the **genetics**society

REVIEW ARTICLE

Recent population genomic insights into the genetic basis of arsenic tolerance in humans: the difficulties of identifying positively selected loci in strongly bottlenecked populations

Mario Apata[1] · **Susanne P. Pfeifer** [1]

Abstract

Recent advances in genomics have enabled researchers to shed light on the evolutionary processes driving human adaptation, by revealing the genetic architectures underlying traits ranging from lactase persistence, to skin pigmentation, to hypoxic response, to arsenic tolerance. Complicating the identification of targets of positive selection in modern human populations is their complex demographic history, characterized by population bottlenecks and expansions, population structure, migration, and admixture. In particular, founder effects and recent strong population size reductions, such as those experienced by the indigenous peoples of the Americas, have severe impacts on genetic variation that can lead to the accumulation of large allele frequency differences between populations due to genetic drift rather than natural selection. While distinguishing the effects of demographic history from selection remains challenging, neglecting neutral processes can lead to the incorrect identification of candidate loci. We here review the recent population genomic insights into the genetic basis of arsenic tolerance in Andean populations, and utilize this example to highlight both the difficulties pertaining to the identification of local adaptations in strongly bottlenecked populations, as well as the importance of controlling for demographic history in selection scans.

Introduction

Arsenic, a naturally occurring element, is acutely toxic to humans. Although low levels of environmental exposure to arsenic are generally considered safe, epidemiological studies indicate that exposure to elevated levels represents an important global public health issue, impacting tens of millions of individuals annually, particularly in developing areas (Naujokas et al. 2013). As arsenic exposure in human populations is the result of both natural and anthropogenic causes, the temporal extent of exposure differs greatly between global populations. Capitalizing on these differences, recent studies have utilized population genomic approaches to begin to unravel the genetic underpinnings of arsenic tolerance in several populations. For example, particular Andean populations show evidence of exposure to arsenic-contaminated drinking water spanning the past 7000 years, and genetic changes have recently been correlated with an increased arsenic metabolization efficiency in these populations (Schlebusch et al. 2013; Apata et al. 2017). Yet, considerable challenges remain to meaningfully connect the measured phenotypic variance observed between individuals and populations, with causative genotypic variants, and ultimately fitness. We here review these recent results, contextualize them more broadly within the field of population genomics, and outline avenues for future research.

A brief overview of global arsenic exposure and population-level phenotypic variation in humans

Arsenic laden environments pose a serious public health issue, with current epidemiological crises in several countries along the Pacific Coast from South-East Asia to the Americas (Nordstrom 2002; Hughes et al. 2011). In these regions, over 90 million people are exposed to arsenic-contaminated water stemming from rivers, wells, and groundwater (Naujokas et al. 2013), with arsenic exposure

✉ Susanne P. Pfeifer
susanne.pfeifer@asu.edu

1   Center for Evolution & Medicine, School of Life Sciences,
    Arizona State University, Tempe, AZ 85821, USA

ranging from 300 to 1000 µg/l (Fig. 1)—far surpassing 10 µg/l (i.e., the level considered safe by the WHO (2011)). This exposure owes to largely natural causes—with alluvial sediments, organic-rich or black shales, thermal springs, and volcanogenic sediments resulting in the dissolution of arsenic-bearing minerals (Nordstrom 2002; Fendorf et al. 2010; López et al. 2012). In addition, anthropogenic factors —for example, mining, mineral extraction, and agriculture (e.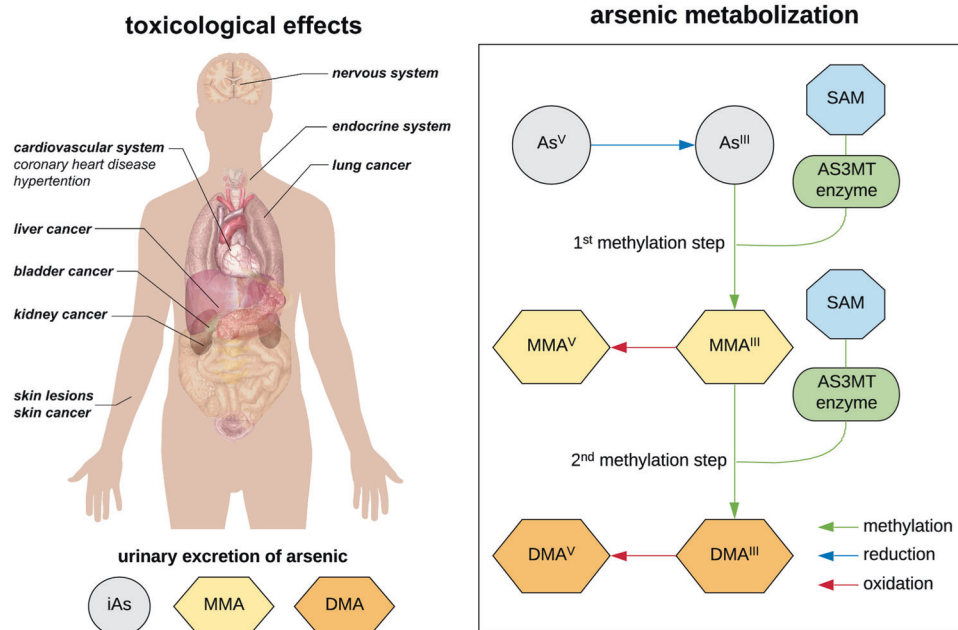g., poultry and swine feed additives and pesticides), can contribute to an enrichment of arsenic in the groundwater (Nordstrom 2002). Chronic arsenic exposure has long been observed to result in genotoxic and carcinogenic effects, increasing morbidity owing to a high prevalence of several cancers in the skin, liver, and bladder (Argos et al. 2010) (Fig. 2). Furthermore, this element may pass through the placenta, increasing rates of spontaneous abortion, low birth weights, and several other conditions (Hopenhayn et al. 2003; Milton et al. 2017).

