Human Adaptation to Arsenic-Rich Environments

Carina M. Schlebusch,[†,1] Lucie M. Gattepaille,[†,1] Karin Engström,[2] Marie Vahter,[3] Mattias Jakobsson,[*,1,4] and Karin Broberg[*,3]

[1]Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

[2]Division of Occupational and Environmental Medicine, Lund University, Lund, Sweden

[3]Unit of Metals and Health, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[4]Science for Life Laboratory, Uppsala University, Uppsala, Sweden

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: mattias.jakobsson@ebc.uu.se; karin.broberg@ki.se.

Associate editor: Joshua Akey.

Abstract

Adaptation drives genomic changes; however, evidence of specific adaptations in humans remains limited. We found that inhabitants of the northern Argentinean Andes, an arid region where elevated arsenic concentrations in available drinking water is common, have unique arsenic metabolism, with efficient methylation and excretion of the major metabolite dimethylated arsenic and a less excretion of the highly toxic monomethylated metabolite. We genotyped women from this population for 4,301,332 single nucleotide polymorphisms (SNPs) and found a strong association between the *AS3MT* (arsenic [+3 oxidation state] methyltransferase) gene and mono- and dimethylated arsenic in urine, suggesting that *AS3MT* functions as the major gene for arsenic metabolism in humans. We found strong genetic differentiation around *AS3MT* in the Argentinean Andes population, compared with a highly related Peruvian population (F_{ST} = 0.014) from a region with much less environmental arsenic. Also, 13 of the 100 SNPs with the highest genome-wide Locus-Specific Branch Length occurred near *AS3MT*. In addition, our examination of extended haplotype homozygosity indicated a selective sweep of the Argentinean Andes population, in contrast to Peruvian and Colombian populations. Our data show that adaptation to tolerate the environmental stressor arsenic has likely driven an increase in the frequencies of protective variants of *AS3MT*, providing the first evidence of human adaptation to a toxic chemical.

Key words: **selection, toxic, AS3MT.**

Introduction

Many organisms have adapted to tolerate toxic chemicals in their environments (Melnick and Parkin 2007; Hanikenne et al. 2008; Song et al. 2010); however, we know little about human adaptation to toxic chemicals. Exposure to inorganic arsenic is associated with multiple severe health effects, including increased morbidity and mortality in early life (Rahman, Persson, et al. 2010; Rahman, Vahter, et al. 2010; Gardner et al. 2013), cancer, cardiovascular and liver toxicity, and probably diabetes (IARC 2012; Maull et al. 2012; Moon et al. 2012). In a few regions in the world, such as the Andean highlands, human populations have lived for thousands of years with drinking water contaminated with arsenic (fig. 1). This raises the question as to whether such populations may have adapted over time to their toxic environment.

The efficiency of arsenic metabolism strongly affects susceptibility to arsenic toxicity. In the body, cellular enzymes methylate inorganic arsenic to monomethylarsonic acid (MMA) and then dimethylarsinic acid (DMA). The fraction of arsenic present as MMA shows a positive association with arsenic toxicity, indicating that MMA is more toxic than DMA (Ahsan et al. 2007; Lindberg et al. 2008; Pierce et al. 2013). By contrast, DMA is more readily excreted in urine and expelled from the body (Gardner et al. 2010). The fractions of arsenic metabolites in human urine vary in different populations (fraction of MMA: 2–30%) (Vahter 2002). Indigenous populations in the Andes, including in the Argentinean village of San Antonio de los Cobres (SAC; fig. 1), show uniquely low urinary excretion of MMA (Vahter et al. 1995).

The enzyme arsenic (+3 oxidation state) methyltransferase (AS3MT) plays a key role in arsenic methylation (Engstrom et al. 2011). Polymorphisms in *AS3MT* are associated with the arsenic methylation as shown in several populations, e.g., in Bangladesh, Argentina, Mexico, and Taiwan (Chung et al. 2009; Fujihara et al. 2009; Gomez-Rubio et al. 2010; Engstrom et al. 2011; Pierce et al. 2012). In particular, the *AS3MT* alleles associated with efficient arsenic methylation vary markedly in frequency (Fujihara et al. 2009; Schlebusch et al. 2013). Individuals from SAC and surrounding villages have higher frequencies of inferred protective *AS3MT* haplotypes than other Native American and Asian populations (Schlebusch et al. 2013). This observation led to the hypothesis that natural selection has favored *AS3MT* haplotypes that associate with more efficient arsenic metabolism in populations that have lived with arsenic exposure for many generations. In this study, we performed a genome-wide

FIG. 1. Study populations. Sites for study populations and evidence for pre-Columbian arsenic exposure in humans.

association study (GWAS) using dense, genome-wide markers, and well-characterized arsenic metabolism phenotypes to demonstrate that *AS3MT* is likely to be the leading gene for arsenic methylation in humans. In strong support of our hypothesis we found that in the people from SAC, the genomic region around *AS3MT* shows dramatic signs of selection, indicating adaptation to arsenic-rich environments.

Results

We used the Illumina 5M Omni chip to genotype 124 arsenic-exposed women from SAC who had a wide range of percentages of the monomethylated arsenic metabolite MMA (%MMA) in their urine (table 1). Their distribution of MMA was representative of the distribution observed in a larger sample of the population of SAC (supplementary fig. S1, Supplementary Material online). We performed GWAS on 1,258,737 filtered single nucleotide polymorphisms (SNPs) for all arsenic metabolism phenotypes (%MMA, %DMA, and %inorganic arsenic, all adjusted for total arsenic in urine). In a GWA scan unadjusted for further covariates and population structure, we found clear associations for

Table 1. Argentinean Women from SAC Selected for the Study ($N = 124$) and Some of Their Characteristics.

Variable	Median	Range
%iAs	13.7	3.3–33
%MMA	7.5	1.2–22
%DMA	78	56–93
Total urinary arsenic	270	61–660
Age	35	14–71
BMI	25	16–36

chromosome 10 (MMA P-value = 4.658e-08, FDR corrected $q = 0.00072$; DMA $P = 1.002e-07$, FDR $q = 0.0035$) and chromosome 21 (MMA $P = 1.238e-05$, FDR $q = 0.019$; DMA $P = 1.204e-05$, FDR $q = 0.040$) in all scans involving %DMA and %MMA (fig. S2, supplementary tables S1–S3, Supplementary Material online). Zoomed views of the peaks (supplementary figs. S3–S5, Supplementary Material online) show many SNPs that are elevated, in particular upstream of *AS3MT* on chromosome 10, where active regulatory

elements are found (supplementary fig. S6, Supplementary Material online), and fewer SNPs for the peak on chromosome 21. We also found some significant but less pronounced peaks: for %MMA on chromosomes 3, 6, and 13, and for %DMA on chromosomes 12 and 13 (supplementary table S1 and fig. S4 and S5, Supplementary Material online). We found no peaks for %inorganic arsenic that reached FDR-corrected significance (supplementary fig. S7, Supplementary Material online). The quantile–quantile plots for all GWA analyses are given in supplementary figure S8 (Supplementary Material online).

To account for population structure and relatedness, we adjusted the data for possible stratification as well as for possible influential covariates. The results showed that clear, significant (FDR-corrected) peaks remained visible on the following chromosomes: %MMA: 2, 10, 21; and %DMA: 2, 10, 12, 13, 21 (fig. 2, including zoomed views of the peak on chromosome 10). Moreover, examination of population structure using principal component analysis (Patterson et al. 2006) and population structure analysis (Alexander et al. 2009) did not reveal any significant population structure within the SAC group (supplementary fig. S9, Supplementary Material online). This analysis also showed limited admixture from Iberians (Iberian cluster membership in SAC population at K = 4, mean = 2.6%, STD = 5.2%, supplementary fig. S9, Supplementary Material online).

