# Does arsenic contamination affect DNA methylation patterns in a wild bird population? An experimental approach

Veronika N. Laine[1], Mark Verschuuren[1], Kees van Oers[1], Silvia Espín[2,3], Pablo Sánchez-Virosta[2,3], Tapio Eeva[3], Suvi Ruuskanen[3]

[1]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands

[2]Area of Toxicology, Department of Socio-Sanitary Sciences, University of Murcia, Spain

[3]Department of Biology, University of Turku, Finland

Abstract

Pollutants, like toxic metals, negatively influence organismal health and performance, even leading to population collapses. Studies in model organisms have shown that epigenetic marks, such as DNA methylation, can be modulated by various environmental factors, including pollutants, influencing gene expression and various organismal traits. Yet experimental data on the effects of pollution on DNA methylation from wild animal populations is largely lacking. We here investigated experimentally for the first time the effects of early-life exposure to environmentally relevant levels of a key pollutant, arsenic (As) on genome-wide DNA methylation status in a wild bird population. We experimentally exposed nestlings of great tits (*Parus major*) to arsenic during their post-natal developmental period (3 to 14 days post-hatching) and compared their DNA methylation levels to those of respective controls. In contrast to predictions, we found no overall hypomethylation in the arsenic group. We found evidence for loci to be differentially methylated between the treatment groups, but for five CpG sites only. This may be explained by the heterogenous

environment, relatively low dose (compared to previous laboratory studies), or choice of tissue. Further studies are needed to evaluate whether epigenetic dysregulation is a commonly observed phenomena in polluted populations, and what are its consequences for organism functioning and for populations.

Keywords: pollution, *Parus major*, environmental epigenetics, ecological epigenetics, ecotoxicology

Running title: Arsenic pollution and DNA methylation in the wild

## Introduction

Environmental pollution can negatively affect organisms at multiple level of organization, from molecular and physiological level to performance, and even lead to population collapses (Dudgeon et al., 2006; McCrink-Goode, 2014; Saaristo et al., 2018; Vos et al., 2000). In wild populations, a largely unexplored mechanism mediating such pollution effects is the potential influence of the epigenome, such as DNA methylation. In human and animal models, it has been discovered that methylation patterns can be changed by various environmental factors, including metal and organic pollutants and other early-life stressors (reviewed by Baccarelli & Bollati, 2009; Cheng, Choudhuri, & Muldoon-Jacobs, 2012; Head, 2014; Head, Dolinoy, & Basu, 2012; Ray, Yosim, & Fry, 2014; Ruiz-Hernandez et al., 2015; Sharavanan et al.). DNA methylation is the addition of a methyl (-CH3) group to the 5' carbon site of cytosines catalyzed by DNA-methyltransferases, and is generally found to be negatively associated with gene expression (Goldberg, Allis, & Bernstein, 2007). Variation in DNA methylation is linked to variation in phenotypes and behavior, and associated with the prevalence of various diseases, including cancers in humans and model animals (Rosenfeld, 2010; Salameh, Bejaoui, & El Hajj, 2020; Skinner, Manikkam, & Guerrero-Bosagna, 2010). Epigenetic changes from early-life environment may persist and affect health throughout life-time and may even be transmitted to future generations (Horsthemke, 2018), which could potentially contribute to explaining delayed or persistent effects of pollutants (e.g. Head 2014). Yet the effects of pollutants on the epigenome have hardly been explored, and epigenetic research in wild animal populations is only emerging (Bossdorf, Richards, & Pigliucci, 2008; Head et al., 2012; Laine et al., 2016; Liebl, Schrey, Richards, & Martin, 2013; Riyahi, Sanchez-Delgado, Calafell, Monk, & Senar, 2015; Rubenstein et al., 2016; Sepers et al., 2019; Verhoeven, Vonholdt, & Sork, 2016; Wenzel & Piertney, 2014)

Arsenic (As) is a global, persistent pollutant, distributed in the environment due to natural and anthropogenic sources such as mining, industrial activities or coal combustion (Mandal & Suzuki, 2002) and the most highly ranked hazardous substances for animals and plants (ATSDR, 2019). Across organisms, arsenic can have negative consequences for basically all organ systems, often via causing oxidative stress, i.e. the imbalance between harmful reactive oxygen species (ROS) and antioxidant defenses, and cancer (Jomova et al., 2011; Sánchez-Virosta, Espín, García-Fernández, & Eeva, 2015).

Arsenic has been repeatedly observed to also modulate patterns of DNA methylation *in vitro*, in laboratory animal models (with large doses) and in studies on human populations (Alegria-Torres et al., 2016; Bjorklund, Aaseth, Chirumbolo, Urbina, & Uddin, 2018; Jensen, Novak, Wnek, Gandolfi, & Futscher, 2009). The major pathway for arsenic to change the DNA methylation of a cytosine, is via the depletion of the cellular availability of methyl groups, since biotransformation of arsenic to less toxic forms includes the addition of methyl group(s) (Minatel et al., 2018). The main methyl donor during the methylation of both arsenic and DNA is s-adenosylmethionine (SAM). The high demand imposed on this molecule during the biotransformation process can then lead to global DNA hypomethylation, as shown in multiple biomedical studies (Baccarelli & Bollati, 2009; Hubaux et al., 2013). Additionally, ROS created during arsenic biotransformation have been suggested to influence DNA methylation by creating aberrant modifications (Ziech, Franco, Pappa, & Panayiotidis, 2011). Pre/postnatal exposure to arsenic is associated with epigenetic modifications related to early onset of diseases, which could have long-term consequences (reviewed in Bantock, Prys-Jones, & Lee, 2008; Gangopadhyay, Sharma, Chauhan, & Srivastava, 2019). For example, prenatal arsenic exposure led to global hypomethylation of inflammatory and tumor suppressor genes (Kile et al., 2014) and interfered with de novo methylation (Broberg et al., 2014) in humans. Despite the extensive data on model animals and humans, the potential

effects of environmental arsenic on wild animals via epigenetic dysregulation has not been studied up to date.