Fig. 1 Map of endemic areas with high arsenic levels in natural water sources around the globe (Arriaza et al. 2018), with regions indicated on the red-scale frequently having significant public health issues associated with exposure. The regions denoted in green boxes have evidence of increased tolerance, perhaps owing to long-term exposure, and thus the local populations are fruitful examples in which to study the evolutionary response to arsenic

Fig. 2 Toxicological effects on human health and the arsenic metabolization pathway. As described in the text, the relative proportion of monomethylarsonic acid (MMA) and dimethylarsinic acid (DMA) has arisen as a useful biomarker for predicting clinical outcomes. The figure is adapted from the 'human body diagram' image, released into the public domain by Mikael Häggström

From an evolutionary perspective, it is firstly interesting to note that the general levels of arsenic tolerance, and consequently suffered toxicological effects, appear to vary between exposed populations across the globe, though the underlying evolutionary causes of these differences are only now beginning to be elucidated. As these toxicological and epidemiological effects have been well reviewed in the literature (see Minatel et al. 2018), we here particularly focus on the recent insights pertaining to the evolutionary response to long-term exposure, and discuss population genetic insights pertaining to the identified mutations observed to be associated with an increased arsenic tolerance.

Beginning with a broad view, most mammals use methyltransferases to metabolize inorganic arsenic into dimethylarsinic acid (DMA) via the highly toxic intermediate metabolite monomethylarsonic acid (MMA) (Fig. 2), though the ability to do so differs strongly across species (Palmgren et al. 2017). Within humans specifically, it is helpful to distinguish between populations experiencing short- vs. long-term exposure. In terms of the former, populations in Bangladesh are thought to have been exposed for only a few decades (Argos et al. 2010), and display characteristic symptoms including skin lesions as well as skin cancer (Pierce et al. 2011). In the Americas, short-term exposure, and the related effects, have been described in populations across the US, Mexico, and Latin America (McClintock et al. 2012). One of the most well-studied examples comes from northern Chile, in the city of Antofagasta. This population was shown to suffer severe effects upon the change of their primary drinking water source to the toxic Toconce River (arsenic level: 860 µg/l). The rapid increase in several arsenic-related diseases in this population, even 40 years after high exposures ceased, continues to be the subject of study (Ferreccio and Sancha 2006; Steinmaus et al. 2014; Roh et al. 2018).

Unlike these recent exposure events, Andean populations in South America have experienced a long-term exposure to arsenic. The Andeans are composed of several different groups including the Aymara, Atacameño, Quechua, and Colla, who have been living along the highlands and the Atacama Desert region (i.e., southern Peru, Bolivia, northern Chile, and Argentina) for thousands of years. Bio-archeological evidence suggests that pre-Columbian populations were likely suffering arsenic poisoning in the Atacama Desert at least 7000 years ago (Arriaza et al. 2010, 2018). Interestingly, modern populations from these regions have demonstrated reduced sensitivity to arsenic (Hopenhayn-Rich et al. 1996; McClintock et al. 2012), with certain diagnostic signs of chronic arsenic exposure (hyperkeratosis, hyperpigmentation, and skin cancer) being only rarely observed in the Aymara and Atacamaño populations (Sancha et al. 1992; Smith et al. 2000; De Loma et al. 2019). Indeed, recent studies indicate that modern Andean populations from the Camarones Valley of Chile, an area where arsenic levels in the water reach in excess of 200 µg/l, are able to metabolize arsenic efficiently and more rapidly expel it from their bodies via urination (Schlebusch et al. 2013; Apata et al. 2017).

Before addressing the genotypic underpinnings of this trait, it is first worth clarifying the nature of the phenotype. The primary detoxification pathway in humans occurs in the liver, during which MMA and DMA metabolites are produced (Fig. 2; Lin et al. 2002; Dheeman et al. 2014), and subsequently varying proportions of MMA, DMA, and inorganic arsenic are excreted (Agusa et al. 2011). The relative proportions differ between individuals and populations, and, as this variation is related to arsenic toxicity, this represents a sort of detoxification biomarker. Indeed, higher percentages of MMA are associated with various adverse health effects (Chen et al. 2003; Ahsan et al. 2006; Chung et al. 2009; Pierce et al. 2011, 2012), while higher levels of DMA are correlated with reduced toxicity. As expected, short-term exposed populations have higher MMA production, while cohorts from populations with long-term exposure have shown higher levels of DMA, including indigenous populations from Argentina and the highlands of Bolivia and Chile (Engström et al. 2011; Muñoz et al. 2018; De Loma et al. 2019).

Connecting phenotype to genotype: searching for the genetic basis of efficient arsenic metabolization

Given the above phenotype, different genes encoding reductases and methyltransferases have been hypothesized to be the source of the inter-individual and inter-population variation in arsenic susceptibility (Fujihara et al. 2009, 2010; Agusa et al. 2011; Antonelli et al. 2014). For example, mutations in the genes *AS3MT* (arsenic [+3 oxidation state] methyltransferase), *GSTP1* (glutathione s-transferase pi 1), and *PNP* (purine nucleoside phosphorylase) have been found to alter metabolite profiles (Antonelli et al. 2014). Thereby, *AS3MT* has received particular attention owing to its key role in the catalysis of arsenic methylation (Sumi and Himeno 2012)—with a high methylation capacity being associated with higher levels of DMA and lower levels of MMA in urine (Gardener et al. 2011) (Fig. 2), and several mutations in this gene have been associated with related diseases (Antonelli et al. 2014).