We furthermore scanned for loci that significantly influence variance (Yang et al. 2010) of %MMA and %DMA and thus might be involved in controlling the expression of main effect genes. SNPs associated with variance in %MMA, and more weakly in %DMA, were found on chromosome 1 (supplementary fig. S10, Supplementary Material online); the associated SNPs are not located in a known gene but intersect with an active regulatory region (supplementary fig. S11, Supplementary Material online). Although no SNPs were significantly associated with variance of %DMA, we found a clear peak in chromosome 15 within the ryanodine receptor 3 (*RYR3*) gene (supplementary fig. S10, Supplementary Material online).

Evidence for Selection

To detect regions of the genome that may have been targets of selection, we used a statistic that captures greater than typical levels of differentiation in one population compared with two other populations (the locus-specific branch length [LSBL] statistic [Shriver et al. 2004]). The LSBL values in SAC, using Peruvian and Colombian individuals (PEL and CLM in figures) from the 1000 Genomes Project data as comparative populations, revealed a strong peak on chromosome 10 in the region of *AS3MT* (fig. 3*A*). This peak was absent from the LSBL scan focusing on the Peruvian population (compared with Colombian and SAC populations, supplementary fig. S12, Supplementary Material online), despite the strong genetic similarities between the Peruvian and SAC populations (genome-wide average $F_{ST} = 0.014$, supplementary fig. S9, Supplementary Material online). The SNP in the peak with the greatest LSBL value for SAC had a genome-wide percentile value of 99.9976% (only 34 SNPs out of the 1,456,054 SNPs that were polymorphic in all three populations have a greater LSBL value). Among the 100 SNPs with greatest genome-wide LSBL values, 13 occurred within the particular peak near *AS3MT*.

Additional evidence for positive selection for arsenic tolerance in the SAC population was revealed by the haplotype homozygosity statistic iHS (Voight et al. 2006). We found elevated iHS values in the *AS3MT* region for the SAC population (using either 3,705,093 SNPs or the reduced set combined with the 1000 Genomes set of 2,023,892 SNPs), whereas there was no signal for the Peruvian population (fig. 3*B*). The individual SNP iHS values were not extraordinarily large around *AS3MT* for the SAC population; however, all SNPs in this region were generally elevated, compared with the background level of iHS, as revealed by a moving average over windows of 1 Mb (fig. 3*C*). The greatest 1 Mb-window |iHS| value in the region of *AS3MT* was in the top 97-percentile (genome-wide) for the SAC population, in contrast to the top 89-percentile for the Peruvian populations (fig. 3*D*). The Peruvian population did not show an elevated signal around the *AS3MT* gene. Besides the *AS3MT* region on chromosome 10, all other regions which showed significant association to %MMA and %DMA during GWA scans (such as the chromosome 21 region), did not show significant evidence of selection in the SAC population. Genes previously linked to arsenic metabolism, e.g., *GSTO1/2* and *MTHFR* (Broberg et al 2015), did not show any significant GWAS associations and therefore were not further evaluated in selection scans.

The exact time of onset of arsenic-related selective pressure in the SAC population is unknown, but it probably lies between the putative date of settlement of the population in the region, estimated at 11,000 years ago (Núñez et al. 1991), and 7,000 years ago, the estimated age of a regionally excavated mummy that displayed high concentrations of arsenic in hair, suggesting strong exposure to arsenic (Arriaza et al. 2010). Neglecting effects of recent genetic drift since divergence between the SAC and the Peruvian populations, we estimate the selection coefficient due to arsenic exposure to range between 0.003 and 0.005 in the SAC population (supplementary fig. S13, Supplementary Material online).

Global Distribution of Haplotypes Associated with Arsenic Metabolism Phenotypes

We wanted to compare the frequencies of putative protective variants between the SAC population and comparable population groups. However, since many SNPs on chromosome 10q24 are strongly associated with arsenic metabolism and contribute to the phenotypes in a similar, additive manner (supplementary fig. S14, Supplementary Material online), it is possible that these SNPs are linked in one high-frequency protective haplotype. Rather than comparing frequencies of single, highly significant SNPs, we inferred a putative protective haplotype by combining strongly associated SNPs in the chr 10 peak, and compared the frequencies of this haplotype in different populations. To identify the protective haplotype,

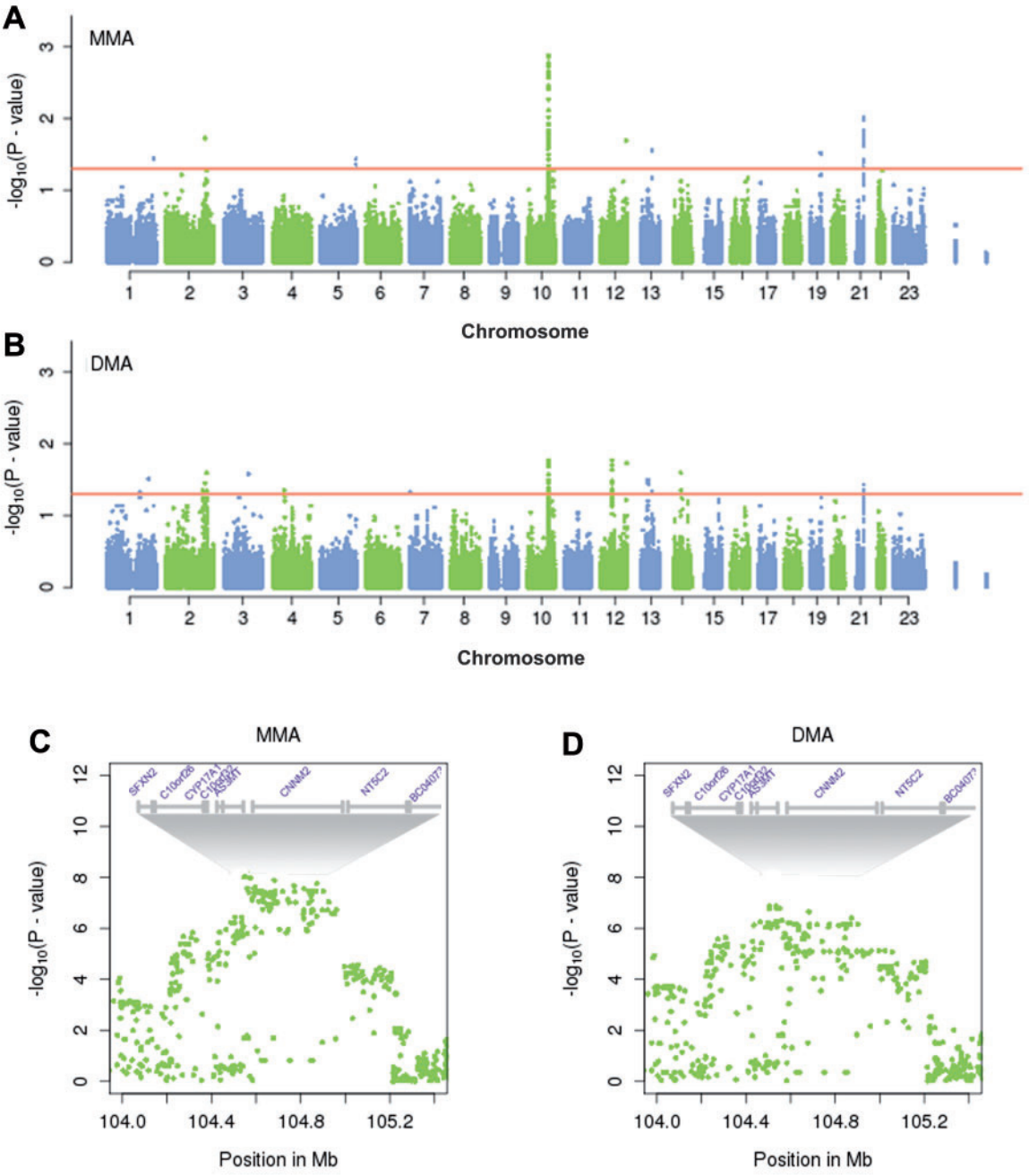

FIG. 2. GWAS study of arsenic metabolism phenotypes. Manhattan plots showing FDR-adjusted *q*-values for genome-wide associations taking possible population stratification into account as well as influential covariates between (*A*) fraction of MMA in urine and (*B*) fraction of DMA in urine. All phenotypes were adjusted for total arsenic in urine. Zoom-in figure for the peak of significant SNPs on chromosome 10 associated with (*C*) %MMA and (*D*) %DMA.

we extracted all SNPs with highly significant FDR-corrected *q*-values ($q < 0.01$) for the %MMA GWA scan, identified the specific alleles associated with low %MMA and linked these SNP alleles into a putative "MMA-based protective haplotype." We found that this putative protective haplotype exactly matched 58.4% of the phased haplotypes in the SAC population. In the comparative data, we observed the greatest frequencies of exact matches in Peruvians (29.1% of all phased haplotypes) and East Asian populations (26.8% of phased haplotypes in the Vietnamese [Kinh in Ho Chi Minh City] population). For haplotypes that did not exactly match the protective haplotype, we counted the number of mutational differences to the putative protective haplotype and visualized the differences as violin plots for the different populations (fig. 4). Violin plots resemble box plots but show the probability density of the data at different values on the *y*-axis.

