). In addition, CS was funded by the European Research Council (ERC StG AfricanNeo, grant no. 759933). The SNP&SEQ Technology Platform and the Uppsala Genome Center are supported by the Swedish Research Council for Infrastructures and Science for Life Laboratory (Sweden). The SNP&SEQ Technology Platform is also supported by the Knut and Alice Wallenberg Foundation.

## Author contributions

Jessica De Loma: Data curation, Formal analysis, Investigation, Project administration, Visualization, Roles/Writing – original draft, Writing – review & editing. Mário Vicente: Data curation, Formal analysis, Visualization, Roles/Writing – original draft, Writing – review & editing. Noemi Tirado: Investigation, Resources, Roles/Writing – original draft, Writing – review & editing. Franz Ascui: Investigation, Roles/Writing – original draft, Writing – review & editing. Marie Vahter: Investigation, Supervision, Roles/Writing – original draft, Writing – review & editing. Jacques Gardon: Investigation, Roles/Writing – original draft, Writing – review & editing. Carina M. Schlebusch: Funding acquisition, Methodology, Resources, Software, Supervision, Roles/Writing – original draft, Writing – review & editing. Karin Broberg: Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Roles/Writing – original draft.

## 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 warmly thank all the participants who took part in this study. We would also like to thank the Servicio Departamental de Salud de Oruro (SEDES), the program of Salud Familiar Comunitaria e Intercultural (SAFCI), as well as the fieldwork teams and workers at local health centers in Argentina and Bolivia for assisting during recruitment. We are thankful to Gabriela Concha for her contribution during the recruitment in Argentina, to Michael Levi for measuring the arsenic metabolites, to Chiara Barbieri for sharing her data prior to publication, to Yanjun Zan and Örjan Carlborg for their insights regarding genotype-phenotype associations, and to Helena Korres de Paula for her help with the DNA extraction. Support by Vincent van Hoef as part of the National Bioinformatics Infrastructure Sweden (NBIS) is gratefully acknowledged. The

genotyping and sequencing were performed by the SNP&SEQ Technology Platform in Uppsala. We acknowledge the support of the National Genomic Infrastructure/Uppsala Genome Center for helping with massive parallel sequencing and computational infrastructure. The computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multi-disciplinary Center for Advanced Computational Sciences (UPPMAX).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chemosphere.2022.134764>.

## References

- Ameer, S.S., Xu, Y., Engström, K., Li, H., Tallving, P., Nermell, B., Boemo, A., Parada, L. A., et al., 2016. Exposure to inorganic arsenic is associated with increased mitochondrial DNA copy number and longer telomere length in peripheral blood. *Front. Cell Dev. Biol.* 4, 87. <https://doi.org/10.3389/fcell.2016.00087>.
- Antonelli, R., Shao, K., Thomas, D.J., Sams, R. 2nd, Cowden, J., 2014. *AS3MT*, *GSTO*, and *PNP* polymorphisms: impact on arsenic methylation and implications for disease susceptibility. *Environ. Res.* 132, 156–167. <https://doi.org/10.1016/j.envres.2014.03.012>.
- Apata, M., Arriaza, B., Llop, E., Moraga, M., 2017. Human adaptation to arsenic in andean populations of the Atacama Desert. *Am. J. Phys. Anthropol.* 163, 192–199. <https://doi.org/10.1002/ajpa.23193>.
- Balakrishnan, P., Vaidya, D., Franceschini, N., Voruganti, V.S., Gribble, M.O., Haack, K., Laston, S., Umans, J.G., et al., 2017. Association of cardiometabolic genes with arsenic metabolism biomarkers in American Indian communities: the Strong Heart Family Study (SHFS). *Environ. Health Perspect.* 125, 15–22. <https://doi.org/10.1289/EHP251>.
- Barbieri, C., Barquera, R., Arias, L., Sandoval, J.R., Acosta, O., Zurita, C., Aguilar-Campos, A., Tito-Álvarez, A.M., et al., 2019. The current genomic landscape of western South America: Andes, Amazonia and Pacific Coast. *Mol. Biol. Evol.* 36, 2698–2713. <https://doi.org/10.1093/molbev/msz174>.
