Signs of adaptation to trace metal contamination in a common urban bird

Samuel C. Andrew[1*]; Mark Patrick Taylor[2]; Sarah Lundregan[3]; Sigbjørn Lien[4]; Henrik Jensen[3]; and Simon C. Griffith[1]

[1]Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia

[2]Department of Environmental Sciences, Macquarie University, Sydney, NSW 2109, Australia

[3]Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway.

[4]Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, NO-1432 Ås, Norway.

Key words: lead exposure, avian, ecotoxicity, local adaptation, *Passer domesticus.*

* To whom correspondence should be addressed

Samuel C. Andrew

Department of Biological Sciences, Macquarie University, New South Wales, 2109, Australia

E-mail: samuel.andrew@students.mq.edu.au

*Highlights (for review : 3 to 5 bullet points (maximum 85 characters including spaces per bullet point)

Signs of adaptation to trace metal contamination in a common urban bird

Research Highlights

- Trace metal contamination can be a stressor that has negative impacts on wildlife but, we lack the empirical evidence to test if species can adapt to these stressful conditions over time.
- Using Genomic data from house sparrow populations and estimates of lead contamination from locations across Australia we find evidence that selection could be helping this invasive species adapt to heavily contaminated mining locations.
- Several of our candidate genes have links to lead and have previously been shown to have changes in their expression levels with exposure to lead.

Abstract

Metals and metalloids at elevated concentrations can be toxic to both humans and wildlife. In particular, lead exposure can act as a stressor to wildlife and cause negative effects on fitness. Any ability to adapt to stress caused by the negative effects of trace metal exposure would be beneficial for species living in contaminated environments. However, mechanisms for responding adaptively to metal contamination are not fully understood in free-living organisms. The Australian populations of the house sparrow (*Passer domesticus*) provides an excellent opportunity to study potential adaptation to environmental lead contamination because they have a commensal relationship with humans and are distributed broadly across Australian settlements including many long-term mining and smelting communities. To examine the potential for an evolutionary response to long-term lead exposure, we collected genomic SNP data using the house sparrow 200K SNP array, from 11 localities across the Australian distribution including two mining sites (Broken Hill and Mount Isa, which are two genetically independent populations) that have well-established elevated levels of lead contamination as well as trace metals and metalloids. We contrast these known contaminated locations to other lesser-contaminated environments. Using an ecological association genome scan method to identify genomic differentiation associated with estimates of lead contamination we identified 60 outlier loci across three tests. A total of 39 genes were found to be physically linked (within 20 kbps) of all outliers in the house sparrow reference genome. The linked candidate genes included 12 genes relevant to lead exposure, such as two metal transporters that can transport metals including lead and zinc across cell membranes. These candidate genes provide targets for follow up experiments comparing resilience to lead exposure between populations exposed to varied levels of lead contamination.

1. Introduction

Environmental contaminants such as lead and other trace metals pose of significant risk of harm for humans and wildlife when found in high concentrations (Peterson et al., 2017; Ray, Yosim, & Fry, 2014). The effects of exposure to trace metal contaminants on health has led to a large body of research aimed towards understanding how humans and other species respond to exposures from their environment (Lanphear, 2015; Lanphear, Vorhees, & Bellinger, 2005a; Lattin, Ngai, & Romero, 2014; Li et al., 2017; Pierron et al., 2011; Varian-Ramos, Swaddle, & Cristol, 2014). These contaminants are not always lethal and can have subtle negative effects on health. For example, the accumulation of metals inside individuals living in contaminated environments is known to be a stressor for animals (Lattin et al., 2014; Romero & Wikelski, 2001; Wikelski et al., 2002). Environmental stressors can cause physiological responses that have negative affects on fitness and reproduction (Cyr & Romero, 2007; L. Michael Romero, 2004). The adverse affects of metal-related stress on fitness has been shown to be a threatening process due to the loss of fitness causing population decline (Wikelski & Cooke, 2006). Beneficial physiological responses to stressors can also alleviate the negative effects of stress. In response to the negative effects of stress on fitness there should be positive selection for adaptations that reduce the impact of exposure in contaminated environments. A recent example of such selection was found in killifish (*Fundulus heteroclitus*) in response to pollution, using whole genome resequencing and transcriptomics they found convergent selection on the aryl hydrocarbon receptor–based signalling pathways (Reid et al., 2016). However, examples of this kind are limited and studying adaptation in urban environments is a growing field of research (Johnson & Munshi-South, 2017).

One of the most studied metal contaminants is lead (Pb) because of its toxicity and its known adverse impacts on human health (Lanphear, 2015; National Toxicology Program, 2012). Environmental lead in aerosols, dusts and soils, amoung other things, are elevated in most global urban city environments due to the former massive emissions of lead from industrial sources with a majority coming from leaded petrol combustion (Kristensen, 2015; Kristensen, Taylor, & Flegal, 2017; Laidlaw et al., 2017; Mielke, Laidlaw, & Gonzales, 2011; Olszowy et al., 1995; Rouillon et al., 2017). These industrial emissions and depositions remain present and bioavailable in the environment (Laidlaw et al., 2017; Mackay et al., 2013) and consequently present a risk of harm to a range of organisms living in urban environments. In Australia, there are a number of locations that have a protracted history of environmental lead emission from lead mining and smelting practices, as well as other trace metals such as cadmium and zinc (Dong et al., 2015; Kristensen & Taylor, 2016; Mackay et al., 2013; Taylor et al., 2010; Taylor et al., 2014b). Environmental exposures in lead producing communities as well as those impacted by former leaded petrol depositions are typically via the ingestion of soils and dusts (Gulson et al., 2014). Unlike dust, soil metal concentrations have national guidelines values promulgated under the NEPM (2013). The most relevant soil lead value is the health investigation level guideline of 300 mg/kg NEPM (2013), which is applied to residential dwellings and to ensure blood lead levels remain below 7.5 µg/dL. However, it is well-accepted that there is no safe lower threshold for blood lead exposure (Lanphear et al., 2005a) and that blood leads as at 2 µg/dL or lower are considered harmful to human health (National Toxicology Program, 2012). Therefore, we use the lowest acceptable upper limit of 300 mg/kg for soil lead provided in the NEPM (2013) as a benchmark for assessing the level of contamination seen at our study sites.