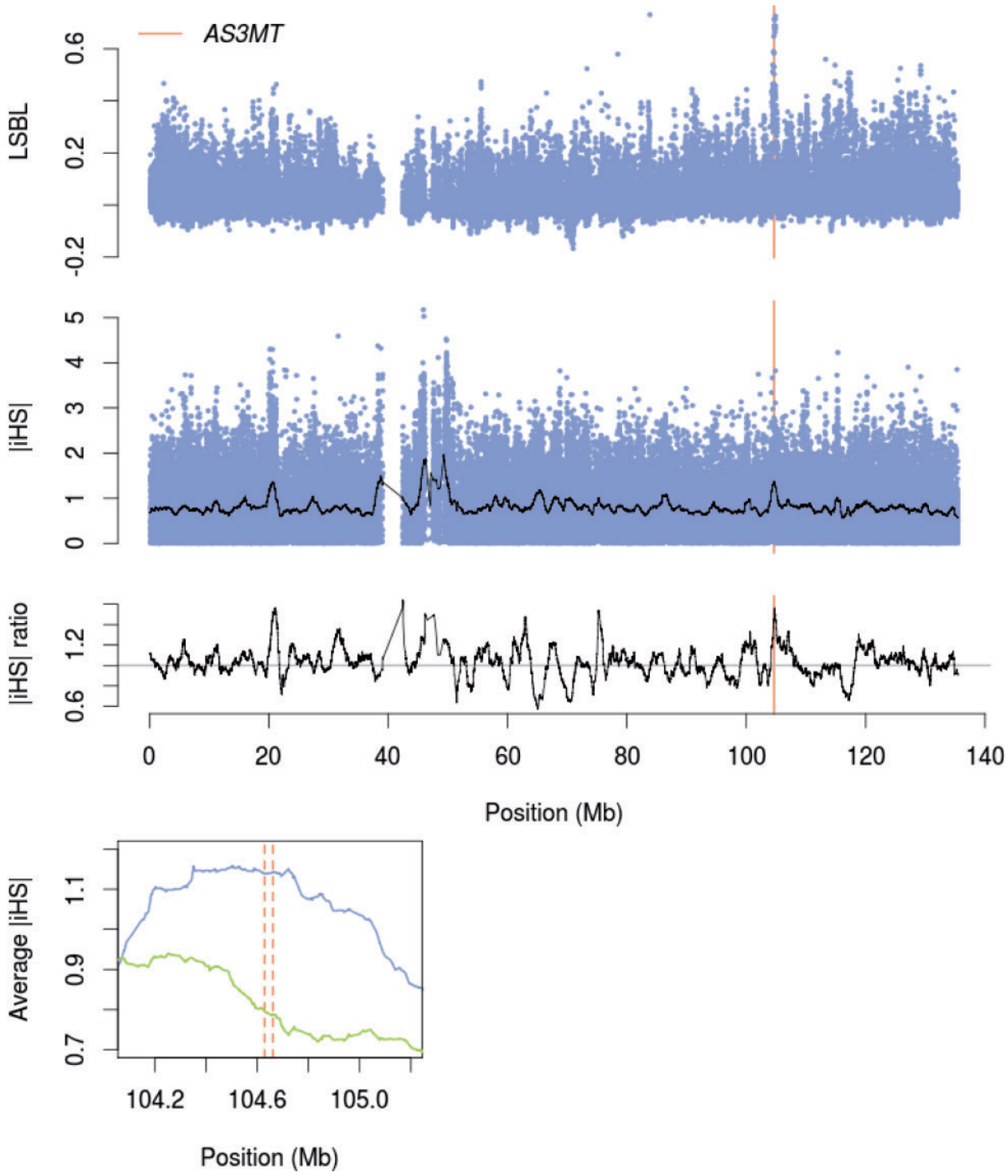

FIG. 3. Results of the selection scan for chromosome 10. (*A*) LSBL signal for the SAC versus PEL and CLM populations. (*B*) Homozygosity around *AS3MT* shown as |iHS| values per SNP for SAC large (blue dots) and sliding average of |iHS| values over 1 Mb windows (black line). (*C*) Ratio of sliding averages over 1 Mb windows of |iHS| values for the combined SNP set between SAC and PEL populations. In (A)–(C), the position of *AS3MT* on chromosome 10 is indicated by a vertical orange line. (*D*) Zoom-in of the region around *AS3MT* (indicated by vertical dashed lines) shown as sliding averages over 1 Mb windows of |iHS| values for the SAC (blue line) and PEL (green line) populations.

Thus, for our data, a complete match to the inferred protective haplotype would be 0 on the *y*-axis; increasing distance from 0 on the *y*-axis indicates an increasing number of differences from the inferred protective haplotype. The width of the violin plot indicates the number of individual haplotypes at each *y*-axis value and the *x*-axis represents the SAC population and the different comparison populations from the 1000 Genomes Project. The SAC populations had the greatest number of exact matches (*y*-axis = 0 in fig. 4) to the inferred protective haplotype, as well as high frequencies of haplotypes

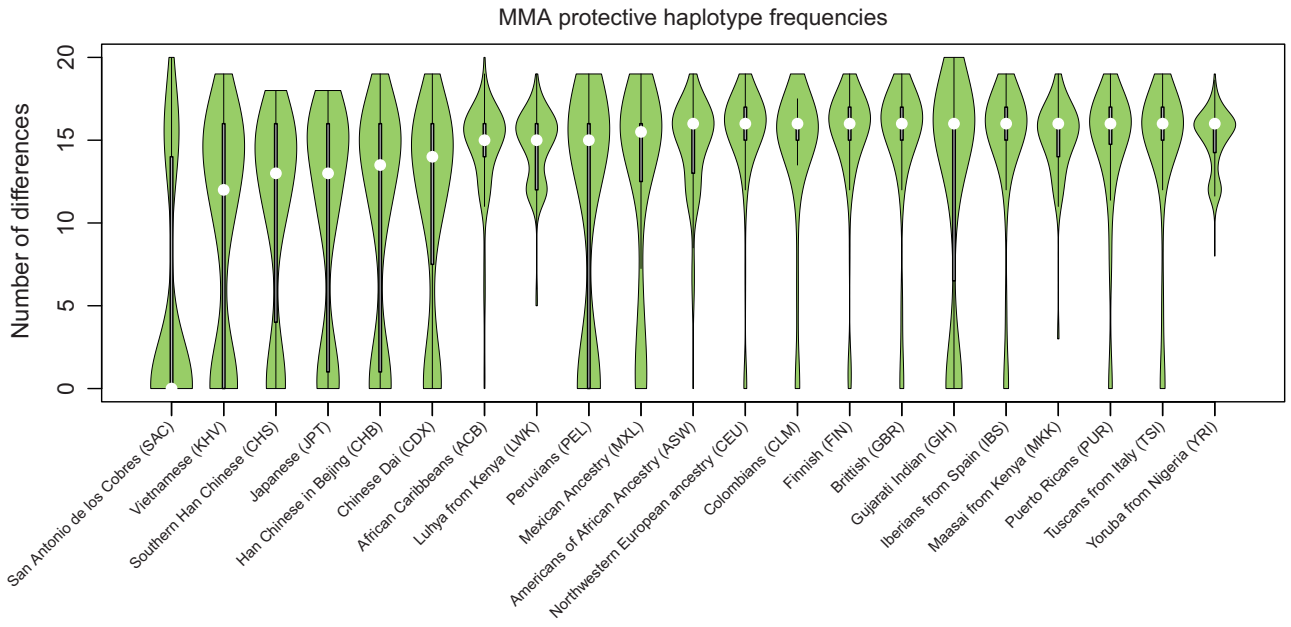

Fig. 4. Global distribution of protective haplotypes. The number of mutational differences (*y*-axis) to the %MMA-based protective haplotypes visualized by the differences in the form of violin plots for the different comparative populations from the 1000 Genome project (*x*-axis).