We here investigated the effects of experimental early-life (post-natal) exposure to arsenic on genome-wide DNA methylation status in a wild population of great tits (*Parus major*). To our knowledge, this is the first study on (i) environmentally relevant arsenic exposure, and (ii) in a wild population, on epigenetic markers. We used a bird model, since birds have been successfully used in biomonitoring of pollution and its effects (e.g. García-Fernández, 2014). Arsenic exposure has been reported to negatively affect multiple fitness-related traits (growth, physiology, behavior and even egg-laying) in several birds species (reviewed in Sánchez-Virosta et al., 2015). For great tits specifically, we have previously reported (using the same experimental protocol) that arsenic exposure increased nestling mortality, reduced wing growth (Sánchez-Virosta et al., 2018) and decreased an intracellular antioxidant, catalase, in nestlings (Sánchez-Virosta et al., 2020). More specifically, we here experimentally exposed nestlings in non-polluted sites to environmentally relevant levels (1 µg/g body mass) of dietary arsenic during the entire post-hatching growth period, and compared their DNA methylation levels to respective controls. We used reduced representation bisulfite sequencing (RRBS) to assess genome-wide methylation and characterized differential methylation across CpG sites between the experimental and the control group. We predict that arsenic exposure will lead to genome-wide hypomethylation, and potentially specifically on gene/hubs related to development.

## Methods

### *Arsenic treatment protocol and sampling*

The study was conducted in a nest box population of great tits (*Parus major)*, a small passerine bird and a popular model species in ecological and evolutionary research. Importantly, as one of the only non-domesticated bird species, both the genome and methylome are available (Derks et al., 2016; Laine et al., 2016; Verhulst et al., 2016).

The arsenic exposure, dosages and sampling are described in detail in Sánchez-Virosta et al. (2020; 2018). In short, the experiment was conducted in a nest-box population in western Finland, with known history of very low pollution levels (Eeva, Ryömä, & Riihimaki, 2005). Breeding was monitored, and from day 3 after hatching until day 14 whole broods were subjected to daily oral dosing with the following treatments: arsenic treatment (1 µg As/g bodymass, N = 16 broods) or control treatment (distilled water, N = 16 broods). The dose aimed to represent environmentally relevant exposure levels occurring in polluted areas in Europe: It was estimated combining data from several sources, such as (i) the lowest-observed-adverse-effect level for different effects on mammals, (ii) fecal arsenic levels reported for great tits at some metal polluted sites (reviewed in Sánchez-Virosta et al., 2015), (iii) arsenic concentrations of food items (moth larvae, spiders and beetles) collected directly from parent great tits feeding their nestlings in the polluted area (Sánchez-Virosta et al., 2018, 2020), and (iv) a pilot experiment Sánchez-Virosta et al., 2018). Fecal matter was sampled 8 days after hatching for metal analyses (see below). DNA methylation was analyzed from red blood cells (RBCs) collected 14 days after hatching to avoid sacrificing the individuals. Ten samples from the arsenic and ten from control treatment were randomly selected for the DNA methylation analyses. These included five females and five males from each treatment (molecularly sexed, following Bantock et al., 2008). Only one nestling per nest was selected to avoid pseudoreplication. The experiment was conducted under licenses from the Animal Experiment Committee of the State Provincial Office of Southern Finland (license number ESAVI/11579/04.10.07/2014) and the Centre for Economic Development,

Transport and the Environment, ELY Centre Southwest Finland (license number VARELY/593/2015).

***Metal analyses***

For detailed analyses, see (Sánchez-Virosta et al., 2018). Briefly, in both experimental groups, two samples (one male and one female) from the same brood were combined to assess brood level metal exposure (total N = 32 broods). We determined the concentrations of As, but also other metals to confirm that the levels of other metals were low and similar across the treatment groups (see Sánchez-Virosta et al., 2018). The determination of pollutants was conducted with ICP-OES with detection limit of 1 ppt (ng/l) and below. Calibration standards and certified reference materials were used for method validation. The levels of other measured metals (Al, Pb, Ni, Zn, Mg, Mn, Fe, Cu) were low, and did not differ among the treatment groups (all $t < 0.88$, all $p < 0.38$).

***DNA isolation***

DNA isolation was performed at the Center of Evolutionary Applications (University of Turku, Finland). We used RBCs given that previous studies suggest that blood shows similar methylation patterns as brain tissue in the study species (e.g. 80% similarity between brain and blood methylation in CpGs Derks et al., 2016; Verhulst et al., 2016). DNA was extracted from 10-20 µl RBCs using the salt extraction method modified from Aljnabi & Martinez (1997). Extracted DNA was treated with RNase-I according to the manufacturer's protocol. DNA concentration was measured fluorometrically with a Qubit High Sensitivity kit (ThermoFisher Scientific) and we assessed DNA integrity by running each DNA sample on an agarose gel.

***RRBS library preparation***

We used a reduced representation bisulfite sequencing (RRBS) approach, which enriches for regions of the genome that have a high CpG content. It was previously shown in the study species that a vast majority of methylated Cs (97%) were derived from CpG sites in blood (Derks et al., 2016). Sequencing was conducted at the Finnish Microarray and Sequencing Center in Turku, Finland. The library preparation was started from 200 ng of genomic DNA and was carried out according to a protocol adapted from Boyle et al. (2012). The first step in the workflow involved the fragmentation of genomic DNA with MspI where the cutting pattern of the enzyme (C^CGG) was used to systematically digest DNA to enrich for CpG dinucleotides. After a fragmentation step a single reaction was carried out to end repair and A-tail (required for the adapter ligation) the MspI digested fragments using Klenow fragment (3’ => 5’ exo) following the purification of A-tailed DNA with bead SPRI clean-up method (AMPure magnetic beads). A unique Illumina TruSeq indexing adapter was then ligated to each sample during adapter ligation step to be able to identify pooled samples of one flow cell lane. To reduce the occurrence of adapter dimers, a lower concentration of adapters (1:10 dilution) was used than recommended by the manufacturer. These ligated DNA fragments were purified with bead SPRI clean-up method before putting samples through bisulfite conversion to achieve C-to-U conversion of unmethylated cytosines, whereas methylated cytosines remain intact. Bisulfite conversion and sample purification were done according to Invitrogen MethylCode Bisulfite Conversion Kit. Aliquots of converted DNA were amplified by PCR (16 cycles) with Taq/Pfu Turbo Cx Polymerase, a proofreading PCR enzyme that does not stall when it encounters uracil, the product of the bisulfite reaction, in the template. PCR-amplified RRBS libraries were purified using two subsequent rounds of SPRI beadclean-ups to minimize primer dimers in the final libraries. The high quality of the libraries was confirmed with Advanced Analytical Fragment Analyzer and the concentrations

of the libraries were quantified with Qubit® Fluorometric Quantitation, Life Technologies. We used an average fragment size of 250-350 bp for sequencing.