Describing how species adapt to trace metal contamination should be a priority for understanding and managing urban ecosystems. The house sparrow's (*Passer domesticus*) introduction to Australia provides a good opportunity for studying adaptation to contaminated urban environments in a common species due to its broad distribution with recently established and isolated populations including lead-contaminated sites. The house sparrow is an obligate commensal species with humans and are typically constrained to urban and rural environments (Anderson, 2006). Sparrows also have sedentary populations with moderate genetic population structure and relatively low gene flow (Jensen et al., 2013; Kekkonen et al, 2011), and this is also true for populations living in broadly spaced human settlements across Australia (Andrew et al., 2017). The invasion of the house sparrow to different mining communities provides an opportunity to observe genetic adaptation to lead contamination through genetic differentiation that is associated with levels of lead contamination. Any selection that has have taken place would have happened over a relative short time frame. For example, sparrows have been present in the mining communities of Broken Hill (New South Wales) and Mount Isa (Queensland) for approximately 100 and 50 years respectively (Andrew & Griffith 2016). Long term mining practices, and relatively high levels of contamination have potentially allowed for selection to take place over dozens of generations (50 and 25 in the two locations respectively, assuming a generation time of approx. two years [Jensen et al. 2013]) to ameliorate the deleterious effects of lead and/or trace metal contamination on fitness. Through previous descriptions of genetic population structure we know these two locations have independent populations (Andrew et al., 2017), and to an extent they are genetically independent replicates in which selection can occur in parallel.

There are many potential physiological mechanisms sparrows could employ to adaptively respond to lead contamination in their environment. Lead (Pb) is not essential to life and is a non-biodegradable element, meaning there are no efficient pathways for it to be metabolized and eliminated, thus it tends to accumulate in the different organs and tissues of individuals exposed to lead (Peakall & Burger, 2003). The first line of defence for species would be through reducing the unintentional uptake of lead into the body through external surfaces like the respiratory system (Ribeirovet al., 2014), the alimentary canal (Madigosky et al., 1991), or the inner ear (Ding et al., 2014). Another suggested trait involved in adapting to lead contamination is increased exclusion through the accumulation of lead in the kidneys and other tissues to reduce the amount lead bioavailable in the body (Ribeiro et al., 2014). Lead is also known to be a stressor to the endoplasmic reticulum of cells, affecting protein production, however this effect can also be countered through different molecular pathways that remove misfolded proteins (Qian & Tiffany-Castiglioni, 2003; Shinkai et al., 2010). Alternatively, sparrows could also respond to lead by trying to avoid contamination through behavioural modifications that reduce exposure.

We have collected genomic data from 11 house sparrow populations across Australia including the mining locations of Broken Hill and Mount Isa, using a 200K SNP array (Lundregan et al., 2018). This is an observational study, in which we aim to explore genetic differentiation across loci that is related to geographical variation in lead contamination. We are particularly interested in the contrast between the heavily contaminated locations of Broken Hill and Mount Isa, and the other nine populations. We predict that some of the loci that are significantly associated with lead contamination in our analyses will be physically linked to

112 genes in the house sparrow genome that are related to known traits associated with lead
113 contamination.

114

115 **2. Methods**

116 *2.1 Sampling and genetic data collection*

117 Adult house sparrows were sampled at urban sites from 11 towns/cities across Australia
118 and 16 male individuals from each of our sampling localities were used for genotyping; 176
119 individuals in total. Sampling was carried out under the Animal Research Authority of the
120 Animal Ethics Committee at Macquarie University (ARA 2014/248).

121 The individuals were genotyped on the house sparrow Affymetrix 200K SNP array
122 (Lundregan et al., 2018). SNP filtering for loci with minor alleles frequencies of less than 5%
123 and loci with more than 9% missing data, was done using the program PLINK (Purcell et al.,
124 2007). After quality filtering there were 162,299 SNPs used in further analyses and 176
125 individuals. The SNP data used here is available via a figshare repository, at:
126 https://figshare.com/s/e632e401fe8fc0a49bcb, in PED file format. Data conversion between PED
127 and BayeScEnv formats was performed using PGDSpider (Lischer & Excoffier, 2012).

128 *2.2 Environmental lead characterisation*

129 Soil lead concentrations are characteristically heterogenous, even over small areas
130 (Rouillon et al., 2017). However, mean soil lead levels in most urban areas around Australia are
131 typically higher than background levels (ca. less than 30 mg/kg [Callender, 2014]). A report by
132 Olszowy *et al.* (1995) illustrates this pattern clearly by describing mean lead levels in

environments with different levels of urbanisation across the main states of Australia. Empirical data has shown that Mount Isa and Broken Hill have much higher than typical average concentrations of lead and other trace metals such as cadmium and zinc, are elevated above the state capital cities of Australia (Kristensen & Taylor, 2016; Mackay et al., 2013; Taylor et al., 2010; Taylor et al., 2014a; Taylor et al., 2014b). This study relies on available soil lead data to characterise concentrations at the different sample sites. We have focused on using soil lead measurements because these soil lead levels are stable over time (Semlali et al., 2004) and house sparrows spend a large amount of time on the ground foraging and dust bathing. Atmospheric sourced lead depositions from automotive vehicles, industry or mining typically accumulate in the uppermost section of the soil profile, i.e. the top 2 cm (Taylor et al., 2010), which forms the most significant component of soil exposed to sparrows. We found relevant site-specific and regional estimates of soil lead measurements for all our sampling localities (Table 1).

The lead data was converted to environmental differentiation for use in the BayeScEnv genome scan analyses (de Villemereuil & Gaggiotti, 2015). This method was chosen because it relates genetic differentiation to environmental differentiation to find loci across the genome that have the strongest covariation with environmental differentiation and does not focus on clinal changes in a variable which you would more normally see with climatic variables. The covariation between genetic and environmental differentiation is likely to be a result of selection. The suggested method for entering the environmental variable into BayeScEnv is as a standardised distances of environmental differentiation. This requires defining a mean environmental condition and calculating how many standard deviations each location is above or below that mean for a given environmental variable. These values should not be more than 3

155 standard deviations above or below the mean to avoid over stressing the model (de Villemereuil
156 & Gaggiotti, 2015).