- Batai, K., Williams, S.R., 2014. Mitochondrial variation among the Aymara and the signatures of population expansion in the central Andes. *Am. J. Hum. Biol. : Off. J. Human. Biol. Council* 26, 321–330. <https://doi.org/10.1002/ajhb.22507>.
- Bräuner, E.V., Nordsborg, R.B., Andersen, Z.J., Tjønneland, A., Loft, S., Raaschou-Nielsen, O., 2014. Long-term exposure to low-level arsenic in drinking water and diabetes incidence: a prospective study of the diet, cancer and health cohort. *Environ. Health Perspect.* 122, 1059–1065. <https://doi.org/10.1289/ehp.1408198>.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4. <https://doi.org/10.1186/s13742-015-0047-8>.
- Chen, Y., Wu, F., Liu, M., Parvez, F., Slavkovich, V., Eunos, M., Ahmed, A., Argos, M., et al., 2013. A prospective study of arsenic exposure, arsenic methylation capacity, and risk of cardiovascular disease in Bangladesh. *Environ. Health Perspect.* 121, 832–838. <https://doi.org/10.1289/ehp.1205797>.
- Concha, G., Nermell, B., Vahter, M., 1998. Metabolism of inorganic arsenic in children with chronic high arsenic exposure in northern Argentina. *Environ. Health Perspect.* 106, 355–359. <https://doi.org/10.1289/ehp.98106355>.
- de la Barra Saavedra, S.Z., Lara Barrientos, G.M., Coca Cru, R.O., 2011. Exclusión y subalteridad de los urus del lago Poopó Discriminación en la relación mayorías y minorías étnicas. *Serie Informes de Investigación PIEB, La Paz*.
- De Loma, J., Gliga, A.R., Levi, M., Ascuí, F., Gardon, J., Tirado, N., Broberg, K., 2020. Arsenic exposure and cancer-related proteins in urine of indigenous Bolivian women. *Front. Public Health* 8, 605123. <https://doi.org/10.3389/fpubh.2020.605123>.
- De Loma, J., Tirado, N., Ascuí, F., Levi, M., Vahter, M., Broberg, K., Gardon, J., 2019. Elevated arsenic exposure and efficient arsenic metabolism in indigenous women around Lake Poopó, Bolivia. *Sci. Total Environ.* 657, 179–186. <https://doi.org/10.1016/j.scitotenv.2018.11.473>.
- Eichstaedt, C., Antao, T., Cardona, A., Pagani, L., Kivisild, T., Mormina, M., 2015. Positive selection of *AS3MT* to arsenic water in Andean populations. *Mutat. Res.* 780, 97–102. <https://doi.org/10.1016/j.mrfmmm.2015.07.007>.
- Engström, K., Vahter, M., Mlakar, S.J., Concha, G., Nermell, B., Raqib, R., Cardozo, A., Broberg, K., 2011. Polymorphisms in arsenic (+III oxidation state) methyltransferase (*AS3MT*) predict gene expression of *AS3MT* as well as arsenic metabolism. *Environ. Health Perspect.* 119, 182–188. <https://doi.org/10.1289/ehp.1002471>.
- Engström, K.S., Hossain, M.B., Lauss, M., Ahmed, S., Raqib, R., Vahter, M., Broberg, K., 2013. Efficient arsenic metabolism - the *AS3MT* haplotype is associated with DNA methylation and expression of multiple genes around *AS3MT*. *PLoS One* 8, e53732. <https://doi.org/10.1371/journal.pone.0053732>.
- Gardner, R., Nermell, B., Kippler, M., Grandér, M., Li, L., Ekström, E.C., Rahman, A., Lönnerdal, B., et al., 2011. Arsenic methylation efficiency increases during the first trimester of pregnancy independent of folate status. *Reprod. Toxicol.* 31, 210–218. <https://doi.org/10.1016/j.reprotox.2010.11.002>.