Two polymorphic sites in particular have been strongly associated with efficient metabolization in modern populations exposed to arsenic (Agusa et al. 2011). The first is rs11191439 (Met287Thr), in which the ancestral T allele is associated with lower levels of the MMA metabolite. Previous studies in miners exposed to arsenic in Antofagasta have shown that carriers of the recessive C allele had higher

percentages of MMA (Fig. 3), and thus a high toxicity risk (Hernández et al. 2008a, 2008b), with similar results from exposed populations studied in central Europe (Lindberg et al. 2007) and Mexico (Gomez-Rubio et al. 2012). The C allele shows a worldwide frequency around 10–14% (Fujihara et al. 2008, 2010), though is found at considerably

lower frequency in populations experiencing long-term arsenic exposure (such as the 1% frequency observed in the Camarones) (Fig. 4). The second strongly associated variant is an intronic SNP, rs3740393 (C/G), which is at moderate frequency (35–40%) in short-term exposed populations (Fujihara et al. 2009, 2010), and higher frequency in long-

Fig. 3 The identified haplotype in the *AS3MT* gene, found to be associated with increased arsenic tolerance. The C-T-T-A haplotype is a specific case of the C-T-A haplotype identified by Schlebusch et al. (2013) which, in addition to the three SNPs rs3740393, rs3740390, and rs10748835, also includes SNP rs11191439. Note that humans not carrying the identified haplotype have a higher toxicity risk
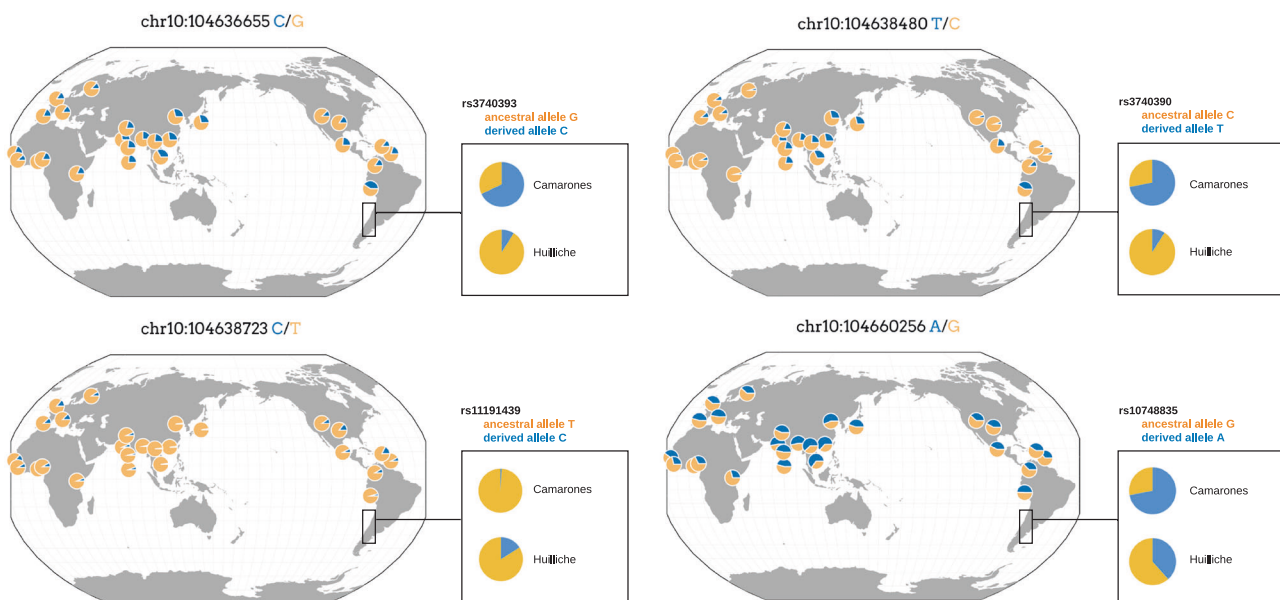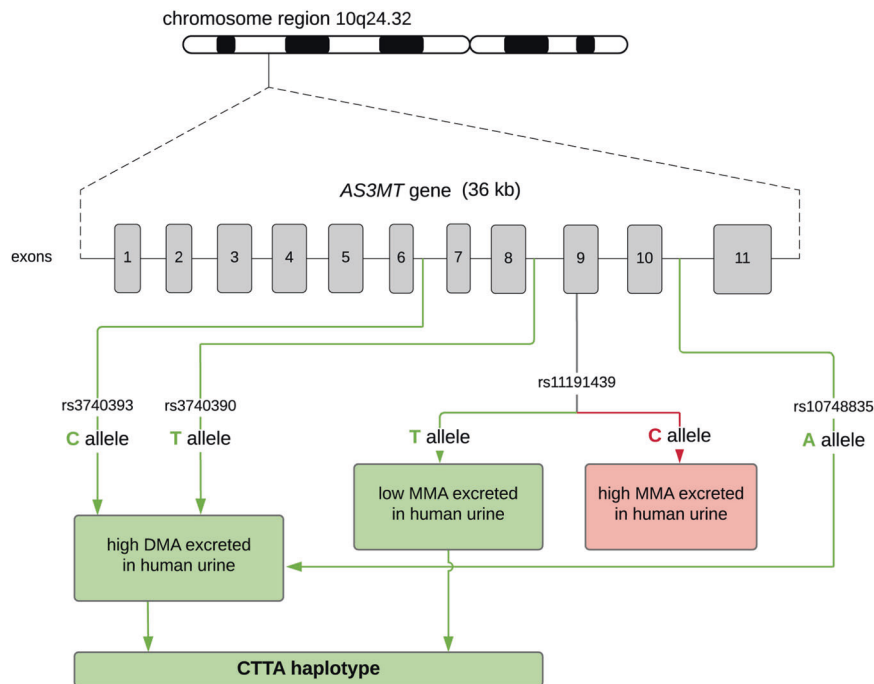

Fig. 4 Population-specific allele frequencies for the four SNPs in the C-T-T-A haplotype in the *AS3MT* gene, found to be associated with increased arsenic tolerance. Allele frequencies for the Chilean populations (i.e., Camarones from the Camarones Valley and Huilliche from southern Chile) were obtained from Apata et al. 2017; allele frequencies for all other populations were obtained from the 1000 Genomes dataset (1000 Genomes Project Consortium 2015) using the Geography of Genetic Variants Browser (Marcus and Novembre 2017). Sites represent a comparison of regions with high and low levels of arsenic in the water, with levels reaching in excess of 200 µg/l in the Camarones Valley but less than 10 µg/l in southern Chile

term exposed populations (e.g., 70% in the Camarones; Fig. 4; Engström et al. 2011; Apata et al. 2017). CC and CG genotypes have been shown to be associated with an increase in the production of the DMA metabolite, and to be in strong linkage disequilibrium (LD) with two other intronic SNPs, rs3740390 (C/T), and rs10748835 (A/G). A joint haplotype of these three, known as C-T-A, is associated with higher percentages of DMA (Schlebusch et al. 2013), and is found at elevated frequency in the Camarones population (Apata et al. 2017).