157 In order to account for inaccuracies in estimates of environmental lead levels we
158 replicated our analysis with predicted and estimated environmental differentiation scores. The
159 first model (Predicted model) uses our predicted values for environmental differentiation based
160 on historical contamination. This model predicts that mining locations had extreme lead
161 contamination and were given a score of 3 and all other locations had an average score of 0. This
162 provided an exploratory model to compare our results from the estimated contamination models.
163 The second model (Standard model) defined our standard environmental condition as the average
164 lead contamination in non-mining locations. We then standardised the distances from this mean
165 by dividing by the standard deviation of non-mining locations (See Table 1). Our standard
166 environmental condition was calculated without the outliers of the mining locations because we
167 wanted to define a standard environment not the average for the sample sites used. Consequently,
168 the locations of Mount Isa and Broken Hill (mining locations) had environmental differentiation
169 scores that were greater than 3, so all of our environmental differentiation scores were scaled
170 down by dividing by three (Table 1). For Broken Hill and Mount Isa we still had to round the
171 environmental differentiation scores to 3 to minimise supplying the model with excessive
172 outliers (Mount Isa only adjusted from 3.1). In the third model (All sites model) all locations,
173 including the outliers of Mount Isa and Broken Hill, were used to calculate the mean
174 environment and the standard deviation to get environmental differentiation. Because Mount Isa
175 and Broken Hill have such high lead contamination it is unlikely that our lead environmental
176 variable is conflated with other environmental variables such as climate or urbanisation and is

therefore the only focus of this study. A full summary of the soil lead values applied to the different locations is detailed in Table 1.

2.3 Statistical analyses

The Program BayeScEnv was used to do Ecological Association (EA) analyses with our estimates of environmental lead levels. The model was run separately using predicted values and estimates of environmental lead contamination (See Table 1). Models were run using the same options: 10 pilot runs with 2 000 iterations; a thinning interval of 10; 5 000 outputted iterations and a burn-in length of 10 000. All other options were kept as the default. All figures and additional analyses were done using R (R core team, 2017).

2.4 Linked genes and gene ontologies

The physical position of outlier SNPs was used to identify annotated genes in the reference genome that are physically linked to the SNPs. The house sparrow reference genome currently has 14 260 known protein coding genes (Elgvin et al., 2017). As the average LD across the house sparrow genome drops to approximately double the background level at ca. 20 kbps (Lundregan et al., 2018), 20 kbps was used as a conservative window for physical linkage between SNPs and potential candidate genes for targets of selection. We recorded all genes within 20 kbps of each SNP and whether the SNP was within a gene or within a known exon of a gene. The proportion of SNPs linked to genes was calculated for all analyses. For the SNPs that were significant in more than one model we looked up the closest protein coding gene within 200 kbps to the SNP.

3. Results

199 The BayeScEnv EA models tested whether there was a significant association between the
200 differentiation in allele frequencies of our 162,299 SNP loci and lead contamination. The three
201 BayeScEnv models identified 60 significant SNPs in total (Predicted model = 19, Standard
202 model = 10, and All sites model = 40 SNPs, using q-values, Figure 2 and Table S1), there were
203 six SNPs significant across multiple models (Figure 3). The 60 Significant SNPs were found to
204 be physically linked to 39 genes, within 20 kbps of the SNP (see details Table 2 and S2). There
205 were two SNPs significant in all three models. SNPi41730 on chromosome 3 is linked to the
206 *FAM167A* gene which is associated with endoplasmic reticulum membrane structure.
207 SNPa160965 on chromosome 5 is within 42 kbps of both *TMEM251* and *UBR7*, which is still a
208 plausible distance for physical linkage (Lundregan et al., 2018, Table S2). One of the significant
209 SNPs from the Predicted model, SNPa52315, was linked to a metal ion transmembrane
210 transporter (solute transporter protein *SLC39A9*) associated with the transport of zinc and other
211 metals. Interestingly. Another SNP that was significant in the All sites model (SNPa222796) was
212 located within 20 kbps of another zinc transporter (*Slc39a8*). In the Standard model, the SNP
213 linked to the *SLC39A9* gene (SNPa52315) was also nearing significance (q-value = 0.061) and
214 the SNP had the 15th smallest q-values out of all 162,299 SNPs.

215 The Predicted model used average lead levels (score of 0) for all localities except the
216 mining locations of Mount Isa and Broken Hill, which were given the maximum lead score of 3.
217 This run identified five significant SNPs on the 1A chromosome that were within 40kbps of each
218 other. This was the only peak of significant SNPs that were linked to each other and interestingly
219 the nearest downstream gene was a third solute transporter protein (*SLC6A15*, ca. 400kbps away,
220 Table S2).

221

4. Discussion

After using162,299 SNPs from across the house sparrow genome, in three alternative models we found 60 significant outlier SNPs associated with estimates of lead pollution in the 11 locations investigated (two with high levels of lead). At this time, it is impossible to determine if the genes physically linked to our outlier SNPs contain a higher than expected number of gene relevant to lead exposure for the house sparrow because no list of relevant genes is available. We have researched the genes linked to our outliers to identify genes most likely to be relevant to lead exposure to help develop a list of potential genes important for biological responses to lead exposure. Some of the significant SNPs identified here were physically linked to 12 genes that were previously described to be linked to traits related to lead contamination in other species. The roles of these genes that are potentially associated with lead include metal ion transmembrane transporters, endoplasmic reticulum function, protein ligase activity, and expression changes in affected tissues due to lead exposure. For example the *SLC39A9* gene that was linked to a significant SNP in the Predicted mode, is a known zinc transporter in chicken B cells (Taniguchi et al., 2013). In sparrows this transporter could also be involved in the transport of lead or other metals that are common in the mining locations of Broken Hill and Mount Isa such as zinc itself (Dong et al., 2015; Taylor et al., 2010; Taylor et al., 2014b). Another zinc transporter that was linked to a significant SNP from the All sites model was *Slc39a8*. This highly conserved gene codes for the ZIP8 zinc transporter which is known to be one of the three primary transporters of lead in many taxa (Liu *et al.* 2008; Nebert *et al.* 2012; Ding *et al.* 2014), although only two of the three transporter genes are present on the house sparrow reference genome (ZIP8 and ZIP14, but not DMT1). The *Slc39a8* transporter can also transport cadmium which is another toxic metal pollutant (Liu et al., 2008; Prozialeck et al., 2008) found in the

mining city of Broken Hill (Dong et al., 2015). The *Slc39a8* gene has been shown to be expressed in the inner ear of rats (*Rattus norvegicus*, Ding et al., 2014) and selection for the down-regulation of these trace metal transporters could result in less lead and other trace metals entering the body. For example, in a transcriptomics study *Slc39a8* was down-regulated in rats exposed to high lead concentrations (Schneider et al., 2012). Another linked gene *OTOGL* to an outlier SNP (Table S2), is known to be important to the development of the inner ear (Yariz et al., 2012). The *SLC39* family of zinc transporters are generally involved in bringing metals into the body rather than excreting them (Jeong & Eide, 2013). Therefore, a likely mechanism for responding to lead contamination is to down-regulate transporters on external surfaces to reduce the amount of lead unintentionally entering the body.