***Sequencing***

The samples were normalized and pooled for the automated cluster preparation which was carried out with Illumina cBot station. The 20 libraries were combined in two pools, 10 samples in each pool (treatments and sexes equally distributed between the pools) and sequenced in two lanes. The samples were sequenced with an Illumina HiSeq 2500 instrument using TruSeq v3 sequencing chemistry. Paired-end sequencing with 2 x 100 bp read length was used with 6 bp index run.

***Sequence data processing and differential methylation expression analysis***

All the reads were checked for quality using FastQC (Babraham Bioinformatics) with multiQC (Ewels, Magnusson, Lundin, & Kaller, 2016), and low-quality sequences were trimmed with Trim Galore v. 0.4.4 (Brabraham Bioinformatics) by using --quality 20 --paired --rrbs settings.

The trimmed reads were mapped to the *Parus major* reference genome build 1.1. (https://www.ncbi.nlm.nih.gov/assembly/ GCF_001522545.2) using Bismark (Krueger & Andrews, 2011) with default parameters. Methylation calling was conducted with Bismark, first with default settings with paired-end mode and overlap removal (--p --no_overlap). After this first calling round, we observed a methylation bias for the samples by plotting the methylation proportion across each possible position in the read. Based on the plotting, the three and two first bases of R1 and R2 respectively of the 5’ prime end were omitted and the first base in the R2 3’ prime end was also omitted in the final methylation calling. Thereafter, Methylkit (Akalin et al., 2012) implemented in R was used for filtering and differential

methylation analysis. We discarded bases that had coverage below 10x. To avoid a possible PCR bias we also discarded bases that had more than 99.9th percentile of coverage in each sample. Before differential methylation analysis we merged read counts from reads covering both strands of a CpG dinucleotide and CpGs needed to be covered with at least 8 samples per group (control and treatment).

Samples were thereafter clustered based on the similarity of their overall methylation profile by (i) using the clustering method ward.D in Methylkit's clusterSamples -function and (ii) using principal component analysis (PCA) with Methylkit's PCASamples -function. We also checked for lane and sex effect by using Methylkit's assocComp -function where it checks which principal components are statistically associated with the potential batch effects such as the used lane and sex of the individuals. For the former, no missing data is allowed, thus we created a separate data object where all the individuals needed to be covered.

For analyzing differential methylation of GpC sites between control and arsenic treatment we used the beta-binomial model from DSS package (Feng, Conneely, & Wu, 2014) which is also included in Methylkit (calculateDiffMethDSS - function). DSS calculates the differential methylation statistics using a beta-binomial model with parameter shrinkage. Bonferroni correction was applied to account for multiple testing.

## Results

### *Arsenic exposure*

As reported in Sánchez -Virosta et al. (2020; 2018), dietary arsenic treatment successfully increased arsenic load as fecal arsenic levels were on average 10 times higher in arsenic exposure compared to control group (average±SD ppm: control 0.51±0.50, As exposure 4.92±4.57, $t_{15.4} = -3.83$, $p = 0.0015$), and increased levels were also found in internal tissues.

***Sequencing and mapping***

The total number of read pairs was 341 million (Supplementary Table 1), varying from 14 million to 20 million per individual. After QC filtering the final number of read pairs was 337 million (Supplementary Table 1). The RRBS individual sequencing data have been deposited in NCBI (Number will be added later). Mapping efficiency was on average 46.15% and on average 3.1 million cytosines were covered before 10x coverage and percentile filtering. After filtering, 1.3 million cytosines were identified in CpG context. When combining the Cs from both strands and restricting our data to at least 8 individuals per group to be covered, we ended up having 652 655 CpGs.

***Sample clustering and differential methylation***

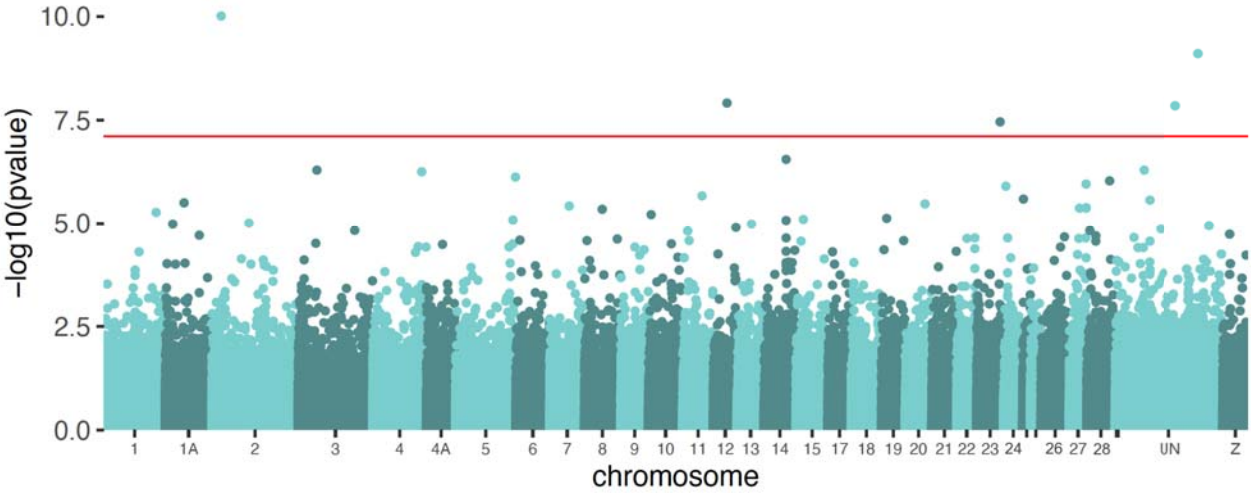
Both the ward.D and PCA clustering methods showed that sample 14 was an outlier in its methylation profile (Supplementary Figure 1). That particular sample also had a low number of reads and showed lower duplication levels (Supplementary Table 1) and we therefore decided to exclude this sample from further analysis. No lane effect was detected, but PC3 was associated with sex after Bonferroni correction (Supplementary Table 2, Supplementary Figure 2), mostly driven by two samples, ctrl_3F and test_16F, since after removing these two female samples from the data, PC3 was not significant anymore. Furthermore, when removing the PC3 from the data, three CpG sites were significant in the differential methylation analysis done with DSS: two of them were the same as when including all the PCs (see below, Table 1). The three other significant sites found below were not covered by all individuals as required in this PC-removal analysis.