Finally, the genomic region 10q24.32 has been generally associated with observed phenotypic variation in arsenic metabolization, with genome-wide association studies in Bangladesh reporting mutations associated with DMA/MMA variation (Pierce et al. 2012), consistent with associated studies in Andean populations associating these regions with higher DMA percentages (Engström et al. 2013; Schlebusch et al. 2015). In addition, this region has been highlighted as containing a potential signature of selective sweeps (Eichstaedt et al. 2015). However, while these mutations in AS3MT are interesting candidates, one must be careful when equating correlation with causation, and the population genetics of this trait and these genotypes are in need of further investigation, as discussed below.

Putative signals of adaptation to arsenic: a case study in the Camarones Valley

Having colonized the Atacama Desert, one of the driest places in the world, around 9000 years ago, the first settlers faced environmental adversities ranging from high UV radiation to hypoxia to high arsenic concentrations in essentially all available drinking water. Yet, peoples from the pre-Columbian period to the present day have survived on this arsenic-contaminated water. The most exposed Andean population located in the Camarones Valley (arsenic level: 1000 µg/l) thus represents a unique natural laboratory to study the genetic underpinnings of any potential adaptation associated with arsenic-rich environments.

First considering bio-anthropological evidence, several studies have revealed severe arsenic concentrations in the inner organs, as well as visible skin lesions, in the remains of individuals from the ancient Chinchorro population—an Incan colony that was settled in the Camarones Valley (Arriaza et al. 2018). Arriaza (2005) posited that arsenic poisoning additionally increased rates of spontaneous abortion among the Chinchorro people, potentially initiating their characteristic artificial mummification practice, as an emotional and cultural response for coping with this loss. This hypothesis is partly based on the oldest ancient Chinchorro mummies (archeological sites of Cam-14 and Cam-17), corresponding predominantly to newborns and children

(Arriaza 2005). Moreover, arsenic measures taken from hair and bones have shown a tendency of decreasing average arsenic levels over time, starting from Chinchorro hunter-gatherer-fishers in the Archaic Period (7000–3000 BP) to later agro-pastoral pre-Hispanic populations (3000–500 BP), to the current populations living in Quebrada Camarones (Yáñez et al. 2005; Arriaza et al. 2010; Byrne et al. 2010; Bartkus et al. 2011; Swift et al. 2015). This has been interpreted as a potential evidence of an adaptive, temporally increasing efficacy of metabolic detoxification. However, it should be noted that further genetic study is required to, for example, directly link the Chinchorro population as the ancestors of the modern Camarones population.

Recently, a haplotype in the AS3MT gene (C-T-<u>T</u>-A—a specific case of the C-T-A haplotype previously identified by Schlebusch et al. (2013) which, in addition to the three SNPs rs3740393, rs3740390, and rs10748835, also includes rs11191439) has been identified in the Camarones population segregating at higher frequency (68%) than in other Amerindian populations with less arsenic exposure (the haplotype is found at 48% frequency in the Azapa Valley (arsenic levels: 10–20 µg/l), and 8% frequency in the Huilliche population (arsenic levels: <10 µg/l)) (Fig. 5). This observation, combined with the reduced frequency of the C risk allele in Met287Thr observed in long-term vs. short-term exposed populations (Camarones (1%) and Azapa (5%), vs. Hulliche (16%) and Antofagasta (14%)), has led to the suggestion that natural selection may be driving these observed frequency differences (Figs. 3–5).

However, while there indeed exist interesting phenotypic differences between long- and short-term exposed populations as described above, and though association studies have produced some promising candidate regions and mutations, considerable work remains in order to establish causal variants and to understand the evolutionary history of these populations. Without a proper population genetic framework accounting for the neutral processes shaping allele frequencies, including population size change, structure, and migration, inappropriate 'adaptive story-telling' remains a danger (Pavlidis et al. 2012; Jensen et al. 2019). In that regard, these small, admixed, highly bottlenecked indigenous populations present multiple challenges for distinguishing demographic and selective effects.

The importance of distinguishing between population genetic signals of demography and selection