Lead is known to be a stressor/suppressor of the functioning of the endoplasmic reticulum inside cells (Shinkai et al., 2010). In all three BayeScEnv models a SNP linked to the *FAM167A* gene was found to be significant and this gene is connected to the structural formation of the endoplasmic reticulum (Table S2). Another SNP significant in all three models (SNPa160965), and another in the All sites model (SNPa160930), was linked to the *UBR7* gene. The gene ontology of UBR7 is ubiquitin protein ligase activity, a process important for breaking down incorrectly folded proteins on the endoplasmic reticulum (Smith, Ploegh, & Weissman, 2011; Teixeira & Reed, 2013) and more specifically E3 ubiquitin ligases like *UBR7* (Christianson & Ye, 2014; Sriram, Kim, & Kwon, 2011). For another example *UBE2J1* has been linked to the task of breaking down terminally misfolded proteins on the endoplasmic reticulum (Burr et al., 2011). The Outlier SNPa291647 is also physical linked to the *UBE2J1* gene on the house sparrow reference genome and was a significant outlier in the All sites model (this SNP also had the 32nd lowest q-value in the Standard model; q-value = 0.108).

There were several other linked genes (listed in Table S2) that have been shown to have changes in their expression level in individuals exposed to lead. The gene *ATP6V1A* has been shown to be downregulated in primary rat proximal tubular (rPT) cells exposed to lead (Song et al., 2017). rPT cells allow for the reabsorption of important nutrients and the excretion of unneeded molecules through the renal system. Selection for upregulation in *ATP6V1A* may have occurred to prevent detrimental lysosomal acidification in lead exposed cells. Lysosomal acidification can also cause neurodegenerative diseases, which is also linked to v-ATPase pumps like *ATP6V1A* (Colacurcio et al., 2017). Another linked gene important to neural development is *IGF2R*, a receptor for IGF2. Pup rats from mothers exposed to lead show decreased expression of *IGF2* in the cerebral cortex that may result in neural damage due to *IGF2*'s important role in neural development (Li et al. 2016). Lead has also been shown to affect the methylation of *IGF2R* and affects methylation during aging (Faulk et al. 2014). The *DUSP12* gene has been shown to have changes in expression in the hippocampus of rats (upregulated in inbred lines and downregulated in outbred lines) exposed to lead (Schneider et al., 2014). Lead has negative effects on bone as well as neural development and exposure to lead is known to suppress bone formation. The suppression of bone development is believed to be linked to the inhibition of the *Wnt* pathway that can impede osteoblast activity and as a result bone development (Beier et al., 2015; Hu et al., 2014). Four genes linked to the *Wnt* pathway were identified as genes linked to outlier SNPs in this study, including *Cby1* (Corbin et al., 2009), *EFEMP1* (Yang et al., 2016), *SKP2* (Tang et al., 20019) and *DACT1* (Hou et al., 2015).

The outlier SNPs that are physically linked genes relating to lead transport and negative impacts on biological processes give support for the presence of local adaptation in Australian populations of the house sparrow to lead contamination. As Broken Hill and Mount Isa are in

separate genetic populations there has likely been two independent occurrences of local adaptation to lead exposure (or other trace metals such as zinc) from the contaminated soil and dust in these environment (Dong et al., 2015; Taylor et al., 2010).

In total 13 of our 60 significant SNPs were physically linked (within ca. 20 kbps) to 12 genes that had prior associations to traits relating to lead pollution (Table S1 and S2). It is unclear at this stage if other significant loci are associated with traits with unknown relationships to how sparrows might be responding to contamination from lead and other metals. Many loci can be a part of a gene networks that are responsible for regulating polygenic traits and loci that are not physically linked to protein coding regions can also be important to gene networks because they are associated with loci important to gene regulation e.g. promoter regions, epigenetic markers and non-coding RNA segments (Allis & Jenuwein, 2016; Gutierrez-arcelus, Ongen, & Lappalainen, 2015; Koch, 2015; Novère, 2015). As we develop our understanding of gene regulation it may become easier to identify links between candidate loci and the target traits, which are predicted to respond to natural selection. This development can happen with improvements to the annotation of the house sparrow reference genome, to include non-protein coding loci that are important to gene expression (Elgvin et al., 2017). However, not all significant loci are going to be true signals of natural section. Indeed, false discovery rates are almost unavoidable for genome scan methods (de Villemereuil, Frichot, Bazin, François, & Gaggiotti, 2014; Francois, Martins, Caye, & Schoville, 2016; Whitlock & Lotterhos, 2015). Demographic history and high neutral genetic drift in introduced species can also increase false discovery rates (de Villemereuil et al., 2014; Hoban et al., 2016; Shultz, Baker, Hill, Nolan, & Edwards, 2016). However, the independent introductions to the two mining locations also adds strength to this ‘natural’ experiment for local adaptation in the house sparrow.

How species adapt to pollutants in their environment is important for their longevity and species diversity in areas most affected by human activity. Further research can extend our understanding of how house sparrows adapt to high lead contamination and this information could help predict the resilience of related species to lead contamination. We predict that sparrows living in mining towns will have higher than average blood lead levels, but this increase would not be linear with increases in soil lead levels, due to adaptations to mitigate lead absorption from the environment. To test this hypothesis a 2x2 common garden experiment could be used. Sparrows from mining and non-mining towns could be collected from the wild and brought to controlled captive environments. After a holding period blood lead levels could be measured before and after low and high lead exposure treatments. We would expect the birds from the mining communities to have a lower accumulation of lead and a faster recovery. Based on the differences in accumulation and recovery rates follow up studies could look at specific mechanisms relating to lead tolerance by observing variation in gene expression and physiology.

From a conservation perspective adapting to pollutants in the environment will be important to the health of populations and ecosystems. The ability to adapt to contamination could reduce the impacts of pollution over time in some species. Some of the strongest examples of evolution have been observed when species are exposed to drastically different conditions to those of their original environment (Reid et al., 2016). Anthropogenic contamination of the environment has occurred over a relatively short evolutionary time scale and its effects on wildlife need to be modelled using a wide range of species to establish more clearly cause and effect. Nevertheless, in this study we have found clear indications that local adaptation is taking place at the genetic level in introduced Australian house sparrow populations, even after a

relatively short period of time. We propose that the house sparrow will be a good model system to study these mechanisms further.