that closely matched the inferred protective haplotype. Exact (or close) matches to the putative protective haplotype are spread globally, but tend to be more frequent in East Asians and Native Americans. Furthermore, to see the correlation between the number of mutational differences from the protective haplotype and the phenotype, we plotted the %MMA versus differences from the protective haplotype (supplementary fig. S15, Supplementary Material online). Differences to the protective haplotype were significantly correlated with %MMA. To visualize haplotypes of individuals in different populations and the distribution of mutational differences to the inferred protective haplotype, we plotted phased haplotypes of individuals against the putative protective haplotype (as reference sequence) in the form of haplotype plots (supplementary fig. S16, Supplementary Material online). The high frequency of the protective haplotype in the SAC population was clearly visible and differences to the protective haplotype in the SAC were distributed toward the ends of the haplotype block. Repeating haplotype analyses with %DMA as the phenotype yielded a similar haplotype as that found with %MMA.

Discussion

In this study, we found very strong and unique GWAS signals for arsenic metabolism phenotypes in the region on chromosome 10q24.32 harboring the *AS3MT* gene; this establishes the association of the arsenic metabolism phenotype with this genomic region and provides strong evidence that AS3MT is likely to be the major contributor to arsenic metabolism in humans. In order to investigate this possibility further, we used multiple different and established tools to look for possible selection for efficient arsenic metabolism. We identified a signal for selection based on extended haplotype homozygosity for *AS3MT* in the Argentinean, but not in the Peruvian or Colombian populations, the latter presumably having lower historic arsenic exposure (Bundschuh et al. 2012). Moreover, we found increased genetic differentiation around the *AS3MT* locus in comparisons between SAC and other South American groups, in contrast to the overall genome-wide level of population differentiation. This additional level of genetic differentiation around *AS3MT* suggests that the *AS3MT* locus has been the target of selection for adaptation to an arsenic-rich environment in a population presumably exposed to arsenic for thousands of years. We estimated the selection coefficient to range between 0.003 and 0.005, which is smaller than the estimated selection coefficients associated with lactase persistence (Tishkoff et al. 2007) and resistance to malaria (Chen and Slatkin 2013), two previously well-established examples of strong positive selection leading to adaptation in humans. Further, we extracted haplotypes containing SNPs significantly associated with the arsenic metabolism phenotypes and saw that these SNPs occur on the same haplotype background (or a very similar background) that is present worldwide, but much more enriched in SAC. The widespread occurrence of the putative protective haplotype could thus be an indication that selection acted on standing variation that was already present in the population that settled in regions with high arsenic in the drinking water. Because the extracted haplotypes were very similar for MMA and DMA, it is likely that recent selection has occurred in the SAC population, increasing the frequency of one protective haplotype in the *AS3MT* region.

We focused on *AS3MT* as different studies have shown it to be important for arsenic methylation: in experimental studies in rat (Lin et al. 2002), human kidney, and liver cells (Drobna et al. 2005, 2006) as well as in vivo studies of a mouse *AS3MT* knockout (Drobna et al. 2009; Chen et al. 2011). Other genes close to *AS3MT* have not been linked to arsenic. This

makes *AS3MT* a very likely candidate gene for the fraction of monomethylated arsenic in urine. The top SNP signals were, however, upstream of *AS3MT*, which points to a potential regulatory function for these SNPs, a notion that is in line with the histone modification pattern found in the top SNP region. Selection signals outside of genes have previously been found, e.g., for lactose tolerance (Tishkoff et al. 2007). The protective haplotype stretches over ~500 kb and could contain one or more protective variants. Identifying which variants drive selection will require functional studies on all variant loci. Strikingly, all mutated alleles show variance in noncoding sequences only. It therefore seems unlikely that differences in the catalytic properties of the encoded proteins can explain the observed phenotypes. However, there is evidence that SNPs in the protective haplotype with *AS3MT* and surrounding genes are functional: we and others previously analyzed the expression of genes in this chromosomal region in whole blood (as a proxy for the liver, which metabolizes arsenic) and found that expression of *AS3MT*, and surrounding genes, was significantly altered in association with an increased number of *AS3MT* protective alleles (Engstrom et al. 2011; Pierce et al. 2012; Engstrom et al. 2013).

The finding that the strongest association signals for arsenic phenotypes %MMA and %DMA occur in and around *AS3MT* agrees with the genomic region reported by Pierce et al. (2012). We also found some other, not previously identified, chromosomal regions associated with arsenic phenotypes. So far, no relationships between genes in these regions and arsenic metabolism have been reported, but one should bear in mind that they might actually represent markers of arsenic toxicity linked to arsenic metabolism, because we used arsenic-metabolizing phenotypes as proxies for arsenic susceptibility. In particular, on chromosome 21q22.3, we observed associations for both %MMA and %DMA. The strongest associations were found for *LSS*, lanosterol synthase (2,3-oxidosqualene-lanosterol cyclase); *MCM3AP*, minichromosome maintenance complex component 3 associated protein; *MCM3AP-AS1*, MCM3AP antisense RNA 1; and *YBEY*, a putative metallopeptidase. LSS catalyzes the formation of lanosterol, the first sterol precursor of cholesterol in humans (Thoma et al. 2004). MCM3AP has been functionally characterized as an acetyltransferase for acetylation of replication protein MCM3 (Takei et al. 2001), but for the other genes little functional information is available at the moment. There were a few other chromosomal peaks that associated with arsenic phenotypes. For %MMA there was a peak near the *CADM2* gene on chromosome 3. For %DMA there was a peak on chromosome 12 encompassing a cluster of genes from the olfactory receptor family 6, subfamily C. The olfactory receptor genes encompass a large multigene family encoding transmembrane signaling proteins required for odorant discrimination. None of these chromosomal peaks overlapped with the chromosomal regions recently reported from a low-density microsatellite marker ($n = 400$) analysis that mapped loci affecting arsenic metabolites in urine (Tellez-Plaza et al. 2013).

The GWAS for explaining variance in the arsenic phenotypes identified novel regions of potential importance for regulation of arsenic metabolism but, as mentioned earlier, possibly also for arsenic toxicity. For DMA, we found a nonsignificant but clear signal within the *RYR3* gene. Ryanodine receptors, such as RYR3, are intracellular calcium ion release channels responsible for the release of Ca^{2+} from intracellular stores following transduction of many different extracellular stimuli. RYR3 is important in muscle contraction, including during the neonatal period (Bertocchini et al. 1997; Perez et al. 2005). There is no known function of Ca^{2+} in arsenic metabolism; however, inorganic arsenic was shown in vitro to induce Ca^{2+} signals and cytotoxicity meditated by ryanodine receptors (Florea et al. 2007).