Strong population bottlenecks, of the sort believed to have been experienced multiple times throughout the evolutionary history of these populations (Lindo et al. 2018), are notoriously difficult to distinguish from genetic hitchhiking
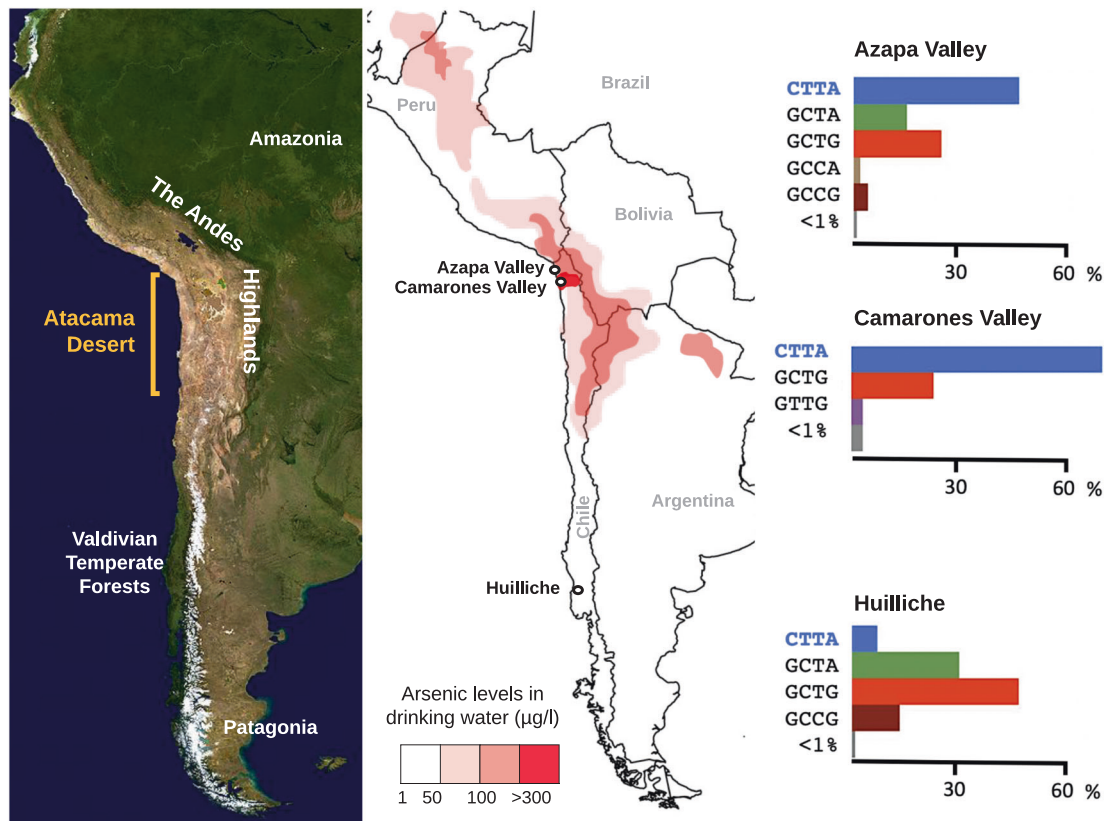
Fig. 5 Observed haplotype frequencies in the *AS3MT* gene between Chilean populations with differing levels of arsenic exposure. The haplotype identified via association studies to be associated with increased arsenic tolerance, C-T-T-A, is shown in blue letters in the vertical axes of the graphs

effects (Barton 1998; Thornton and Jensen 2007; and see the review of Bank et al. 2014). With regards to the particular methodologies thus far utilized in examining positive selection around the *AS3MT* locus in these Andean populations, Eichstaedt et al. (2015) relied on F_{ST} and PBS statistics (Yi et al. 2010), neglecting the demographic history and implementing a simple outlier approach. While common, such outlier approaches are fraught with error, owing to the a priori assumption that a set fraction of loci residing in the tail of a given statistical distribution must be owing to positive selection effects (i.e., neglecting the fact that any model, including neutrality, will have tails in the associated distribution). In addition, tenuous, the authors reported similarly significant F_{ST} values around this locus in populations without previous or current history of arsenic exposure. Moreover, in an earlier study, Eichstaedt et al. (2014) found that this genomic region was not significant with either the iHS or XP-EHH (Voight et al. 2006; Sabeti et al. 2007) approaches—suggesting a lack of evidence supporting the role hitchhiking effects modulating these allele frequencies (i.e., sweep-like patterns in LD do not appear to be present around this locus in these populations).

Similarly investigating putative patterns of positive selection around this locus, Schlebusch et al. (2015),

considering population structure but not the history of population size change, found only weak support using the iHS statistic, and, like Eichstaedt et al. (2015), stronger evidence using a branch-length statistic (in this case, the LSBL statistics (Shriver et al. 2004)). Regardless of this mixed evidence, the performance of sweep-detection statistics under nonequilibrium demographic models is known to be severely comprised, often characterized by low power and high false-positive rates under a wide range of demographic scenarios (e.g., Jensen et al. 2005; Teshima et al. 2006; Excoffier et al. 2009; Bierne et al. 2011; Harris et al. 2018; and see the review of Crisci et al. 2012). Evaluating the performance of many of these statistics specifically, Crisci et al. 2013 found that both frequency spectrum and LD based approaches have false-positive rates often in considerable excess of true-positive rates, under strongly bottlenecked population histories such as these.

Conclusions and future directions

Though results obtained without accounting for neutral processes are troubling in their implications regarding the ability to identify positively selected loci associated with

arsenic exposure (or any other trait) in these populations, there is an upside. There exist increasingly sophisticated approaches for inferring demographic histories (e.g., Gutenkunst et al. 2009; Excoffier et al. 2013), which utilize high-quality, genome-wide data of the sort now available for these populations. While the resulting neutral demographic model may prove difficult to differentiate from selective effects as described above, the utilization of this model can at least greatly reduce mis-inference, and allow for the incorporation of the types of environmental measures that are available pertaining to arsenic exposure (see reviews of Haasl and Payseur 2016; Jensen et al. 2016). Fortunately, recent efforts have already begun to characterize the population history of this region, inferring population split times between low- and high-elevation populations in the Andes, as well as the severity of the population size reduction associated with European contact (Lindo et al. 2018).