Acknowledgements: For funding support: S.C.A. was supported by Macquarie University Research Excellence Scholarships (no. 2013077). S.C.G. was supported by an Australian Research Council Future Fellowship (FT130101253). H.J., I.J.H. and S.L. were supported by the Research Council of Norway (project no. 211956), and by the Research Council of Norway's Centres of Excellence funding scheme (project 223257). SNP array genotyping was conducted at CIGENE Norway. We thank, Monica Awasthy, Amanda D. Griffith and Elizabeth L. Sheldon for their participation with field work. We would also like to thank Lee Ann Rollins for their assistance with DNA quantification. **Author Contributions:** S.C.G. and S.C.A. conceived the study and collected the samples. M.P.T collected lead data. H.J. and S. Lien coordinated genotyping using the SNP array. S.C.A. performed Bioinformatics with input from other co-authors. H.J., and S. Lundregan participated in the interpretation of results and the identification of gene ontology. All authors contributed to writing the manuscript.

References

Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, *17*(8), 487–500.

Anderson, T. R. (2006). Biology of the Ubiquitous House Sparrow: from Genes to Populations. In *University Press New York.* (pp. 31–56). New York: Oxford University Press,Inc.

Andrew, S. C., Awasthy, M., Bolton, P. E., Rollins, L. A., Nakagawa, S., & Griffith, S. C. (2017). The genetic structure of the introduced house sparrow populations in Australia and New Zealand is consistent with historical descriptions of multiple introductions to each country. *Biological Invasions*, *20*, 1507-1522.

Andrew, S. C., & Griffith, S. C. (2016). Inaccuracies in the history of a well-known introduction: a case study of the Australian House Sparrow (Passer domesticus). *Avian Research*, *7*(1), 9.

Beier EE, Sheu TJ, Dang D, Holz JD, Ubayawardena R, Babij P, Puzas JE. (2015). Heavy metal ion regulation of gene expression: Mechanisms by which lead inhibits osteoblastic bone-forming activity through modulation of the Wnt/β-catenin signaling pathway. *Journal of Biological Chemistry*, *290*, 18216–18226.

Burr, M. L., Cano, F., Svobodova, S., Boyle, L. H., Boname, J. M., & Lehner, P. J. (2011). HRD1 and UBE2J1 target misfolded MHC class I heavy chains for endoplasmic reticulum-associated degradation. *PNAS*, *108*(5), 2034–2039.

Callender, E. (2014). Heavy Metals in the Environment – Historical Trends. In *Treatise on Geochemistry* (Vol. 9, pp. 67–105).

Christianson, J. C., & Ye, Y. (2014). Cleaning up in the endoplasmic reticulum: ubiquitin in charge. *Nature Structural & Molecular Biology*, *21*(4), 325–35.

Colacurcio DJ, Nixon RA. (2016). Disorders of lysosomal acidification—The emerging role of v-ATPase in aging and neurodegenerative disease. *Ageing Research Reviews, 32*, 75–88.

376 Corbin, M., de Reyniès, A., Rickman, D. S., Berrebi, D., Boccon-Gibod, L., Cohen-Gogo, S., …
377 Jeanpierre, C. (2009). WNT/β-catenin pathway activation in Wilms tumors: A unifying
378 mechanism with multiple entries? *Genes, Chromosomes and Cancer*, *48*, 816–827.

379 Cyr, N. E., & Romero, M. L. (2007). Chronic stress in free-living European starlings reduces
380 corticosterone concentrations and reproductive success. *General and Comparative*
381 *Endocrinology*, *151*(1), 82–89.

382 de Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan
383 methods against more complex models: when and how much should we trust them?
384 *Molecular Ecology*, *23*(8), 2006–2019.

385 de Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local
386 adaptation using environmental variables. *Methods in Ecology and Evolution*, *6*(11), 1248–
387 1258.

388 Ding, D., Salvi, R., & Roth, J. A. (2014). Cellular localization and developmental changes of
389 Zip8, Zip14 and transferrin receptor 1 in the inner ear of rats. *BioMetals*, *27*(4), 731–744.

390 Dong, C., Taylor, M. P., Kristensen, L. J., & Zahran, S. (2015). Environmental contamination in
391 an Australian mining community and potential influences on early childhood health and
392 behavioural outcomes. *Environmental Pollution*, *207*, 345–356.

393 Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., … Sætre,
394 G.-P. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, *3*, e1602996.

395 Faulk C, Liu K, Barks A, Goodrich JM, Dolinoy DC. (2014). Longitudinal epigenetic drift in
396 mice perinatally exposed to lead. *Epigenetics, 9*, 934–941.

Francois, O., Martins, H., Caye, K., & Schoville, S. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology*, *25*, 454–469.

Gulson, B., Mizon, K., Taylor, A., Korsch, M., Davis, J. M., Louie, H., … Antin, L. (2014). Pathways of Pb and Mn observed in a 5-year longitudinal investigation in young children and environmental measures from an urban setting. *Environmental Pollution*, *191*, 38–49.

Gutierrez-arcelus, M., Ongen, H., & Lappalainen, T. (2015). Tissue-Specific Effects of Genetic and Epigenetic Variation on Gene Regulation and Splicing. *PLOS Genetics*, 1–25.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, *188*(4), 379–397.

Hou, J., Wen, Y.-H., Feng, K.-N., Ma, X.-F., & Yao, J.-P. (2015). DACT1 is involved in human placenta development by promoting Wnt signaling. *Archives of Gynecology and Obstetrics*, *291*(6), 1289–1296.

Hu F, Xu L, Liu ZH, Ge MM, Ruan DY, Wang HL. (2014). Developmental lead exposure alters synaptogenesis through inhibiting canonical wnt pathway in vivo and in vitro. *PLoS ONE*, *9*, 3–10.

Jensen, H., Moe, R., Hagen, I. J., Holand, A. M., Kekkonen, J., Tufto, J., & Saether, B.-E. (2013). Genetic variation and structure of house sparrow populations: is there an island effect? *Molecular Ecology*, *22*(7), 1792–805.

Jeong, J., & Eide, D. J. (2013). The SLC39 family of zinc transporters. *Molecular Aspects of Medicine*, *34*, 612–619.

418 Johnson MTJ, Munshi-South J. (2017). Evolution of life in urban environments. *Science, 358*,
419 eaam8327.

420 Kekkonen, J., Hanski, I. K., Jensen, H., Väisänen, R. a., & Brommer, J. E. (2011). Increased
421 genetic differentiation in house sparrows after a strong population decline: From panmixia
422 towards structure in a common bird. *Biological Conservation*, *144*(12), 2931–2940.

423 Koch, L. (2015). Genomics: Adding another dimension to gene regulation. *Nature Reviews*
424 *Genetics*, *16*(10), 563–563.