In the differential methylation analysis when including all the PCs, five CpG sites showed a significant difference in methylation level with a q-value below 0.05 and percent methylation

difference larger than 10% (Table 1, Figure 1, Supplementary Table 3). Lambda estimation was close to 1 ($\lambda$ = 0.747, SE 0.000136) (Supplementary Figure 3), suggesting no systematic biases ($\lambda$>1 indicates bias). Four of these sites were hypermethylated (higher methylation in the arsenic treatment group) and one was hypomethylated (higher methylation in the control group). Three of the sites were located in gene bodies, namely zinc finger and BTB domain containing 47 (*ZBTB47*), HIVEP zinc finger 3 (*HIVEP3*) and insulin like growth factor 2 mRNA binding protein 1 (*IGF2BP1*) based on NCBI *P. major* annotation report 102.

**Table 1.** The differentially methylated CpG sites between arsenic exposed and control individuals. Methylation diff% refers to the methylation difference, comparing arsenic exposed to control group. Positive values therefore indicate hypermethylation in the arsenic treatment group compared to the control group. PC3 indicates sites that were significant after PC3 removal.

| Chr | Chr Genbank | Position | P-value | q-value | Methylation diff % | Gene | PC3 |
|---|---|---|---|---|---|---|---|
| 2 | NC_031769.1 | 2,448,788 | 1.00E-10 | 6.53E-05 | 36.40 | *ZBTB47* | |
| 12 | NC_031781.1 | 9,949,364 | 1.23E-08 | 8.05E-03 | 57.66 | - | |
| 23 | NC_031791.1 | 5,408,232 | 3.43E-08 | 2.24E-02 | 42.66 | *HIVEP3* | x |
| UN | NW_015379267.1 | 107,660 | 7.99E-10 | 5.21E-04 | 60.43 | *IGF2BP1* | |
| UN | NW_015379318.1 | 39,910 | 1.42E-08 | 9.27E-03 | -52.66 | - | x |
| **Only in PC3 removed** | | | | | | | |
| 14 | NC_031783.1 | 14,025,702 | 1.34E-07 | 0.042 | 27.09 | - | x |


10.0
7.5
5.0
2.5
0.0
−log10(pvalue)
1
1A
2
3
4
4A
5
6
7
8
9
10
11
12
13
14
15
17
18
19
20
21
22
23
24
26
27
28
UN
Z
chromosome

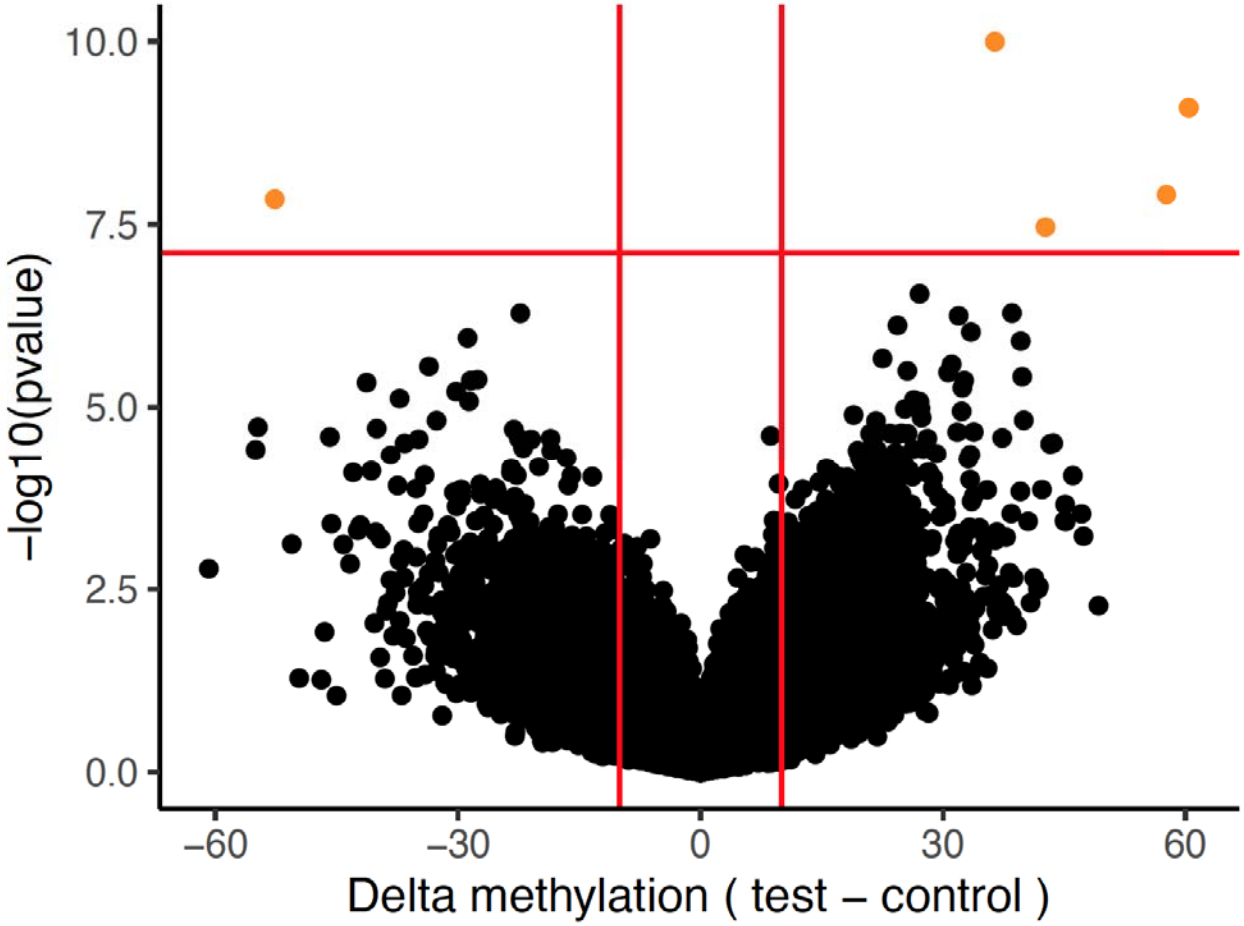

10.0
7.5
5.0
2.5
0.0
−log10(pvalue)
−60
−30
0
30
60
Delta methylation ( test − control )