As demographic inference continues to accumulate for these Andean populations, we propose a multistep process that can improve evolutionary inference in these populations in general, and with regards to the genotypic variants underlying phenotypic traits such as increased arsenic tolerance in particular. Having fit a demographic model, it is next of importance to assess the fit of that model to the genomic data. Namely, it is straightforward to simulate the parameters of the estimated model in order to ensure that it sufficiently replicates patterns of variation and LD observed in the natural population, using, for example, simulation software such as SLiM (Haller and Messer 2019). If the model is found to be a sufficient fit, one may then simulate selective sweeps of varying strengths and ages within the context of that demographic history. In such a way, by applying the planned statistical analyses (be it iHS, XP-EHH, F_{ST}, or any other) to these simulated replicates, one may quantify the performance of that statistic (i.e., the true- and false-positive rates) under a population history actually relevant for the empirical application. With this information in hand, and if the results indeed suggest an ability to accurately detect selection for the population in question, genomic scan approaches may proceed. Relatedly, in the example of arsenic tolerance, given mixed evidence of selection on the candidate loci identified via association studies to date, as well as their unusual intermediate frequencies observed in both arsenic-exposed and nonexposed populations (Fig. 4; Eichstaedt et al. 2015), it appears worth also explicitly considering the potential of polygenic adaptation—a model which may not result in classic selective sweep signatures, and indeed may be simply characterized by subtle changes in allele frequency (Jain and Stephan 2017). However, differentiating these subtle frequency changes from those resulting from the underlying demographic history is expected to be even more challenging than for classic selective sweep models.

In closing, other systems in which genotype, phenotype, and fitness have been successfully connected may serve as useful models for future analyses pertaining to arsenic tolerance in humans. One of the best-known examples in mammals pertains to the evolution of cryptic coloration (see review of Harris et al. (2019)). As with the phenotypic observation of different levels of tolerance corresponding with the underlying levels of arsenic exposure in humans, in mouse populations the correspondence between coat coloration and soil color has long been noted (Dice 1947). In this instance, the correlation is likely owing to avian predation, in which cryptically colored individuals are more likely to avoid visual detection. Also, as in the arsenic example, large-scale association studies were conducted in order to identify genotypic variants that correlate with the phenotype in question (Linnen et al. 2009, 2013). As such, the cryptic coloration literature may serve as an informative roadmap outlining the avenues of future research. In that literature, genome-wide polymorphism data from light (derived-state) and dark (ancestral-state) mouse populations were next utilized to fit a demographic history including population size change, structure, and migration, and simulation was used to assess the fit of the model to the data and to characterize sweep-detection performance under the relevant demographic history (Pfeifer et al. 2018). Utilizing this model as the appropriate null distribution for characterizing the expected performance of summary statistics under neutrality compared to selection, well-supported candidate loci were identified. It is important to note, that certain mouse populations were found to have demographic histories characterized by bottlenecks of such severity that genomic scans for selection were simply not feasible, owing to the anticipated true- and false-positive rates (Poh et al. 2014). In populations for which genomic scans were applicable, statistically identified mutations were tested and functionally validated in *Mus*, thereby completing the link between genotype and phenotype using this mix of population genomics, association mapping, and functional assays (Barrett et al. 2019).

Thus, we propose that the way forward for the study of arsenic tolerance will involve more extensive demographic modeling, large-scale simulation studies to identify statistics that may be useful for identifying patterns associated with genetic hitchhiking/polygenic adaptation within the context of the inferred population history, and finally an examination of the correspondence between large-scale association studies and genomic scans in order to identify top candidate loci. With regards to the feasibility of functional studies to similarly validate candidates, it is noteworthy that tractable animal models indeed exist for the study of the *AS3MT* locus, with previous studies demonstrating its importance for arsenic

methylation in rats (Lin et al. 2002) as well as in a mouse knockout (Drobna et al. 2009; Chen et al. 2011)—thereby providing potential study systems for the examination of existing and yet-to-be-identified candidate mutations.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1000 Genomes Project Consortium (2015) A global reference for human genetic variation. Nature 526:68–74

Agusa T, Fujihara J, Takeshita H, Iwata H (2011) Individual variations in inorganic arsenic metabolism associated with AS3MT genetic polymorphisms. Int J Mol Sci 12(4):2351–2382

Ahsan H, Chen Y, Parvez F, Argos M, Hussain AI, Momotaj H et al. (2006) Health effects of arsenic longitudinal study (HEALS): description of a multidisciplinary epidemiologic investigation. J Expo Sci Environ Epiemiol 16:191–205

Antonelli R, Shao K, Thomas DJ, Sams R, Cowden J (2014) AS3MT, GSTO, and PNP polymorphisms: impact on arsenic methylation and implications for disease susceptibility. Environ Res 132:156–167

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

Argos M, Kalra T, Rathouz PJ, Chen Y, Pierce B, Parvez F et al. (2010) Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in Bangladesh (HEALS): a prospective cohort study. Lancet 376:252–258

Arriaza BT (2005) Arseniasis as an environmental hypothetical explanation for the origin of the oldest artificial mummification practice in the world. Chungara, Rev De Antropoĺa Chil 37:255–260

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

Arriaza B, Amarasiriwardena D, Standen V, Yáñez J, Van Hoesen J, Figueroa L (2018) Living in poisoning environments: Invisible risks and human adaptation. Evol Anthropol 27:188–196

Bank C, Foll M, Ferrer-Admetlla A, Ewing G, Jensen JD (2014) Thinking too positive? Revisiting current methods in population genetic selection inference. Trends Genet 30:540–546

Barrett RDH, Laurent S, Mallarino R, Pfeifer SP, Xu CC, Foll M et al. (2019) Linking a mutation to survival in wild mice. Science 363:499–504

Barton NH (1998) The effect of hitch-hiking on neutral genealogies. Genet Res Camb 72:123–133

Bartkus L, Amarasiriwardena D, Arriaza B, Bellis D, Yáñez J (2011) Exploring lead exposure in ancient Chilean mummies using a single strand of hair by laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS). Microchem J 98:267–274

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans fail to map local adaptation genes. Mol Ecol 20:2044–2072

Byrne S, Amarasiriwardena D, Bandak B, Bartkus L, Kane J, Jones J et al. (2010) Were Chinchorro exposed to arsenic? Arsenic determination in Chinchorro mummies' hair by laser ablation inductively coupled plasma-mass spectrometry (LA-ICP-MS). Microchem J 94:28–35

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 by drinking water. Toxicol Sci 124:320–326