The selection for efficient arsenic metabolism could act through different mechanisms and during different periods in life. Studies of animals and children showed that arsenic affects the immune system (Fry et al. 2007; Kozul et al. 2009; Ahmed et al. 2011) and increases infant morbidity and mortality (Rahman, Persson, et al. 2010; Rahman, Vahter, et al. 2010; Gardner et al. 2013). Arsenic crosses the placenta and arsenic exposure during pregnancy enhances placental inflammatory responses, reduces placental T cells, alters cord blood cytokines, and impairs thymic function in the newborn (Fry et al. 2007; Ahmed et al. 2012). In Bangladesh and in the United States, the risk of upper and lower respiratory tract infections and diarrhea in infants increased with higher maternal arsenic during pregnancy (Rahman et al. 2011; Farzan et al. 2013). Selection for a protective *AS3MT* haplotype could also be caused by adverse effects of arsenic later in life, such as hepatoxicity, cardiovascular disease, and impaired lung function (Smith et al. 2006; Moon et al. 2012; Parvez et al. 2013), which may result in reduced reproduction. Given the severe deleterious health effects of arsenic in both children and adults, individuals who carry the arsenic-tolerance haplotype and thus can metabolize arsenic faster with reduced exposure to toxic metabolites could have a very strong selective advantage in high-arsenic environments. Thus, these adverse effects of arsenic before and during reproductive age may provide the mechanism for adaptation to arsenic-rich environments.

The historical arsenic concentrations in the drinking water of the Peruvian and the Columbian populations are not known, but the present-day levels are generally much lower than levels in the northern part of Argentina, apart from some areas where mining activities have resulted in elevated levels during the last century (Bundschuh et al. 2008, 2012; Cooke and Abbott 2008). Elevated concentrations of arsenic in drinking water seem to be quite common in some areas of the Andes Mountains (Smith et al. 2006; Van Den Bergh et al. 2010) and several reports show that other Native American populations living in Chile and Mexico in areas with historical arsenic exposure have efficient arsenic methylation (Hopenhayn-Rich et al. 1996; Meza et al. 2005; Gomez-Rubio et al. 2010, 2012).

There are a few well-known cases of selection in humans, such as lactase persistence, which occurred independently in the same gene (lactase, *LCT*) in Africa and Europe (Tishkoff et al. 2007); copy number variation in the amylase gene (*AMY1*), which improved the capacity to digest starch-rich

diets (Perry et al. 2007); resistance to malaria (Kwiatkowski 2005) and cholera (Karlsson et al. 2013); adaptation to living at high altitudes (Simonson et al. 2010); the production of long-chain fatty acids (Ameur et al. 2012); and light skin pigmentation (Basu Mallick et al. 2013). However, data on human adaptation to toxic compounds remain scarce. The xenobiotic metabolizing N-aryl acetyltransferase family has, based on the molecular evolutionary history, been suggested to be subject to adaptive selection in vertebrates (Sabbagh et al. 2013), but the genetic data did not correlate to a phenotype. In this context, adaptation to an arsenic-rich habitat through the *AS3MT* gene is the first case of human selection for tolerance to a toxic chemical. The potential use of the results of this study in relation to public health will be explored in further studies.

Materials and Methods

Argentinean Study Population

The study site encompassed SAC (3,800 m above sea level) in the Puna area of the Andes Mountains (fig. 1). In this region, arsenic in the volcanic bedrock is released into the groundwater used as drinking water and there are no anthropogenic arsenic sources that affect the water. This arsenic comes from natural reservoirs and modern and pre-Columbian mining activities (Núñez et al. 1991; Idrovo 2005; Bundschuh et al. 2012). The drinking water in SAC contained ~200 μg arsenic/l with small variations over time up until 2012 when a filter system was installed to remove the arsenic (Concha et al. 2006, and unpublished data). The people in SAC are of Atacameño descent and the Atacameños, who once occupied the north of Chile and of Argentina, have lived in the region for 11,000 years (Núñez et al. 1991). There are traces of human settlements in this region of northern Argentina from 1500 BC (Normando Cruz 2011).

Individuals from SAC were interviewed and sampled in 2004, 2005, and 2008. The samples of blood and urine were collected with informed consent (oral and written) and the protocol was approved by the Health Ministry of Salta, Argentina, and the Ethics Committee of the Karolinska Institutet, Sweden. The study subjects were, based on interviews, mainly of indigenous (Atacameño) origin with little ancestry from Hispanics. The families of the participants had lived in the area for at least 2–3 generations.

Water and urine samples were obtained for determination of exposure to inorganic arsenic and the individual arsenic metabolite pattern. Speciation of arsenic metabolites in urine was performed using HPLC hyphenated with hydride generation and inductively coupled plasma mass spectrometry (Agilent 1100 series system; Agilent 7500ce; Agilent Technologies, Japan and Germany), employing adequate and high-quality control (Schlawicke Engstrom et al. 2007). For estimation of exposure to inorganic arsenic, we used the sum of inorganic arsenic, MMA, and DMA concentrations in urine. Urinary arsenic concentrations were adjusted to the mean specific gravity to adjust for variation in urine dilution. The relative amounts (%) of arsenic metabolites in urine were use as measure of methylation efficiency (Vahter 2002).

In total, 385 women had both DNA and urinary arsenic metabolite data. Of those, we selected non-first-degree-related women with a wide range of %MMA in urine ($N = 124$). The individuals selected did not differ in %inorganic arsenic, %MMA, %DMA from the whole study population (supplementary fig. S1, Supplementary Material online). Exclusion criteria were pregnancy, diseases potentially affecting arsenic metabolism, and very high values of urinary arsenic (>600 μg/l; which may inhibit the methylation of arsenic) (table 1).

Whole blood or buccal swabs were collected for DNA extraction (Schlawicke Engstrom et al. 2007; Engstrom et al. 2011). DNA was extracted with the Qiagen Blood Mini kit (Qiagen, Hilden, Germany). Eighteen samples were whole-genome amplified with the REPLI-g Mini kit (Qiagen). The genome-wide genotyping was performed on 400 ng DNA per sample, as determined by Pico green, on the Illumina HumanOmni 5M BeadChip according to the manufacturer's protocol (Illumina, San Diego, CA). Genotyping was performed by the SNP&SEQ Technology Platform in Uppsala, Sweden (www.genotyping.se). Genotype data are available from the EGA data repository (https://www.ebi.ac.uk/ega/home) under the project name "IMM project arsenic adaptation" for researchers who meet the criteria for access to confidential data.

Peruvian and Colombian Study Populations

The selection of samples and the genetic analysis for the 1000 Genomes Project is described in Abecasis et al. (2012). The population used in this study consists of Peruvians ($n = 86$) from Lima, Peru (called PEL) and Colombians ($n = 95$) in Medellin, Colombia (called CLM) (fig. 1) (http://www.1000genomes.org/about#ProjectDesign). All samples were collected from individuals providing a broad informed consent, and there are no individual identifiers linked to genetic data or phenotypes.

Association Studies

Genome-wide analysis was done in GenAbel v. 1.7-6 (Aulchenko et al. 2007). Data were filtered in GenAbel by removing individuals with a call rate below 85%, individuals with too high autosomal heterozygosity (FDR < 1%) and individuals with a pairwise identity by state (IBS) value above 0.9 (excluding highly similar individuals—i.e., duplicate individuals). One individual (one with the lowest call rate) in the pairs of individuals that fail the IBS threshold was removed, leaving 108 individuals. Markers were filtered for a call rate below 90% (217,030 SNPs), a minimum allele frequency below 5% (2,810,794 SNPs), and loci that fail Hardy-Weinberg Equilibrium (HWE) test (P-value $< 1 \times 10^{-8}$) (1,607 SNPs). Furthermore, prior to filtering, duplicate SNPs (55,085 SNPs), SNPs with chromosome position 0 (8,970 SNPs), and indel loci (4,472 SNPs) were removed, leaving 1,258,737 SNPs for association analysis.

The phenotype distributions were close to normally distributed. We performed the GWAS on 1,258,737 filtered SNPs for all arsenic metabolism phenotypes (%MMA, %DMA, and

%inorganic arsenic) using the qtscore function in GenAbel. All statistical analyses were adjusted for total arsenic in urine to account for the fact that higher arsenic exposure has been associated with higher %MMA and lower %DMA (Vahter 2002). FDR corrected *q*-values were calculated using the Benjamini-Hochberg procedure as implemented in GenAbel using the "qvaluebh95" function. Population structure and relatedness in the data were handled by adjustment for principal components of the genomic kinship matrix using the egscore function in GenAbel. Further adjustments were performed for covariates that significantly influence arsenic phenotypes as tested by Pearson's correlation coefficient (%MMA: age, BMI, and year; %DMA: age and BMI; %inorganic arsenic: age and height). We furthermore scanned for loci that significantly influence variance, and thus might be involved in controlling expression of main effect genes, for the arsenic phenotypes %MMA and %DMA, using the zscore method introduced by Yang et al. (2010).