**Figure 1.** Plots of significance of CpG sites from the differential methylation analysis conducted with DSS implemented in Methylkit. (a) Manhattan plot with the significance of differential methylation of the arsenic treatment against the control pool, against the great tit reference genome version 1.1. The orange line depicts the genome-wide threshold based on a Bonferroni correction: 7.11. (b) A volcano plot of the significance against the absolute difference in methylation between the two pools, with delta methylation is arsenic treatment – control. Orange points are the genome-wide significant sites after Bonferroni correction and filtering for Delta methylation >10%.

## Discussion

We studied for the first time whether early-life exposure to environmentally relevant levels of experimental arsenic affects DNA methylation in a wild vertebrate population. The experimental treatment increased arsenic levels significantly, but contrary to predictions, did not lead to overall hypomethylation. We did find that treated individuals showed hypermethylation in four and hypomethylation in one CpG site, indicating that increased levels of arsenic exposure may affect methylation at specific parts of the genome only. Yet also at these sites, the assumption of general hypomethylation was not met.

The lack of overall or site-specific hypomethylation may be explained by various factors: first, contrary to our predictions, the methyl donor s-adenosylmethionine needed for methylation may not have been limiting, potentially because oxidative status was not altered dramatically in all individuals. Indeed, as reported from the exact same experiment and samples by Sánchez-Virosta et al. (2020), most biomarkers of oxidative status and damage in blood were only slightly (but not statistically significantly) elevated, and only antioxidant enzyme catalase showed significant decrease. In the future, sampling before and after exposure to e.g. pollutants may be advised to associate DNA methylation changes directly to

changes in oxidative status, for example in adult birds (in contrast to developing animals where measurements are confounded by the changes in growth and associated changes in physiology).

Second, the response is likely to depend on the tissue type studied. For example, global hypomethylation in response to arsenic exposure is not consistently reported in blood: in humans, where blood leucocytes have been used to characterize arsenic associated changes no evidence for global hypo or hypermethylation was detected, yet arsenic was repeatedly reported to induce hypermethylation in various genes (especially promoters) (Koestler, Avissar-Whiting, Houseman, Karagas, & Marsit, 2013), whereas global hypomethylation was detected in hepatic cells (Chen et al., 2001). Given that arsenic metabolism and SAM production mostly takes part in liver, we may expect tissue-dependent hypomethylation especially in liver, but not necessarily in other tissues. Unfortunately, we lack oxidative status measurements from the liver in this experiment. Thus measuring methylation levels from red blood cells might not be ideal in this case. Although studies have shown correlating methylation values between different tissue (e.g. between absolute blood methylation levels in blood and absolute methylation levels in liver, kidney and brain  Derks et al., 2016; McKay et al., 2011), tissue-specific methylation patterns were more prominent for genes that are expressed in a tissue-specific way (Derks et al., 2016).

Furthermore, contrary to many previous studies in laboratory animals, this experiment was conducted with relatively low doses, mimicking exposure in polluted environments, whereas effects via SAM may only be apparent when levels are higher. Finally as the experiment was conducted in a wild population, in comparison to previous studies in laboratory, the environmental variability and potential variability across sexes may have masked some effects of the experimental treatments. The effects of (low) pollution may further be evident only if there are other concurrent stressors.

We could annotate three of the five differentially methylated sites to genes. One of the genes, *IGF2BP1* is especially interesting as it is associated with development and growth: it has been showed that *IGF2BP1* plays important roles in various aspects of cell function, such as cell proliferation, differentiation, migration, morphology and metabolism (Chen et al., 2019; Hansen et al., 2004) but also embryogenesis and potentially even arsenic-related carcinogenesis (Huang et al., 2018; Noubissi, Yedjou, Spiegelman, & Tchounwou, 2018). *IGF2BP1* is abundantly expressed in fetal and neonatal tissues (Hansen et al., 2004). Furthermore, two of the genes, *ZBTB47* and *HIVEP3* are both zinc-finger domains and are associated with transcriptional regulation (Hicar, Liu, Allen, & Wu, 2001). Epigenetic regulation of both *ZBTB47* and *HIVEP3* is known to be associated cancer (Lubecka et al., 2018; Sarkar, Leung, Baguley, Finlay, & Askarian-Amiri, 2015)

All the three gene related differentially methylated CpG sites were found in the gene body region, in both intron (*IGF2BP1*) and in exons (*ZBTB47* and *HIVEP3*). Hypermethylation at CpG sites at promoter regions represses transcription of genes which is a well-known mechanism operating in many scenarios. DNA methylation at intergenic regions and gene bodies and its impact on gene expression is gaining more attention especially in cancer studies (Jones, 2012). Interestingly, a recent study done with corals showed that gene body methylation was altered by the environmental factors, which facilitated acclimatization and adaptation to different habitats (Dixon, Liao, Bay, & Matz, 2018). However, in great tits the DNA methylation observed in gene body CpGs does not seem to affect gene expression (Derks et al., 2016) thus future studies are needed to determine the role of gene body methylation in gene expression control.

In conclusion, our study shows that early-life exposure to a toxic metal, arsenic, potentially affects fitness via DNA methylation changes in specific pathways, but not via an overall

hypomethylation in the red blood cells. The effect might more profound in other tissues that are more relevant to arsenic metabolism, such as liver. Thus, future studies should inspect also other tissues as well. Other pathways of epigenetic alterations, known to be subject to arsenic-related alternations in vitro, such as histone acetylation (Jensen (Jensen et al., 2009) and micro-RNAs (Cao et al., 2011) could be further explored.