- Gautier, M., Vitalis, R., 2012. Rehm an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28, 1176–1177. <https://doi.org/10.1093/bioinformatics/bts115>.
- Gomez-Rubio, P., Meza-Montenegro, M.M., Cantu-Soto, E., Klimecki, W.T., 2010. Genetic association between intronic variants in *AS3MT* and arsenic methylation efficiency is focused on a large linkage disequilibrium cluster in chromosome 10. *J. Appl. Toxicol.* 30, 260–270. <https://doi.org/10.1002/jat.1492>.
- Grau-Perez, M., Kuo, C.C., Gribble, M.O., Balakrishnan, P., Jones Spratlen, M., Vaidya, D., Francesconi, K.A., Goessler, W., et al., 2017. Association of low-moderate arsenic exposure and arsenic metabolism with incident diabetes and insulin resistance in the Strong Heart Family Study. *Environ. Health Perspect.* 125, 127004. <https://doi.org/10.1289/EHP2566>.
- Harari, F., Langeén, M., Casimiro, E., Bottai, M., Palm, B., Nordqvist, H., Vahter, M., 2015. Environmental exposure to lithium during pregnancy and fetal size: a longitudinal study in the Argentinean Andes. *Environ. Int.* 77, 48–54. <https://doi.org/10.1016/j.envint.2015.01.011>.
- Hopenhayn-Rich, C., Biggs, M., Lou, Smith, A.H., Kalman, D.A., Moore, L.E., 1996. Methylation study of a population environmentally exposed to arsenic in drinking water. *Environ. Health Perspect.* 104, 620–628. <https://doi.org/10.1289/ehp.96104620>.
- Howe, C.G., Niedzwiecki, M.M., Hall, M.N., Liu, X., Ilievski, V., Slavkovich, V., Alam, S., Siddique, A.B., et al., 2014. Folate and cobalamin modify associations between S-adenosylmethionine and methylated arsenic metabolites in arsenic-exposed Bangladeshi adults. *J. Nutr.* 144, 690–697. <https://doi.org/10.3945/jn.113.188789>.
- IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012. *Arsenic, metals, fibres, and dusts*. IARC monographs on the evaluation of carcinogenic risks to humans/World Health Organization. *Int. Agency Res. Canc.* 100, 11–465 <https://doi.org/10.9782/9283213208>.
- Jacovas, V.C., Couto-Silva, C.M., Nunes, K., Lemes, R.B., de Oliveira, M.Z., Salzano, F.M., Bortolini, M.C., Hünemeier, T., 2018. Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. *Sci. Rep.* 8, 12733. <https://doi.org/10.1038/s41598-018-31100-6>.
- Khan, K.M., Chakraborty, R., Bundschuh, J., Bhattacharya, P., Parvez, F., 2020. Health effects of arsenic exposure in Latin America: an overview of the past eight years of research. *Sci. Total Environ.* 710, 136071. <https://doi.org/10.1016/j.scitotenv.2019.136071>.
- Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P. H., Schraiber, J.G., et al., 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513, 409–413. <https://doi.org/10.1038/nature13673>.
- Li, H., Engström, K., Vahter, M., Broberg, K., 2012. Arsenic exposure through drinking water is associated with longer telomeres in peripheral blood. *Chem. Res. Toxicol.* 25, 2333–2339. <https://doi.org/10.1021/tx300222t>.
- Lin, S., Shi, Q., Nix, F.B., Styblo, M., Beck, M.A., Herbin-Davis, K.M., Hall, L.L., Simeonsson, J.B., et al., 2002. A novel S-Adenosyl-L-methionine:Arseenic(III) methyltransferase from rat liver cytosol. *J. Biol. Chem.* 277, 10795–10803. <https://doi.org/10.1074/jbc.M110246200>.
- Lindberg, A.L., Ekström, E.C., Nermell, B., Rahman, M., Lönnerdal, B., Persson, L.Å., Vahter, M., 2008a. Gender and age differences in the metabolism of inorganic arsenic in a highly exposed population in Bangladesh. *Environ. Res.* 106, 110–120. <https://doi.org/10.1016/j.envres.2007.08.011>.