Chen GQ, Zhou L, Styblo M, Walton F, Jing Y, Weinberg R et al. (2003) Methylated metabolites of arsenic trioxide are more potent than arsenic trioxide as apoptotic but not differentiation inducers in leukemia and lymphoma cells. Cancer Res 63:1853–1859

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

Crisci JL, Poh Y, Bean A, Simkin A, Jensen JD (2012) Recent progress in polymorphism-based population genetic inference. J Hered 103:287–296

Crisci J, Poh Y-P, Mahajan S, Jensen JD (2013) The impact of equilibrium assumptions on tests of selection. Front Genet 4:235

De Loma J, Tirado N, Ascui F, Levi M, Vahter M, Broberg K et al. (2019) Elevated arsenic exposure and efficient arsenic metabolism in indigenous women around Lake Poopó, Bolivia. Sci Total Environ 657:179–186

Dheeman DS, Packianathan C, Pillai JK, Rosen BP (2014) Pathway of human AS3MT arsenic methylation. Chem Res Toxicol 27:1979–1989

Dice LR (1947) Effectiveness of selection by owls of deer mice (Peromyscus maniculatus) which contrast in color with their background. Contrib Lab Vertebrate Biol Univ Mich 34:1–20

Drobna Z, Naranmandura H, Kubachka KM, Edwards BC, Herbin-Davis K, Styblo M 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

Eichstaedt CA, Antao T, Pagani L, Cardona A, Kivisild T, Mormina M (2014) The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. PLoS One 9:e93314

Eichstaedt CA, 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

Engström K, Vahter M, Mlakar SJ, Concha G, Nermell B, Raquib R et al. (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

Engström K, Hossain MB, Lauss M, Ahmed S, Raqib R, Vahter M et al. (2013) Efficient arsenic metabolism—the AS3MT haplotype is associated with DNA methylation and expression of multiple genes around AS3MT. PLoS ONE 8:e53732

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. Heredity 103:285–298

Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. PLoS Genet 9:e1003905

Fendorf S, Michael HA, van Geen A (2010) Spatial and temporal variations of groundwater arsenic in south and southeast Asia. Science 328:1123–1127

Ferreccio C, Sancha AM (2006) Arsenic exposure and its impact on health in Chile. J Health Popul Nutr 24:164–175

Fujihara J, Soejima M, Koda Y, Kunito T, Takeshita H (2008) Asian specific low mutation frequencies of the M287T polymorphism in the human arsenic (+3 oxidation state) methyltransferase (AS3MT) gene. Mutat Res 654:158–161

Fujihara J, Fujii Y, Agusa T, Kunito T, Yasuda T, Moritani T et al. (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

Fujihara J, Soejima M, Yasuda T, Koda Y, Agusa T, Kunito T et al. (2010) Global analysis of genetic variation in human arsenic (+3 oxidation state) methyltransferase (AS3MT). Toxicol Appl Pharmacol 243:292–299

Gardener RM, Nermell B, Kippler M, Grander M, Li L, Ekström EC et al. (2011) Arsenic methylation efficiency increases during the first trimester of pregnancy independent of folate status. Reprod Toxicol 31:210–218

Gomez-Rubio P, Klimentidis YC, Cantu-Soto E, Meza-Montenegro MM, Billheimer D, Lu Z et al. (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

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet 5:e1000695

Haasl RJ, Payseur BA (2016) Fifteen years of genome-wide scans for selection: trends, lessons and unaddressed genetic sources of complication. Mol Ecol 25:5–23

Haller BC, Messer PW (2019) SLiM 3: Forward genetic simulations beyond the Wright-Fisher model. Mol Biol Evol 36:632

Harris RB, Sackman A, Jensen JD (2018) On the unfounded enthusiasm for soft selective sweeps II: examining recent evidence from humans, flies, and viruses. PLoS Genet 14:e1007859

Harris RB, Irwin KK, Jones MR, Laurent S, Barrett RDH, Nachman MW, et al. (2019) The population genetics of crypsis in vertebrates. Heredity. https://doi.org/10.1038/s41437-019-0257-4

Hernández A, Xamena N, Sekaran C, Tokunaga H, Sampayo-Reyes A, Quinteros D et al. (2008) High arsenic metabolic efficiency in AS3MT287Thr allele carriers. Pharmacogenet Genomics 18:349–355

Hernández A, Xamena N, Surrallés J, Sekaran C, Tokunaga H, Quinteros D et al. (2008) Role of the Met(287)Thr polymorphism in the AS3MT gene on the metabolic arsenic profile. Mutat Res 637:80–92

Hopenhayn C, Ferreccio C, Browning SR, Huang B, Peralta C, Gibb H et al. (2003) Arsenic exposure from drinking water and birth weight. Epidemiology 14:593–602

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 104:9

Hughes MF, Beck BD, Chen Y, Lewis AS, Thomas DJ (2011) Arsenic exposure and toxicology: a historical perspective. Toxicol Sci 123:305–332

Jain K, Stephan W (2017) Modes of rapid polygenic adaptation. Mol Biol Evol 34:3169–3175

Jensen JD, Kim Y, Bauer DuMont V, Aquadro CF, Bustamante CD (2005) Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics 170:1401–1410

Jensen JD, Foll M, Bernatchez L (2016) Introduction: the past, present, and future of genomic scans for selection. Mol Ecol 25:1–4

Jensen JD, Payseur BA, Stephan W, Aquadro CF, Lynch M, Charlesworth D et al. (2019) The importance of the Neutral Theory in 1968 and 50 years on: a response to Kern & Hahn 2018. Evolution 73:111–114

Lin S, Shi Q, Nix FB, Styblo M, Beck MA, Herbin-Davis KM et al. (2002) A novel S-adenosyl-l-methionine: arsenic(III) methyltransferase from rat liver cytosol. J Biol Chem 277:10795–10803

Lindberg AL, Kumar R, Goessler W, Thirumaran R, Gurzau E, Koppova K et al. (2007) Metabolism of low-dose inorganic arsenic in a central European population: Influence of sex and genetic polymorphisms. Environ Health Perspect 115:1081–1086