ASSOCIATED CONTENT

**Figure S1.** Clustering of samples based on ward.D (a) and principal component analysis (PCA) (b) in Methylkit

**Table S1**. Number of reads before and after read trimming with mapping and methylation calling success of Bismark. CpG site filtering was done with Methylkit.

**Table S2.** Results from the tests on the effects of sequencing lane and sex.

**Table S3**. A full description of all differentially methylated sites.

AUTHOR INFORMATION

Corresponding author: Suvi Ruuskanen (skruus@utu.fi)

**Conflict of interest**

We have no conflict of interest to declare.

**Data accessibility**

Data will be deposited in Dryad and in Genbank upon acceptance.

**Author contributions**

SR, SE, PSV, VNL, MV, KvO and TE designed the study. TE, SE and PSV collected the data. SR and KvO designed the sequencing. VNL and MV conducted the bioinformati

analyses. SE and PSV conducted metal analyses. KvO and VNL provided the genome resources. VNL, MV, KvO and SR interpreted the data. VNL and SR wrote the first draft. All authors contributed to writing the manuscript.

**Funding sources**

TE, SE, PSV were funded by Academy of Finland (to TE), SR was also funded by Academy of Finland and Turku University Foundation.

**Acknowledgements**

We thank Miia Rainio and Jorma Nurmi for their efforts in helping us with field work and Dr. Fleur Gawehns-Bruning and William Sies for assistance with the methylation analysis. We also thank the Center for Evolutionary Applications for molecular work and Finnish Functional Genomics Centre for sequencing services. Our study was financed by KONE foundation (SR) and Academy of Finland (TE: project 265859).

**References**

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology, 13*(10), 9. doi:10.1186/gb-2012-13-10-R87

Alegria-Torres, J. A., Carrizales-Yanez, L., Diaz-Barriga, F., Rosso-Camacho, F., Motta, V., Tarantini, L., & Bollati, V. (2016). DNA Methylation Changes in Mexican Children Exposed to Arsenic From Two Historic Mining Areas in San Luis Potosi. *Environmental and Molecular Mutagenesis, 57*(9), 717-723. doi:10.1002/em.22062

Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research, 25*(22), 4692-4693. doi:10.1093/nar/25.22.4692

ATSDR. (2019). The ATSDR 2019 Substance Priority List. In: Agency for Toxic Substances and Disease Registry, U.S. Public Health Service, U.S.A.

Baccarelli, A., & Bollati, V. (2009). Epigenetics and environmental chemicals. *Current Opinion in Pediatrics, 21*(2), 243-251. doi:10.1097/MOP.0b013e32832925cc

Bantock, T. M., Prys-Jones, R. P., & Lee, P. L. M. (2008). New and improved molecular sexing methods for museum bird specimens. *Molecular Ecology Resources, 8*(3), 519-528. doi:10.1111/j.1471-8286.2007.01999.x

Bjorklund, G., Aaseth, J., Chirumbolo, S., Urbina, M. A., & Uddin, R. (2018). Effects of arsenic toxicity beyond epigenetic modifications. *Environmental Geochemistry and Health, 40*(3), 955-965. doi:10.1007/s10653-017-9967-9

Bossdorf, O., Richards, C. L., & Pigliucci, M. (2008). Epigenetics for ecologists. *Ecology Letters, 11*(2), 106-115. doi:10.1111/j.1461-0248.2007.01130.x

Boyle, P., Clement, K., Gu, H. C., Smith, Z. D., Ziller, M., Fostel, J. L., . . . Meissner, A. (2012). Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biology, 13*(10), 10. doi:10.1186/gb-2012-13-10-R92

Broberg, K., Ahmed, S., Engstrom, K., Hossain, M. B., Mlakar, S. J., Bottai, M., . . . Vahter, M. (2014). Arsenic exposure in early pregnancy alters genome-wide DNA methylation in cord blood, particularly in boys. *Journal of Developmental Origins of Health and Disease, 5*(4), 288-298. doi:10.1017/s2040174414000221

Cao, Y., Yu, S. L., Wang, Y., Guo, G. Y., Ding, Q. A., & An, R. H. (2011). MicroRNA-dependent regulation of PTEN after arsenic trioxide treatment in bladder cancer cell line T24. *Tumor Biology, 32*(1), 179-188. doi:10.1007/s13277-010-0111-z

Chen, H., Liu, J., Zhao, C. Q., Diwan, B. A., Merrick, B. A., & Waalkes, M. P. (2001). Association of c-myc overexpression and hyperproliferation with arsenite-induced malignant transformation. *Toxicology and Applied Pharmacology, 175*(3), 260-268. doi:10.1006/taap.2001.9253

Chen, J. H., Ren, X. Y., Li, L. M., Lu, S. Y., Chen, T., Tan, L. T., . . . Luo, W. (2019). Integrative Analyses of mRNA Expression Profile Reveal the Involvement of IGF2BP1 in Chicken Adipogenesis. *International Journal of Molecular Sciences, 20*(12), 15. doi:10.3390/ijms20122923

Cheng, T. F., Choudhuri, S., & Muldoon-Jacobs, K. (2012). Epigenetic targets of some toxicologically relevant metals: a review of the literature. *Journal of Applied Toxicology, 32*(9), 643-653. doi:10.1002/jat.2717

Derks, M. F. L., Schachtschneider, K. M., Madsen, O., Schijlen, E., Verhoeven, K. J. F., & van Oers, K. (2016). Gene and transposable element methylation in great tit (Parus major) brain and blood. *Bmc Genomics, 17*, 13. doi:10.1186/s12864-016-2653-y

Dixon, G., Liao, Y., Bay, L. K., & Matz, M. V. (2018). Role of gene body methylation in acclimatization and adaptation in a basal metazoan. *Proceedings of the National Academy of Sciences of the United States of America, 115*(52), 13342-13346. doi:10.1073/pnas.1813749115

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Leveque, C., . . . Sullivan, C. A. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews, 81*(2), 163-182. doi:10.1017/s1464793105006950

Eeva, T., Ryoma, M., & Riihimaki, J. (2005). Pollution-related changes in diets of two insectivorous passerines. *Oecologia, 145*(4), 629-639. doi:10.1007/s00442-005-0145-x