425 Kristensen, L. J. (2015). Quantification of atmospheric lead emissions from 70 years of leaded
426 petrol consumption in Australia. *Atmospheric Environment*, *111*, 195–201.

427 Kristensen, L. J., & Taylor, M. P. (2016). Unravelling a "miner"s myth' that environmental
428 contamination in mining towns is naturally occurring. *Environmental Geochemistry and*
429 *Health*, *38*, 1015–1027.

430 Kristensen, L. J., Taylor, M. P., & Evans, A. J. (2016). Tracing changes in atmospheric sources
431 of lead contamination using lead isotopic compositions in Australian red wine.
432 *Chemosphere*, *154*, 40–47.

433 Kristensen, L. J., Taylor, M. P., & Flegal, A. R. (2017). An odyssey of environmental pollution:
434 The rise, fall and remobilisation of industrial lead in Australia. *Applied Geochemistry*, *83*,
435 3–13.

436 Laidlaw, M. A. S., Gordon, C., & Ball, A. S. (2018). Preliminary assessment of surface soil lead
437 concentrations in Melbourne, Australia. *Environmental Geochemistry and Health*, *40*(2),
438 637–650.

439 Laidlaw, M. A. S., Mohmmad, S. M., Gulson, B. L., Taylor, M. P., Kristensen, L. J., & Birch, G.
440 (2017). Estimates of potential childhood lead exposure from contaminated soil using the US
441 EPA IEUBK Model in Sydney, Australia. *Environmental Research*, *156*, 781–790.

442 Lanphear, B. P. (2015). The Impact of Toxins on the Developing Brain. *Annual Review of Public*
443 *Health*, *36*(1), 211–230.

444 Lanphear, B. P., Hornung, R., Khoury, J., Yolton, K., Baghurst, P., Bellinger, D. C., … Roberts,
445 R. (2005a). Low-level environmental lead exposure and children's intellectual function: An
446 international pooled analysis. *Environmental Health Perspectives*, *113*, 894–899.

447 Lanphear, B. P., Vorhees, C. V, & Bellinger, D. C. (2005b). Protecting children from
448 environmental toxins. *PLoS Medicine*, *2*(3), e61.

449 Lattin, C. R., Ngai, H. M., & Romero, L. M. (2014). Evaluating the Stress Response as a
450 Bioindicator of Sub-Lethal Effects of Crude Oil Exposure in Wild House Sparrows (Passer
451 domesticus). *PLoS ONE*, *9*(7), e102106.

452 Li N, Yang G, Wang Y, Qiao M, Zhang P, Shao J, Yang G. (2016). Decreased IDE and IGF2
453 expression but increased Aβ40in the cerebral cortex of mouse pups by early life lead
454 exposure. *Brain Research Bulletin, 121*, 84–90.

455 Li, X., Xing, M., Chen, M., Zhao, J., Fan, R., Zhao, X., … Xu, S. (2017). Effects of selenium-
456 lead interaction on the gene expression of inflammatory factors and selenoproteins in
457 chicken neutrophils. *Ecotoxicology and Environmental Safety*, *139*(November 2016), 447–
458 453.

459 Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for

460 connecting population genetics and genomics programs. *Bioinformatics*, *28*, 298–299.

461 Liu, Z., Li, H., Soleimani, M., Girijashanker, K., Reed, J. M., He, L., … Nebert, D. W. (2008).
462 Cd2+ versus Zn2+ uptake by the ZIP8 HCO3--dependent symporter: Kinetics,
463 electrogenicity and trafficking. *Biochemical and Biophysical Research Communications*,
464 *365*(4), 814–820.

465 Lundregan, S. L., Hagen, I. J., Gohli, J., Niskanen, A. K., Kemppainen, P., Ringsby, T. H., …
466 Jensen, H. (2018). Inferences of genetic architecture of bill morphology in house sparrow
467 using a high density SNP array point to a polygenic basis. *Molecular Ecology, in press*.

468 Mackay, A. K., Taylor, M. P., Munksgaard, N. C., Hudson-Edwards, K. A., & Burn-Nunes, L.
469 (2013). Identification of environmental lead sources and pathways in a mining and smelting
470 town: Mount Isa, Australia. *Environmental Pollution*, *180*, 304–311.

471 Madigosky, S. R., Alvarez-hernandez, X., & Glass, J. (1991). Lead, Cadmium, and Aluminum
472 Accumulation in the Red Swamp Crayfish Procambarus clarkii G. Collected from Roadside
473 Drainage Ditches in Louisiana. *Arch. Environ. Contam. Toxicol.*, *20*, 253–258.

474 Mielke, H. W., Laidlaw, M. A. S., & Gonzales, C. R. (2011). Estimation of leaded (Pb)
475 gasoline's continuing material and health impacts on 90 US urbanized areas. *Environment*
476 *International*, *37*(1), 248–257.

477 National Toxicology Program. (2012). National Toxicology Program: health effects of low-level
478 lead. *NTP Monograph*, (June), 1–148. Retrieved from
479 https://www.niehs.nih.gov/health/topics/agents/lead/%5Cnhttp://www.ncbi.nlm.nih.gov/pub
480 med/23964424%5Cnhttps://ntp.niehs.nih.gov/ntp/ohat/lead/final/monographhealtheffectslo
481 wlevellead_newissn_508.pdf

Nebert, D. W., Gálvez-Peralta, M., Hay, E. Ben, Li, H., Johansson, E., Yin, C., … Soleimani, M. (2012). ZIP14 and ZIP8 zinc/bicarbonate symporters in Xenopus oocytes: characterization of metal uptake and inhibition. *Metallomics*, *4*, 1218–25.

NEPM. (2013). Schedule B1: Guideline on Investigation for Soil and Groundwater. In *National Environmental Protection Measure (NEPM)* (p. 57).

Novère, N. Le. (2015). Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics*, *16*(3), 146–158.

Olszowy, H., Torr, P., Imray, P., Smith, P., Hegarty, J., & Hastie, G. (1995). *Trace Element Concentrations in Soils from Rural and Urban Areas of Australia. Contaminated Sites Monograph Series. No.4.* Retrieved from http://www.urbanleadpoisoning.com/Trace%2520Elements%25 20Surface%2520Soils%2520Urban%2520onurbna.pdf.

Peakall, D., & Burger, J. (2003). Methodologies for assessing exposure to metals: Speciation, bioavailability of metals, and ecological host factors. *Ecotoxicology and Environmental Safety*, *56*, 110–121.