- Lindberg, A.L., Rahman, M., Persson, L.Å., Vahter, M., 2008b. The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. *Toxicol. Appl. Pharmacol.* 230, 9–16. <https://doi.org/10.1016/j.taap.2008.02.001>.
- Lindo, J., Haas, R., Hofman, C., Apata, M., Moraga, M., Verdugo, R.A., Watson, J.T., Llave, C.V., et al., 2018. The genetic prehistory of the Andean highlands 7000 years BP through European contact. *Sci. Adv.* 4, eaau4921 <https://doi.org/10.1126/sciadv.aau4921>.
- Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., Chen, W.M., Barrett, J., 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>.
- McLaren, W., Gil, L., Hunt, S.E., Riat, S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F., 2016. The Ensembl variant effect predictor. *Genome Biol.* 17, 122. <https://doi.org/10.1186/s13059-016-0974-4>.
- Nelson, R.M., Pettersson, M.E., Li, X., Carlborg, Ö., 2013. Variance heterogeneity in *Saccharomyces cerevisiae* expression data: trans-regulation and epistasis. *PLoS One* 8 (11), e79507. <https://doi.org/10.1371/journal.pone.0079507>.
- Núñez, L., Grosjean, M., Cartajena, I., 2002. Human occupations and climate change in the Puna de Atacama, Chile. *Science* 298, 821–824.
- Pierce, B.L., Kibriya, M.G., Tong, L., Jasmine, F., Argos, M., Roy, S., Paul-Brutus, R., Rahaman, R., et al., 2012. Genome-wide association study identifies chromosome 10q24.32 variants associated with arsenic metabolism and toxicity phenotypes in Bangladesh. *PLoS Genet.* 8, e1002522 <https://doi.org/10.1371/journal.pgen.1002522>.
- Pu, Y.S., Yang, S.M., Huang, Y.K., Chung, C.J., Huang, S.K., Chiu, A.W.H., Yang, M.H., Chen, C.J., et al., 2007. Urinary arsenic profile affects the risk of urothelial carcinoma even at low arsenic exposure. *Toxicol. Appl. Pharmacol.* 218, 99–106. <https://doi.org/10.1016/j.taap.2006.09.021>.
- Rahman, A., Persson, L.Å., Nermell, B., El Arifeen, S., Ekström, E.C., Smith, A.H., Vahter, M., 2010. Arsenic exposure and risk of spontaneous abortion, stillbirth, and infant mortality. *Epidemiology* 21, 797–804. <https://doi.org/10.1097/EDE.0b013e3181f56a0d>.
- Rahman, A., Vahter, M., Ekström, E.C., Persson, L.Å., 2011. Arsenic exposure in pregnancy increases the risk of lower respiratory tract infection and diarrhea during infancy in Bangladesh. *Environ. Health Perspect.* 119, 719–724. <https://doi.org/10.1289/ehp.1002265>.
- Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., et al., 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837. <https://doi.org/10.1038/nature01140>.

- Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., et al., 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913–918. <https://doi.org/10.1038/nature06250>.
- Sandoval, J.R., Lacerda, D.R., Jota, M.S.A., Salazar-Granara, A., Paulo, P., Vieira, R., Acosta, O., Cuellar, C., et al., 2013. The genetic history of indigenous populations of the Peruvian and Bolivian Altiplano: the legacy of the Uros. *PLoS One* 8, e73006. <https://doi.org/10.1371/journal.pone.0073006>.
- Schlebusch, C.M., Gattepaille, L.M., Engström, K., Vahter, M., Jakobsson, M., Broberg, K., 2015. Human adaptation to arsenic-rich environments. *Mol. Biol. Evol.* 32, 1544–1555. <https://doi.org/10.1093/molbev/msv046>.
- Schlebusch, C.M., Lewis, C.M., Vahter, M., Engström, K., Tito, R.Y., Obregón-Tito, A.J., Huerta, D., Polo, S.I., et al., 2013. Possible positive selection for an arsenic-protective haplotype in humans. *Environ. Health Perspect.* 121, 53–58. <https://doi.org/10.1289/ehp.1205504>.