Ewels, P., Magnusson, M., Lundin, S., & Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics, 32*(19), 3047-3048. doi:10.1093/bioinformatics/btw354

Feng, H., Conneely, K. N., & Wu, H. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Research, 42*(8), 11. doi:10.1093/nar/gku154

Gangopadhyay, S., Sharma, V., Chauhan, A., & Srivastava, V. (2019). Potential facet for prenatal arsenic exposure paradigm: linking endocrine disruption and epigenetics. *Nucleus-India, 62*(2), 127-142. doi:10.1007/s13237-019-00274-3

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A landscape takes shape. *Cell, 128*(4), 635-638. doi:10.1016/j.cell.2007.02.006

Hansen, T. V. O., Hammer, N. A., Nielsen, J., Madsen, M., Dalbaeck, C., Wewer, U. M., . . . Nielsen, F. C. (2004). Dwarfism and impaired gut development in insulin-like growth factor II mRNA-

binding protein 1-deficient mice. *Molecular and Cellular Biology, 24*(10), 4448-4464. doi:10.1128/mcb.24.10.4448-4464.2004

Head, J. A. (2014). Patterns of DNA Methylation in Animals: An Ecotoxicological Perspective. *Integrative and Comparative Biology, 54*(1), 77-86. doi:10.1093/icb/icu025

Head, J. A., Dolinoy, D. C., & Basu, N. (2012). Epigenetics for ecotoxicologists. *Environmental toxicology and chemistry / SETAC, 31*(2), 221-227. doi:10.1002/etc.1707

Hicar, M. D., Liu, Y. L., Allen, C. E., & Wu, L. C. (2001). Structure of the human zinc finger protein HIVEP3: Molecular cloning, expression, exon-intron structure, and comparison with paralogous genes HIVEP1 and HIVEP2. *Genomics, 71*(1), 89-100. doi:10.1006/geno.2000.6425

Horsthemke, B. (2018). A critical view on transgenerational epigenetic inheritance in humans. *Nature Communications, 9*. doi:10.1038/s41467-018-05445-5

Huang, X. W., Zhang, H., Guo, X. R., Zhu, Z. X., Cai, H. B., & Kong, X. Y. (2018). Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1) in cancer. *Journal of Hematology & Oncology, 11*, 15. doi:10.1186/s13045-018-0628-y

Hubaux, R., Becker-Santos, D. D., Enfield, K. S. S., Rowbotham, D., Lam, S., Lam, W. L., & Martinez, V. D. (2013). Molecular features in arsenic-induced lung tumors. *Molecular Cancer, 12*, 11. doi:10.1186/1476-4598-12-20

Jensen, T. J., Novak, P., Wnek, S. M., Gandolfi, A. J., & Futscher, B. W. (2009). Arsenicals produce stable progressive changes in DNA methylation patterns that are linked to malignant transformation of immortalized urothelial cells. *Toxicology and Applied Pharmacology, 241*(2), 221-229. doi:10.1016/j.taap.2009.08.019

Jomova, K., Jenisova, Z., Feszterova, M., Baros, S., Liska, J., Hudecova, D., . . . Valko, M. (2011). Arsenic: toxicity, oxidative stress and human disease. *Journal of Applied Toxicology, 31*(2), 95-107. doi:10.1002/jat.1649

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics, 13*(7), 484-492. doi:10.1038/nrg3230

Kile, M. L., Houseman, E. A., Baccarelli, A. A., Quamruzzaman, Q., Rahman, M., Mostofa, G., . . . Christiani, D. C. (2014). Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood. *Epigenetics, 9*(5), 774-782. doi:10.4161/epi.28153

Koestler, D. C., Avissar-Whiting, M., Houseman, E. A., Karagas, M. R., & Marsit, C. J. (2013). Differential DNA Methylation in Umbilical Cord Blood of Infants Exposed to Low Levels of Arsenic in Utero. *Environmental Health Perspectives, 121*(8), 971-977. doi:10.1289/ehp.1205925

Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics, 27*(11), 1571-1572. doi:10.1093/bioinformatics/btr167

Laine, V. N., Gossmann, T. I., Schachtschneider, K. M., Garroway, C. J., Madsen, O., Verhoeven, K. J. F., . . . Great Tit HapMap, C. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications, 7*. doi:10.1038/ncomms10474

Liebl, A. L., Schrey, A. W., Richards, C. L., & Martin, L. B. (2013). Patterns of DNA Methylation Throughout a Range Expansion of an Introduced Songbird. *Integrative and Comparative Biology, 53*(2), 351-358. doi:10.1093/icb/ict007

Lubecka, K., Flower, K., Beetch, M., Qiu, J., Kurzava, L., Buvala, H., . . . Stefanska, B. (2018). Loci-specific differences in blood DNA methylation in HBV-negative populations at risk for hepatocellular carcinoma development. *Epigenetics, 13*(6), 605-626. doi:10.1080/15592294.2018.1481706

Mandal, B. K., & Suzuki, K. T. (2002). Arsenic round the world: a review. *Talanta, 58*(1), 201-235. doi:10.1016/s0039-9140(02)00268-0

McCrink-Goode, M. (2014). Pollution: A global threat. *Environment International, 68*, 162-170. doi:10.1016/j.envint.2014.03.023

McKay, J. A., Xie, L., Harris, S., Wong, Y. K., Ford, D., & Mathers, J. C. (2011). Blood as a surrogate marker for tissue-specific DNA methylation and changes due to folate depletion in post-partum female mice. *Molecular Nutrition & Food Research, 55*(7), 1026-1035. doi:10.1002/mnfr.201100008

Minatel, B. C., Sage, A. P., Anderson, C., Hubaux, R., Marshall, E. A., Lam, W. L., & Martinez, V. D. (2018). Environmental arsenic exposure: From genetic susceptibility to pathogenesis. *Environment International, 112*, 183-197. doi:10.1016/j.envint.2017.12.017

Noubissi, F. K., Yedjou, C. G., Spiegelman, V. S., & Tchounwou, P. B. (2018). Cross-Talk between Wnt and Hh Signaling Pathways in the Pathology of Basal Cell Carcinoma. *International Journal of Environmental Research and Public Health, 15*(7), 13. doi:10.3390/ijerph15071442