Peterson, E. K., Buchwalter, D. B., Kerby, J. L., LeFauve, M. K., Varian-Ramos, C. W., & Swaddle, J. P. (2017). Integrative behavioral ecotoxicology: bringing together fields to establish new insight to behavioral ecology, toxicology, and conservation. *Current Zoology*, *63*(2), 185–194.

Pierron, F., Normandeau, E., Defo, M. A., Campbell, P. G. C., Bernatchez, L., & Couture, P. (2011). Effects of chronic metal exposure on wild fish populations revealed by high-throughput cDNA sequencing. *Ecotoxicology*, *20*(6), 1388–1399.

Prozialeck, W. C., Edwards, J. R., Nebert, D. W., Woods, J. M., Barchowsky, A., & Atchison, W. D. (2008). The vascular system as a target of metal toxicity. *Toxicological Sciences*, *102*(2), 207–218.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575.

Qian, Y., & Tiffany-Castiglioni, E. (2003). Lead-induced endoplasmic reticulum (ER) stress responses in the nervous system. *Neurochemical Research*, *28*(1), 153–162.

R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.r-project.org/

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*(JUL), 1–26.

Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., … Whitehead, A. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science*, *354*(6317), 1305–1308.

Ribeiro, A. M., Risso, W. E., Fernandes, M. N., & Martinez, C. B. R. (2014). Lead accumulation and its effects on the branchial physiology of Prochilodus lineatus. *Fish Physiology and Biochemistry*, *40*(3), 645–657.

Romero, L. M. (2004). Physiological stress in ecology: Lessons from biomedical research. *Trends in Ecology and Evolution*, *19*(5), 249–255.

Romero, L. M., & Wikelski, M. (2001). Corticosterone levels predict survival probabilities of Galapagos marine iguanas during El Nino events. *Proceedings of the National Academy of Sciences*, *98*(13), 7366–7370.

Rouillon, M., Gore, D. B., & Taylor, M. P. (2013). The nature and distribution of Cu, Zn, Hg, and Pb in urban soils of a regional city: Lithgow, Australia. *Applied Geochemistry*, *36*, 83–91.

Rouillon, M., Harvey, P. J., Kristensen, L. J., George, S. G., & Taylor, M. P. (2017). VegeSafe: A community science program measuring soil-metal contamination, evaluating risk and providing advice for safe gardening. *Environmental Pollution*, *222*, 557–566.

Schneider, J. S., Anderson, D. W., Talsania, K., Mettil, W., & Vadigepalli, R. (2012). Effects of developmental lead exposure on the hippocampal transcriptome: Influences of sex, developmental period, and lead exposure level. *Toxicological Sciences*, *129*(1), 108–125.

Schneider JS, Talsania K, Mettil W, Anderson DW. (2014). Genetic Diversity Influences the Response of the Brain to Developmental Lead Exposure. *Toxicological Sciences, 141,* 29–43.

Semlali, R. M., Dessogne, J. B., Monna, F., Bolte, J., Azimi, S., Navarro, N., … van Oort, F. (2004). Modeling Lead Input and Output in Soils Using Lead Isotopic Geochemistry. *Environ. Sci. Technol*, *38*, 1513–1521.

Shinkai, Y., Yamamoto, C., & Kaji, T. (2010). Lead induces the expression of endoplasmic reticulum chaperones GRP78 and GRP94 in vascular endothelial cells via the JNK-AP-1 pathway. *Toxicological Sciences*, *114*(2), 378–386.

546 Shultz, A. J., Baker, A. J., Hill, G. E., Nolan, P. M., & Edwards, S. V. (2016). SNPs across time
547 and space: population genomic signatures of founder events and epizootics in the House
548 Finch (*Haemorhous mexicanus*). *Ecology and Evolution*, *6*, 7475–7489.

549 Smith, M. H., Ploegh, H. L., & Weissman, J. S. (2011). Road to Ruin: Targeting Proteins for
550 Degradation in the Endoplasmic Reticulum. *Science*, *334*(6059), 1086–1090.

551 Song X Bin, Liu G, Liu F, Yan ZG, Wang ZY, Liu ZP, Wang L. (2017). Autophagy blockade
552 and lysosomal membrane permeabilization contribute to lead-induced nephrotoxicity in
553 primary rat proximal tubular cells. *Cell death & disease, 8*, e2863.

554 Sriram, S. M., Kim, B. Y., & Kwon, Y. T. (2011). The N-end rule pathway: emerging functions
555 and molecular principles of substrate recognition. *Nat Rev Mol Cell Biol*, *12*(11), 735–747.

556 Tang, Y., Simoneau, A. R., Liao, W. -x., Yi, G., Hope, C., Liu, F., … Zi, X. (2009). WIF1, a
557 Wnt pathway inhibitor, regulates SKP2 and c-myc expression leading to G1 arrest and
558 growth inhibition of human invasive urinary bladder cancer cells. *Molecular Cancer*
559 *Therapeutics*, *8*(2), 458–468.

560 Taniguchi, M., Fukunaka, A., Hagihara, M., Watanabe, K., Kamino, S., Kambe, T., …
561 Hiromura, M. (2013). Essential Role of the Zinc Transporter ZIP9/SLC39A9 in Regulating
562 the Activations of Akt and Erk in B-Cell Receptor Signaling Pathway in DT40 Cells. *PLoS*
563 *ONE*, *8*(3), e58022.

564 Taylor, M. P. (2015). Atmospherically deposited trace metals from bulk mineral concentrate port
565 operations. *Science of the Total Environment*, *515–516*, 143–152.

566 Taylor, M. P., Davies, P. J., Kristensen, L. J., & Csavina, J. L. (2014a). Licenced to pollute but

not to poison: The ineffectiveness of regulatory authorities at protecting public health from atmospheric arsenic, lead and other contaminants resulting from mining and smelting operations. *Aeolian Research*, *14*, 35–52.

Taylor, M. P., Mackay, A. K., Hudson-Edwards, K. A., & Holz, E. (2010). Soil Cd, Cu, Pb and Zn contaminants around Mount Isa city, Queensland, Australia: Potential sources and risks to human health. *Applied Geochemistry*, *25*(6), 841–855.

Taylor, M. P., Mould, S. A., Kristensen, L. J., & Rouillon, M. (2014b). Environmental arsenic, cadmium and lead dust emissions from metal mine operations: Implications for environmental management, monitoring and human health. *Environmental Research*, *135*, 296–303.

Teixeira, L. K., & Reed, S. I. (2013). Ubiquitin Ligases and Cell Cycle Control. *Annual Review of Biochemistry*, *82*(1), 387–414. doi:10.1146/annurev-biochem-060410-105307

Varian-Ramos, C. W., Swaddle, J. P., & Cristol, D. A. (2014). Mercury reduces avian reproductive success and imposes selection: An experimental study with adult- or lifetime-exposure in zebra finch. *PLoS ONE*, *9*(4), e95674.