- Shih, Y.H., Islam, T., Hore, S.K., Sarwar, G., Shahriar, M.H., Yunus, M., Graziano, J.H., Harjes, J., et al., 2017. Associations between prenatal arsenic exposure with adverse pregnancy outcome and child mortality. *Environ. Res.* 158, 456–461. <https://doi.org/10.1016/j.envres.2017.07.004>.
- Shriver, M.D., Kennedy, G.C., Parra, E.J., Lawson, H.A., Sonpar, V., Huang, J., Akey, J. M., Jones, K.W., 2004. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum. Genom.* 1, 274–286. <https://doi.org/10.1186/1479-7364-1-4-274>.
- Smith, A.H., Marshall, G., Liaw, J., Yuan, Y., Ferreccio, C., Steinmaus, C., 2012. Mortality in young adults following in utero and childhood exposure to arsenic in drinking water. *Environ. Health Perspect.* 120, 1527–1531. <https://doi.org/10.1289/ehp.1104867>.
- Smith, A.H., Yunus, M., Khan, A.F., Ercumen, A., Yuan, Y., Smith, M.H., Liaw, J., Balmes, J., et al., 2013. Chronic respiratory symptoms in children following in utero and early life exposure to arsenic in drinking water in Bangladesh. *Int. J. Epidemiol.* 42, 1077–1086. <https://doi.org/10.1093/ije/dyt120>.
- Steinmaus, C., Bates, M.N., Yuan, Y., Kalman, D., Atallah, R., Rey, O.A., Biggs, M.L., Hopenhayn, C., et al., 2006. Arsenic methylation and bladder cancer risk in case-control studies in Argentina and the United States. *J. Occup. Environ. Med.* 48, 478–488. <https://doi.org/10.1097/01.jom.0000200982.28276.70>.
- Steinmaus, C., Yuan, Y., Kalman, D., Rey, O.A., Skibola, C.F., Dauphine, D., Basu, A., Porter, K.E., et al., 2010. Individual differences in arsenic metabolism and lung cancer in a case-control study in Cordoba, Argentina. *Toxicol. Appl. Pharmacol.* 247, 138–145. <https://doi.org/10.1016/j.taap.2010.06.006>.
- Tseng, C.H., Huang, Y.K., Huang, Y.L., Chung, C.J., Yang, M.H., Chen, C.J., Hsueh, Y.M., 2005. Arsenic exposure, urinary arsenic speciation, and peripheral vascular disease in Blackfoot disease-hyperendemic villages in Taiwan. *Toxicol. Appl. Pharmacol.* 206, 299–308. <https://doi.org/10.1016/j.taap.2004.11.022>.
- Vahter, M., 2002. Mechanisms of arsenic biotransformation. *Toxicology* 181–182, 211–217. [https://doi.org/10.1016/S0300-483X\(02\)00285-8](https://doi.org/10.1016/S0300-483X(02)00285-8).
- Vahter, M., Concha, G., 2001. Role of metabolism in arsenic toxicity. *Pharmacol. Toxicol.* 89, 1–5. <https://doi.org/10.1111/j.1600-0773.2001.890101.x>.
- Vahter, M., Concha, G., Nermell, B., Nilsson, R., Dulout, F., Natarajan, A.T., 1995. A unique metabolism of inorganic arsenic in native Andean women. *Eur. J. Pharmacol.* 293, 455–462. [https://doi.org/10.1016/0926-6917\(95\)90066-7](https://doi.org/10.1016/0926-6917(95)90066-7).
- Vicente, M., Jakobsson, M., Ebbesen, P., Schlebusch, C.M., 2019. Genetic affinities among Southern Africa hunter-gatherers and the impact of admixing farmer and herder populations. *Mol. Biol. Evol.* 36, 1849–1861. <https://doi.org/10.1093/molbev/msz089>.
- Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4, e72. <https://doi.org/10.1371/journal.pbio.0040072>.