Selection Scan

1) **Datasets.** We performed a genome-wide iHS scan on the SAC data for two different sets of SNPs: a large set, containing 3,705,093 SNPs (*large*), and a second set obtained after merging with 1000 Genomes Project data from a SNP array (*combined*) containing 2,023,892 SNPs. Filters for missingness on genotypes and HWE filters were all performed using PLINK (Purcell et al. 2007).

   Dataset *large*: a 15% genotype missingness threshold was applied to individuals and a 10% genotype missingness threshold was applied to the SNPs. In addition, a HWE rejection confidence level was set to 0.001. We defined the relatedness cutoffs for including / excluding individuals as follows: closest relatives were identified using pairwise IBS and using a threshold of 0.25; we removed the minimum number of individuals guaranteeing that all pairs of remaining individuals have an IBS value lower than 0.25. Because some of the samples were whole-genome amplified whereas others were not, we removed the SNPs with especially different allele frequencies between the two types of samples (see supplementary fig. S17, Supplementary Material online). A/T and C/G SNPs were removed to prevent strand issues in the identification of the ancestral allele (see later) and only SNPs with ancestral information were retained. After filtering, the *large* dataset contained 95 individuals, and 3,705,093 SNPs.

   Dataset *combined*: we downloaded SNP data from the 1000 Genomes Project, at ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120131_omni_genotypes_and_intensities. There were originally 2,141 individuals in the data, from 21 populations worldwide. After filtering for genotype missingness (15%) and relatives using an IBS threshold (IBS2 + 0.5 * IBS1 > population median value + 0.01, with IBS2 being the proportion of SNPs where both alleles are identical by state between the two individuals, and IBS1 being the proportion of SNPs where only one allele is identical), 1,664 individuals remained. SNPs with genotype missingness exceeding 10% were removed and we filtered the SNPs based on a HWE test within the Peruvian samples, with a threshold of 0.001. Consequently, only the SNPs intersecting the *large* dataset were retained, to allow merging between the SAC samples and the 1000 Genomes Project data. The *combined* dataset included a total of 2,023,892 SNPs.
2) **Ancestral State**. We computed the ancestral allele using the Chimpanzee reference genome (panTro3 from UCSC Genome Browser), the Gorilla reference genome (gorGor3 from UCSC Genome Browser), the Bonobo genome (Prufer et al. 2012) and the Denisova genome (Meyer et al. 2012). For each SNP in the *large* dataset, the ancestral allele was identified as the most often observed allele in the outgroup genomes. When both alleles were seen in the same number of outgroups, the ancestral state was declared as unknown and the SNP rejected from the *large* dataset. The vast majority of SNPs had the same allele in all four outgroups.
3) **Genetic Map**. A genetic map was found from the HapMap project at: http://hapmap.ncbi.nlm.nih.gov/downloads/recombination/2011-01_phaseII_B37/. To lower the stochastic effect of a finite sample on the genetic map, and still retaining local variation of the recombination rate, we smoothed the values of the original genetic map by computing local recombination rates using linear regression on sliding 1 Mb windows around each position in the map. Based on this smoothed genetic map, we computed the genetic position of each SNP in the *large* dataset, using piecewise linear functions.
4) **Phasing**. Phasing was done separately on each dataset (*large* and *combined*) using the software Hapi-Ur v1.01 (Williams et al. 2012). As recommended by the developers, we launched three independent phasing runs for each dataset and computed the consensus phase from the three runs using vote-phase (program released together with Hapi-Ur). The window size used was 190 markers in the two types of datasets.
5) **iHS Computations**. All iHS computations were conducted using the R package rehh (Gautier and Vitalis 2012). The values for *limehh* and *limehhs* in the function *scan_hh* were both set to 0.02.
6) **LSBL Scans.** To scan for regions of increased genetic differentiation in the SAC sample, we used the LSBL (Shriver et al. 2004) on the *combined* dataset. The three groups considered were SAC, Peruvians (PEL), and Colombians (CLM). LSBL is a statistic based on pairwise F_{ST} values between samples and measures the population specific genetic differentiation at each position. Here, we estimated F_{ST} values using Hudson's estimator, as defined in (Bhatia et al. 2013).

Selection Coefficient

We used equation (2) from Ohta and Kimura (1975), which links the selection coefficient to current allele frequency, allele frequency at onset of selection and number of generations

since onset of selection. We used the top SNP associated with %MMA (rs486955) in the region of *AS3MT* and assumed a generation time of 25 years. Because onset of selection is likely to have occurred between 11,000 and 7,000 years ago, we looked at the range of selection coefficients obtained between those two curves (supplementary fig. S13, Supplementary Material online). We assumed the effect of drift, since divergence to the Peruvian population, to have had little impact on the frequency of the beneficial allele in Peruvians (which is thus assumed to have evolved neutrally at that position in the genome), and took current PEL frequency as a proxy for the allele frequency at onset of selection.

Haplotype Correlations

To determine whether a specific haplotype was the target of selection (suggested by iHS scans as well as similar additive patterns in boxplots of phenotypes versus significantly associated SNPs), we extracted all SNPs with significant FDR-corrected *q*-values for MMA-based GWA scans. From these SNPs, we constructed a putative "MMA-based protective haplotype" by linking together the specific SNP alleles, significantly associated (FDR $q < 0.01$) with low %MMA during the GWA scan. The MMA protective haplotype stretched over 554 kb (97 SNPs). We used the phased dataset merged with comparative data to extract all phased haplotypes spanning these regions for all SAC and comparative populations. Since the comparative data were typed on a 2.5 Mchip, the haplotype for this region in the comparative dataset was reduced to 50 SNPs. Exact matches to the putative protective haplotype were counted in the SAC and all comparative populations. For haplotypes that did not match the protective haplotype, differences to the protective haplotype were counted. The differences of individual haplotypes to the "protective haplotype" were visualized as violin plots for the SAC and all comparative populations. An exact match to the putative "protective haplotype" is seen at 0 on the *y*-axis. These differences of individual haplotypes to the protective haplotype were also plotted against MMA urinary fractions for the SAC population and a straight line was fitted to the scatterplot. The correlation coefficients (Kendall's Tau) between differences of individual haplotypes to protective haplotype versus phenotypes in SAC, were also calculated and are shown as insets on the scatterplots. Furthermore, to visualize individual haplotypes and their differences to the protective haplotype the haplotypes were visualized for SAC and three comparative populations (Peruvian PEL, Iberian IBS, and African American ASW) using the putative protective haplotype as reference. The phased haplotypes of individuals were sorted within each population according to their differences to the putative protective reference haplotype, which was printed on top of the haplotype plot.

Supplementary Material

Supplementary tables S1–S3 and figures S1–S17 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank participants who donated blood and urine samples. C.M.S. thank R.M. Nelson for useful discussions and suggestions. Genotyping was performed by the SNP&SEQ Technology Platform in Uppsala, Sweden (www.genotyping.se). The platform is part of Science for Life Laboratory at Uppsala University and supported as a national infrastructure by the Swedish Research Council. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2012222. This work was supported by grants from the Swedish Council for Working Life and Social Research to K.B., the Karolinska Institutet M.V., Eric Philip Sörensen's Foundation to K.B., Kungliga Fysiografiska Sällskapet to K.B., the EU within the Sixth Framework Programme for RTD ("PHIME" contract no FOOD-CT-2006-016253 [K.B. and M.V.]), the Wenner-Gren Foundations and the Swedish Research Council (C.M.S.), the Swedish Research Council Formas to M.J., and the Swedish Research Council for Science to M.J.

References

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.