Ray, P. D., Yosim, A., & Fry, R. C. (2014). Incorporating epigenetic data into the risk assessment process for the toxic metals arsenic, cadmium, chromium, lead, and mercury: strategies and challenges. *Frontiers in Genetics, 5*, 201. doi:10.3389/fgene.2014.00201

Riyahi, S., Sánchez-Delgado, M., Calafell, F., Monk, D., & Senar, J. C. (2015). Combined epigenetic and intraspecific variation of the DRD4 and SERT genes influence novelty seeking behavior in great tit Parus major. *Epigenetics, 10*(6), 516-525. doi:10.1080/15592294.2015.1046027

Rosenfeld, C. S. (2010). Animal Models to Study Environmental Epigenetics. *Biology of Reproduction, 82*(3), 473-488. doi:10.1095/biolreprod.109.080952

Rubenstein, D. R., Skolnik, H., Berrio, A., Champagne, F. A., Phelps, S., & Solomon, J. (2016). Sex-specific fitness effects of unpredictable early life conditions are associated with DNA methylation in the avian glucocorticoid receptor. *Molecular Ecology, 25*(8), 1714-1728. doi:10.1111/mec.13483

Ruiz-Hernandez, A., Kuo, C. C., Rentero-Garrido, P., Tang, W. Y., Redon, J., Ordovas, J. M., . . . Tellez-Plaza, M. (2015). Environmental chemicals and DNA methylation in adults: a systematic review of the epidemiologic evidence. *Clinical Epigenetics, 7*, 24. doi:10.1186/s13148-015-0055-7

Saaristo, M., Brodin, T., Balshine, S., Bertram, M. G., Brooks, B. W., Ehlman, S. M., . . . Arnold, K. E. (2018). Direct and indirect effects of chemical contaminants on the behaviour, ecology and evolution of wildlife. *Proceedings of the Royal Society B-Biological Sciences, 285*(1885), 10. doi:10.1098/rspb.2018.1297

Salameh, Y., Bejaoui, Y., & El Hajj, N. (2020). DNA Methylation Biomarkers in Aging and Age-Related Diseases. *Frontiers in Genetics, 11*, 11. doi:10.3389/fgene.2020.00171

Sanchez-Virosta, P., Espin, S., Garcia-Fernandez, A. J., & Eeva, T. (2015). A review on exposure and effects of arsenic in passerine birds. *Science of the Total Environment, 512*, 506-525. doi:10.1016/j.scitotenv.2015.01.069

Sanchez-Virosta, P., Espin, S., Ruiz, S., Panda, B., Ilmonen, P., Schultz, S. L., . . . Eeva, T. (2020). Arsenic-related oxidative stress in experimentally-dosed wild great tit nestlings. *Environmental Pollution, 259*, 7. doi:10.1016/j.envpol.2019.113813

Sanchez-Virosta, P., Espin, S., Ruiz, S., Salminen, J. P., Garcia-Fernandez, A. J., & Eeva, T. (2018). Experimental manipulation of dietary arsenic levels in great tit nestlings: Accumulation pattern and effects on growth, survival and plasma biochemistry. *Environmental Pollution, 233*, 764-773. doi:10.1016/j.envpol.2017.10.113

Sarkar, D., Leung, E. Y., Baguley, B. C., Finlay, G. J., & Askarian-Amiri, M. E. (2015). Epigenetic regulation in human melanoma: past and future. *Epigenetics, 10*(2), 103-121. doi:10.1080/15592294.2014.1003746

Sepers, B., van den Heuvel, K., Lindner, M., Viitaniemi, H., Husby, A., & van Oers, K. (2019). Avian ecological epigenetics: pitfalls and promises. *Journal of Ornithology, 160*(4), 1183-1203. doi:10.1007/s10336-019-01684-5

Sharavanan, V. J., Sivaramakrishnan, M., Sivarajasekar, N., Senthilrani, N., Kothandan, R., Dhakal, N., . . . Naushad, M. Pollutants inducing epigenetic changes and diseases. *Environmental Chemistry Letters*, 19. doi:10.1007/s10311-019-00944-3

558 Skinner, M. K., Manikkam, M., & Guerrero-Bosagna, C. (2010). Epigenetic transgenerational actions
559 of environmental factors in disease etiology. *Trends in Endocrinology and Metabolism, 21*(4),
560 214-222. doi:10.1016/j.tem.2009.12.007
561 Verhoeven, K. J. F., Vonholdt, B. M., & Sork, V. L. (2016). Epigenetics in ecology and evolution: what
562 we know and what we need to know INTRODUCTION. *Molecular Ecology, 25*(8), 1631-1638.
563 doi:10.1111/mec.13617
564 Verhulst, E. C., Mateman, A. C., Zwier, M. V., Caro, S. P., Verhoeven, K. J. F., & Van Oers, K. (2016).
565 Evidence from pyrosequencing indicates that natural variation in animal personality is
566 associated with DRD4 DNA methylation. *Molecular Ecology, 25*(8), 1801-1811.
567 doi:10.1111/mec.13519
568 Vos, J. G., Dybing, E., Greim, H. A., Ladefoged, O., Lambre, C., Tarazona, J. V., . . . Vethaak, A. D.
569 (2000). Health effects of endocrine-disrupting chemicals on wildlife, with special reference
570 to the European situation. *Critical Reviews in Toxicology, 30*(1), 71-133.
571 doi:10.1080/10408440091159176
572 Wenzel, M. A., & Piertney, S. B. (2014). Fine-scale population epigenetic structure in relation to
573 gastrointestinal parasite load in red grouse (Lagopus lagopus scotica). *Molecular Ecology,*
574 *23*(17), 4256-4273. doi:10.1111/mec.12833
575 Ziech, D., Franco, R., Pappa, A., & Panayiotidis, M. I. (2011). Reactive Oxygen Species (ROS)-Induced
576 genetic and epigenetic alterations in human carcinogenesis. *Mutation Research-*
577 *Fundamental and Molecular Mechanisms of Mutagenesis, 711*(1-2), 167-173.
578 doi:10.1016/j.mrfmmm.2011.02.015

579

580