Lindo J, Haas R, Hofman C, Apata M, Moraga M, Verdugo RA et al. (2018) The genetic prehistory of the Andean highlands 7000 years BP though European contact. Sci Adv 4:eaau4921

Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an adaptive allele in Peromyscus mice. Science 325:1095–1098

Linnen CR, Poh Y-P, Peterson BK, Barrett RDH, Larson JG, Jensen JD et al. (2013) Adaptive evolution of multiple traits through multiple mutations at a single gene. Science 339:1312–1316

López DL, Bundschuh J, Birkle P, Armienta MA, Cumbal L, Sracek O et al. (2012) Arsenic in volcanic geothermal fluids of Latin America. Sci Total Environ 429:57–75

Marcus JH, Novembre J (2017) Visualizing the geography of genetic variants. Bioinformatics 33:594–595

McClintock TR, Chen Y, Bundschuh J, Oliver JT, Navoni J, Olmos V et al. (2012) Arsenic exposure in Latin America: biomarkers, risk assessments and related health effects. Sci Total Environ 429:76–91

Milton A, Hussain S, Akter S, Rahman M, Mouly T, Mitchell K (2017) A review of the effects of chronic arsenic exposure on adverse pregnancy outcomes. Int J Environ Res Public Health 14:556

Minatel BC, Sage AP, Anderson C, Hubaux R, Marshall EA, Lam WL et al. (2018) Environmental arsenic exposurE: From genetic susceptibility to pathogenesis. Environ Int 112:183–197

Muñoz M, Valdés M, Muñoz-Quezada M, Lucero B, Rubilar P, Pino P et al. (2018) Urinary inorganic arsenic concentration and gestational diabetes mellitus in pregnant women from Arica, Chile. Int J Environ Res Public Health 15:1418

Naujokas MF, Anderson B, Ahsan H, Aposhian HV, Graziano JH, Thompson C et al. (2013) The broad scope of health effects from chronic arsenic exposure: update on a worldwide public health problem. Environ Health Perspect 121:295–302

Nordstrom DK (2002) Worldwide occurrences of arsenic in ground water. Science 296:2143–2145

Palmgren M, Engström K, Hallström BM, Wahlberg K, Søndergaard DA, Säll T et al. (2017) AS3MT-mediated tolerance to arsenic evolved by multiple independent horizontal gene transfers from bacteria to eukaryotes. PLoS ONE 12:e0175422

Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012) A critical assessment of story-telling: go categories and the importance of validating genomic scans. Mol Biol Evol 29:3237–3248

Pfeifer SP, Laurent S, Sousa VC, Linnen CR, Foll M, Excoffier L et al. (2018) The evolutionary history of Nebraska deer mice: local adaptation in the face of strong gene flow. Mol Biol Evol 35:792–806

Pierce BL, Argos M, Chen Y, Melkonian S, Parvez F, Islam T et al. (2011) Arsenic exposure, dietary patterns, and skin lesion risk in Bangladesh: a prospective study. Am J Epidemiol 173:345–354

Pierce BL, Kibriya MG, Tong L, Jasmine F, Argos M, Roy S 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

Poh Y-P, Domingues V, Hoekstra HE, Jensen JD (2014) On the prospect of identifying adaptive loci in recently bottlenecked populations: a case study in beach mice. PLoS ONE 9:e110579

Roh T, Steinmaus C, Marshall G, Ferreccio C, Liaw J, Smith AH (2018) Age at exposure to arsenic in water and mortality 30–40 years after exposure cessation. Am J Epidemiol 187:2297–2305

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C et al. (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918

Sancha AM, Vega F, Venturino H, Fuentes S, Salazar AM, Moreno V et al. (1992). The arsenic health problem in northern Chile evaluation and control. A case study preliminary report. In: Proceedings of the international seminar. Arsenic in the Environment and its incidence on health. Universidad to Chile, Santiago, Chile, pp 187–202

Schlebusch CM, Lewis Jr CM, Vahter M, Engström K, Tito RY, Obregón-Tito AJ et al. (2013) Possible positive selection for an arsenic-protective haplotype in humans. Environ Health Perspect 12:53–58

Schlebusch CM, Gattepaille LM, Engström K, Vahter M, Jakobsson M, Broberg K (2015) Human adaptation to arsenic-rich environments. Mol Biol Evol 32:1544–1555

Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J et al. (2004) The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Hum Genomics 1:274–286

Smith AH, Lingas EO, Rahman M (2000) Contamination of drinking water by arsenic in Bangladesh: a public health emergency. Bull World Health Organ 78:1093–1103

Steinmaus C, Ferreccio C, Acevedo J, Yuan Y, Liaw J, Durán V et al. (2014) Increased lung and bladder cancer incidence in adults after in utero and early-life arsenic exposure. Cancer Epidemiol Biomark Prev 23:1529–1538

Sumi D, Himeno S (2012) Role of arsenic (+3 oxidation state) methyltransferase in arsenic metabolism and toxicity. Biol Pharm Bull 35:1870–1875

Swift J, Cupper ML, Greig A, Westaway MC, Carter C, Santoro CM et al. (2015) Skeletal arsenic of the pre-Columbian population of Caleta Vitor, northern Chile. J Archaeol Sci 58:31–45

Teshima KM, Coop G, Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps? Genome Res 16:702–712

Thornton KR, Jensen JD (2007) Controlling the false-positive rate in multi-locus genome scans for selection. Genetics 175:737–750

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4:e72

World Health Organization (2011) Guidelines for drinking-water quality, 4th edn. WHO, Gutenberg, Malta, http://www.who.int/water_sanitation_health/publications/2011/9789241548151_toc.pdf

Yáñez J, Fierro V, Mansilla H, Figueroa L, Cornejo L, Barnes RM (2005) Arsenic speciation in human hair: a new perspective for epidemiological assessment in chronic arsenicism. J Environ Monit 7:1335–1341

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329:75–78