Ahmed S, Ahsan KB, Kippler M, Mily A, Wagatsuma Y, Hoque AM, Ngom PT, El Arifeen S, Raqib R, Vahter M. 2012. In utero arsenic exposure is associated with impaired thymic function in newborns possibly via oxidative stress and apoptosis. *Toxicol Sci.* 129:305–314.

Ahmed S, Mahabbat-e Khoda S, Rekha RS, Gardner RM, Ameer SS, Moore S, Ekstrom EC, Vahter M, Raqib R. 2011. Arsenic-associated oxidative stress, inflammation, and immune disruption in human placenta and cord blood. *Environ Health Perspect.* 119:258–264.

Ahsan H, Chen Y, Kibriya MG, Slavkovich V, Parvez F, Jasmine F, Gamble MV, Graziano JH. 2007. Arsenic metabolism, genetic susceptibility, and risk of premalignant skin lesions in Bangladesh. *Cancer Epidemiol Biomarkers Prev.* 16:1270–1278.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.

Ameur A, Enroth S, Johansson A, Zaboli G, Igl W, Johansson AC, Rivas MA, Daly MJ, Schmitz G, Hicks AA, et al. 2012. Genetic adaptation of fatty-acid metabolism: a human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. *Am J Hum Genet.* 90:809–820.

Arriaza B, Amarasiriwardena D, Cornejo L, Standen V, Byrne S, Bartkus L, Bandak B. 2010. Exploring chronic arsenic poisoning in pre-Columbian Chilean mummies. *J Archaeol Sci.* 37:1274–1278.

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23: 1294–1296.

Basu Mallick C, Iliescu FM, Mols M, Hill S, Tamang R, Chaubey G, Goto R, Ho SY, Gallego Romero I, Crivellaro F, et al. 2013. The light skin allele of *SLC24A5* in South Asians and Europeans shares identity by descent. *PLoS Genet.* 9:e1003912.

Bertocchini F, Ovitt CE, Conti A, Barone V, Scholer HR, Bottinelli R, Reggiani C, Sorrentino V. 1997. Requirement for the ryanodine receptor type 3 for efficient contraction in neonatal skeletal muscles. *Embo J.* 16:6956–6963.

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* 23: 1514–1521.

Broberg K, Engström K, Ameer S. 2015. Gene-environment for metals. In: Nordberg GF, Fowler B, Nordberg M, editors. Handbook on the toxicology of metals. Oxford: Academic Press. p. 242–247.

Bundschuh J, Pérez-Carrera A, Litter MI. 2008. Distribución del arsénico en las regiones Ibérica e Iberoamericana. Editorial Programa Iberoamericano de Ciencia y Tecnologia para el Desarrollo, Buenos Aires, Argentina..

Bundschuh J, Litter MI, Parvez F, Roman-Ross G, Nicolli HB, Jean JS, Liu CW, Lopez D, Armienta MA, Guilherme LR, et al. 2012. One century of arsenic exposure in Latin America: a review of history and occurrence from 14 countries. *Sci Total Environ.* 429:2–35.

Chen B, Arnold LL, Cohen SM, Thomas DJ, Le XC. 2011. Mouse arsenic (+3 oxidation state) methyltransferase genotype affects metabolism and tissue dosimetry of arsenicals after arsenite administration in drinking water. *Toxicol Sci.* 124:320–326.

Chen H, Slatkin M. 2013. Inferring selection intensity and allele age from multilocus haplotype structure. *G3* 3:1429–1442.

Chung CJ, Hsueh YM, Bai CH, Huang YK, Huang YL, Yang MH, Chen CJ. 2009. Polymorphisms in arsenic metabolism genes, urinary arsenic methylation profile and cancer. *Cancer Causes Control* 20: 1653–1661.

Concha G, Nermell B, Vahter M. 2006. Spatial and temporal variations in arsenic exposure via drinking-water in northern Argentina. *J Health Popul Nutr.* 24:317–326.

Cooke CA, Abbott MB. 2008. A paleolimnological perspective on industrial-era metal pollution in the central Andes, Peru. *Sci Total Environ.* 393:262–272.

Drobná Z, Waters SB, Devesa V, Harmon AW, Thomas DJ, Stýblo M. 2005. Metabolism and toxicity of arsenic in human urothelial cells expressing rat arsenic (+3 oxidation state)-methyltransferase. *Toxicol Appl Pharmacol.* 207:147–159.

Drobna Z, Xing W, Thomas DJ, Stýblo M. 2006. shRNA silencing of AS3MT expression minimizes arsenic methylation capacity of HepG2 cells. *Chem Res Toxicol.* 19:894–898.

Drobna Z, Naranmandura H, Kubachka KM, Edwards BC, Herbin-Davis K, Styblo M, Le XC, Creed JT, Maeda N, Hughes MF, et al. 2009. Disruption of the arsenic (+3 oxidation state) methyltransferase gene in the mouse alters the phenotype for methylation of arsenic and affects distribution and retention of orally administered arsenate. *Chem Res Toxicol.* 22:1713–1720.

Engstrom K, Vahter M, Mlakar SJ, 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.

Engstrom KS, Hossain MB, 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.

Farzan SF, Korrick S, Li Z, Enelow R, Gandolfi AJ, Madan J, Nadeau K, Karagas MR. 2013. In utero arsenic exposure and infant infection in a United States cohort: a prospective study. *Environ Res.* 126:24–30.

Florea AM, Splettstoesser F, Busselberg D. 2007. Arsenic trioxide (As2O3) induced calcium signals and cytotoxicity in two human cell lines: SY-5Y neuroblastoma and 293 embryonic kidney (HEK). *Toxicol Appl Pharmacol.* 220:292–301.

Fry RC, Navasumrit P, Valiathan C, Svensson JP, Hogan BJ, Luo M, Bhattacharya S, Kandjanapa K, Soontararuks S, Nookabkaew S, et al. 2007. Activation of inflammation/NF-kappaB signaling in infants born to arsenic-exposed mothers. *PLoS Genet.* 3:e207.

Fujihara J, Fujii Y, Agusa T, Kunito T, Yasuda T, Moritani T, Takeshita H. 2009. Ethnic differences in five intronic polymorphisms associated with arsenic metabolism within human arsenic (+3 oxidation state) methyltransferase (AS3MT) gene. *Toxicol Appl Pharmacol.* 234: 41–46.

Gardner RM, Kippler M, Tofail F, Bottai M, Hamadani J, Grander M, Nermell B, Palm B, Rasmussen KM, Vahter M. 2013. Environmental exposure to metals and children's growth to age 5 years: a prospective cohort study. *Am J Epidemiol.* 177:1356–1367.

Gardner RM, Nermell B, Kippler M, Grander M, Li L, Ekstrom EC, Rahman A, Lonnerdal B, Hoque AM, Vahter M. 2010. Arsenic methylation efficiency increases during the first trimester of pregnancy independent of folate status. *Reprod Toxicol.* 31: 210–218.

Gautier M, Vitalis R. 2012. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28:1176–1177.

Gomez-Rubio P, Klimentidis YC, Cantu-Soto E, Meza-Montenegro MM, Billheimer D, Lu Z, Chen Z, Klimecki WT. 2012. Indigenous American ancestry is associated with arsenic methylation efficiency in an admixed population of northwest Mexico. *J Toxicol Environ Health A.* 75:36–49.

Gomez-Rubio P, Meza-Montenegro MM, Cantu-Soto E, Klimecki WT. 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.

Hanikenne M, Talke IN, Haydon MJ, Lanz C, Nolte A, Motte P, Kroymann J, Weigel D, Kramer U. 2008. Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of *HMA4*. *Nature* 453:391–395.

Hopenhayn-Rich C, Biggs ML, Smith AH, Kalman DA, Moore LE. 1996. Methylation study of a population environmentally exposed to arsenic in drinking water. *Environ Health Perspect.* 104:620–628.

IARC. 2012. A review of human carcinogens: arsenic, metals, fibres, and dusts, IARC monographs on the evaluation of carcinogenic risks to humans. Vol. 100C. Lyon, France: World Health Organization.

Idrovo AJ. 2005. [Possible health effects associated with pre-Columbian metallurgy]. *Biomedica* 25:295–303; discussion 304.

Karlsson EK, Harris JB, Tabrizi S, Rahman A, Shlyakhter I, Patterson N, O'Dushlaine C, Schaffner SF, Gupta S, Chowdhury F, et al. 2013. Natural selection in a Bangladeshi population from the cholera-endemic Ganges river delta. *Sci Transl Med.* 5:192ra186.

Kozul CD, Ely KH, Enelow RI, Hamilton JW. 2009. Low-dose arsenic compromises the immune response to influenza A infection in vivo. *Environ Health Perspect.* 117:1441–1447.

Kwiatkowski DP. 2005. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet.* 77:171–192.

Lin S, Shi Q, Nix FB, Styblo M, Beck MA, Herbin-Davis KM, Hall LL, Simeonsson JB, Thomas DJ. 2002. A novel S-adenosyl-L-methionine:arsenic (III) methyltransferase from rat liver cytosol. *J Biol Chem.* 277:10795–10803.

Lindberg AL, Rahman M, Persson LA, Vahter M. 2008. 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.

Maull EA, Ahsan H, Edwards J, Longnecker MP, Navas-Acien A, Pi J, Silbergeld EK, Styblo M, Tseng CH, Thayer KA, et al. 2012. Evaluation of the association between arsenic and diabetes: a National Toxicology Program workshop review. *Environ Health Perspect.* 120:1658–1670.

Melnick JG, Parkin G. 2007. Cleaving mercury-alkyl bonds: a functional model for mercury detoxification by MerB. *Science* 317:225–227.

Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prufer K, de Filippo C, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338: 222–226.

Meza MM, Yu L, Rodriguez YY, Guild M, Thompson D, Gandolfi AJ, Klimecki WT. 2005. Developmentally restricted genetic determinants of human arsenic metabolism: association between urinary methylated arsenic and *CYT19* polymorphisms in children. *Environ Health Perspect.* 113:775–781.

Moon K, Guallar E, Navas-Acien A. 2012. Arsenic exposure and cardiovascular disease: an updated systematic review. *Curr Atheroscler Rep.* 14:542–555.

Normando Cruz E. 2011. Historia y Etnicidad en las Yungas de la Argentina. Venezuela: Ediciones Polar.

Núñez ALN, Hill HG, Martinez AL. 1991. Guía Museo Arqueológico, Universidad Católica del Norte Chile. San Pedro de Atacama (II Regíon): Instituto de Investigaciones Arqueológicas y Museo RP.

Gustavo Le Paige SJ. Antofagasta, Chile: Universidad Católica del Norte Chile.
Ohta T, Kimura M. 1975. The effect of selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect). *Genet Res.* 25: 313–326.
Parvez F, Chen Y, Yunus M, Olopade C, Segers S, Slavkovich V, Argos M, Hasan R, Ahmed A, Islam T, et al. 2013. Arsenic exposure and impaired lung function. Findings from a large population-based prospective cohort study. *Am J Respir Crit Care Med.* 188:813–819.
Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2:e190.
Perez CF, Lopez JR, Allen PD. 2005. Expression levels of RyR1 and RyR3 control resting free Ca2+ in skeletal muscle. *Am J Physiol Cell Physiol.* 288:C640–C649.
Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat Genet.* 39: 1256–1260.
Pierce BL, Kibriya MG, Tong L, Jasmine F, Argos M, Roy S, Paul-Brutus R, Rahaman R, Rakibuz-Zaman M, Parvez F, 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.
Pierce BL, Tong L, Argos M, Gao J, Farzana J, Roy S, Paul-Brutus R, Rahaman R, Rakibuz-Zaman M, Parvez F, et al. 2013. Arsenic metabolism efficiency has a causal role in arsenic toxicity: Mendelian randomization and gene-environment interaction. *Int J Epidemiol.* 42:1862–1871.
Prufer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G, Kodira C, Winer R, et al. 2012. The bonobo genome compared with the chimpanzee and human genomes. *Nature* 486:527–531.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81:559–575.
Rahman A, Persson LA, Nermell B, El Arifeen S, Ekstrom EC, Smith AH, Vahter M. 2010. Arsenic exposure and risk of spontaneous abortion, stillbirth, and infant mortality. *Epidemiology* 21:797–804.
Rahman A, Vahter M, Ekstrom EC, Persson LA. 2010. Arsenic exposure in pregnancy increases the risk of lower respiratory tract infection and diarrhea during infancy in Bangladesh. *Environ Health Perspect.* 119: 719–724.
Rahman A, Vahter M, Ekstrom EC, Persson LA. 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.
Sabbagh A, Marin J, Veyssiere C, Lecompte E, Boukouvala S, Poloni ES, Darlu P, Crouau-Roy B. 2013. Rapid birth-and-death evolution of the xenobiotic metabolizing NAT gene family in vertebrates with evidence of adaptive selection. *BMC Evol Biol.* 13:62.
Schlawicke Engstrom K, Broberg K, Concha G, Nermell B, Warholm M, Vahter M. 2007. Genetic polymorphisms influencing arsenic metabolism: evidence from Argentina. *Environ Health Perspect.* 115: 599–605.
Schlebusch CM, Lewis CM Jr., Vahter M, Engstrom K, Tito RY, Obregon-Tito AJ, Huerta D, Polo SI, Medina AC, Brutsaert TD, et al. 2013. Possible positive selection for an arsenic-protective haplotype in humans. *Environ Health Perspect.* 121:53–58.
Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW. 2004. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum Genomics* 1:274–286.
Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, et al. 2010. Genetic evidence for high-altitude adaptation in Tibet. *Science* 329:72–75.
Smith AH, Marshall G, Yuan Y, Ferreccio C, Liaw J, von Ehrenstein O, Steinmaus C, Bates MN, Selvin S. 2006. Increased mortality from lung cancer and bronchiectasis in young adults after exposure to arsenic in utero and in early childhood. *Environ Health Perspect.* 114: 1293–1296.
Song WY, Park J, Mendoza-Cozatl DG, Suter-Grotemeyer M, Shim D, Hortensteiner S, Geisler M, Weder B, Rea PA, Rentsch D, et al. 2010. Arsenic tolerance in Arabidopsis is mediated by two ABCC-type phytochelatin transporters. *Proc Natl Acad Sci U S A.* 107: 21187–21192.
Takei Y, Swietlik M, Tanoue A, Tsujimoto G, Kouzarides T, Laskey R. 2001. MCM3AP, a novel acetyltransferase that acetylates replication protein MCM3. *EMBO Rep.* 2:119–123.
Tellez-Plaza M, Gribble MO, Voruganti VS, Francesconi KA, Goessler W, Umans JG, Silbergeld EK, Guallar E, Franceschini N, North KE, et al. 2013. Heritability and preliminary genome-wide linkage analysis of arsenic metabolites in urine. *Environ Health Perspect.* 121:345–351.
Thoma R, Schulz-Gasch T, D'Arcy B, Benz J, Aebi J, Dehmlow H, Hennig M, Stihle M, Ruf A. 2004. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* 432:118–122.
Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet.* 39:31–40.
Vahter M. 2002. Mechanisms of arsenic biotransformation. *Toxicology* 181–182:211–217.
Vahter M, Concha G, Nermell B, Nilsson R, Dulout F, Natarajan AT. 1995. A unique metabolism of inorganic arsenic in native Andean women. *Eur J Pharmacol.* 293:455–462.
Van Den Bergh K, Du Laing G, Montoya JC, De Deckere E, Tack FM. 2010. Arsenic in drinking water wells on the Bolivian high plain: field monitoring and effect of salinity on removal efficiency of iron-oxides-containing filters. *J Environ Sci Health A Tox Hazard Subst Environ Eng.* 45:1741–1749.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.
Williams AL, Patterson N, Glessner J, Hakonarson H, Reich D. 2012. Phasing of many thousands of genotyped samples. *Am J Hum Genet.* 91:238–251.
Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 42:565–569.