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. *The American Naturalist*, *186*, S24–S36.

Wikelski, M., & Cooke, S. J. (2006). Conservation physiology. *Trends in Ecology and Evolution*, *21*(1), 38–46. doi:10.1016/j.tree.2005.10.018

Wikelski, M., Wong, V., Chevalier, B., Rattenborg, N., & Snell, H. L. (2002). Marine iguanas

die from trace oil pollution. *Nature*, *417*(6889), 607–608.

Yang, T., Zhang, H., Qiu, H., Li, B., Wang, J., Du, G., … Wan, X. (2016). EFEMP1 is repressed by estrogen and inhibits the epithelial- mesenchymal transition via Wnt/β-catenin signaling in endometrial carcinoma. *Oncotarget*, *7*(18), 25712–25725.

Yariz KO et al. (2012). Mutations in OTOGL, encoding the inner ear protein otogelin-like, cause moderate sensorineural hearing loss. *American Journal of Human Genetics, 91*, 872–882.

Figures

Figure 1. Map of sampling locations. The locations affected by mining with high levels of lead contamination are marked in orange. The area across Australia where the species has been observed in shaded in grey.

Figure 2. Manhattan plots for BayeScEnv models. a) shows the results for the Predicted model and has an interesting peak on chromosome chr1A with 5 significant SNPs that are very close to each other. The Predicted model scored most of the locations as having normal levels of lead except for Mount Isa and Broken Hill which were allocated high environmental lead levels. b) plots the results from the Standard model which used estimates of environmental differentiation for lead contamination using only non-mining locations to define a standard environment and c) plots the All sites model (see Table 1). These plots display the log transformed q-values and SNPs above the dashed line are significant.

Figure 3. Venn Diagram for the overlap between the significant SNPs from the three models.

Table 1
Click here to download Table: Table_1_STOTEN_Lead_180828.docx

Table 1. Mean lead data and Environmental differentiation calculations. The "Environmental differentiation" (ED) values were used for BayeScEnv analyses. The conservative threshold for safe lead levels in areas with urban dwelling is 300 (mg/kg) (NEPM 2013). The two locations above the threshold are heavily affected by mining (Mount Isa and Broken Hill). The Standard model ED was calculated using the mean (112 mg/kg) and standard deviation (56 mg/kg) of mean lead contamination at non-mining sites. We scaled these values by dividing by 3 to account for very high lead levels in the two mining locations (Standard ED = ((site value – mean)/SD)/3). For the All sites model ED was calculated using the mean (286 mg/kg) and standard deviation (435 mg/kg) of mean lead contamination of all 11 sites (All sites ED = (site value – mean)/SD).

Pop No.	Location	State	Test	Mean Soil lead (mg/kg)	Citation	Standard model ED	All sites model ED	Comments/Notes
1	Hobart	TAS	Soil lead less than 2 cm	92	(Olszowy et al., 1995)	-0.119	-0.446	See Table 4 data in Olszowy *et al.* (1995) for Hobart.
2	Melbourne	VIC	Soil lead mean	173	(Laidlaw, Gordon, & Ball, 2018)	0.363	-0.260	Similar estimate to Olszowy *et al.* (1995) for Victoria.
3	Mt Gambier	SA	Soil (old suburbs, low traffic)	54	(Olszowy et al., 1995)	-0.345	-0.533	Used mean for low traffic areas in SA no data for Mount Gambier specifically.
4	Adelaide	SA	Soil (old suburbs, high traffic)	144	(Olszowy et al., 1995)	0.190	-0.326	Air and soil lead are also relatively low in Kristensen *et al.* (2016) for the wine regions of McLaren Vale which is where we sampled this population.
5	Broken Hill	NSW	Soil lead mean	1500	(Kristensen & Taylor, 2016)	8.262 (3)	2.791	Similar in Dong *et al.* (2015). Background historic levels ca. 100 mg/kg.
6	Sydney	NSW	Soil lead mean	210	(Laidlaw et al., 2017)	0.583	-0.175	Similar in other studies that find about 50% of samples above threshold.
7	Armidale	NSW	Lithgow average	46	(Rouillon, Gore, & Taylor, 2013)	-0.393	-0.552	We used data from the regional centre of Lithgow which is similar to background levels across NSW (Olszowy et al., 1995;

								Rouillon et al., 2017).
8	Toowoomba	QLD	Soil (old suburbs, low traffic)	79	(Olszowy et al., 1995)	-0.196	-0.476	Used QLD mean for old suburbs with low traffic.
9	Townsville	QLD	Soil lead	128	(Mark Patrick Taylor, 2015)	0.095	-0.363	Mean from supplementary Table 4.
10	Longreach	QLD	Soil (old suburbs, low traffic)	79	(Olszowy et al., 1995)	-0.196	-0.476	Used QLD mean for old suburbs with low traffic.
11	Mount Isa	QLD	Soil lead mean	638	(Mackay et al., 2013; Taylor et al., 2010)	3.131 (3)	0.809	Taylor (mean 346, n = 60, < 180 µm grain size) and Mackay (mean = 1560, n = 19) were averaged to get city wide data. Mackay was focused on mining sites.

Table 2
Click here to download Table: Table_2_STOTEN_Lead_180828.docx

Table 2. Summary of significant SNPs from BayeScEnv models. From BayeScEnv significant outliers, due to ecological association with the environmental variable, can be assessed using q-values or the more conservative PEP (Posterior Error Probability) statistic. The Predicted model assumed all localities had average lead levels, score of 0, except for Broken Hill and Mount Isa that had high lead scores of 3. The Standard model used the estimates of environmental differentiation using the mean of non-mining sites and the All sites model uses the mean from all sites (Table 1). We present three models because there is no optimal way to summaries lead pollution with one statistic.

Analysis	Total SNPs	No. SNPs linked to genes	No. of genes linked to SNPs	Percentage (%) of SNPs linked to genes	Percentage of SNPs within genes	Percentage of SNPs within exons
Predicted model – q-value	19	6	9	31.6	10.5	0
PEP	11	3	5	27.3	9.1	0
Standard model – q-value	10	3	5	30	0	0
PEP	3	2	3	66.7	0	0
All sites model – q-value	40	17	27	42.5	2.5	0
PEP	22	9	15	40.9	0	0

Figure 1

Click here to download Figure: Fig_1_Lead_map_9x7.pdf

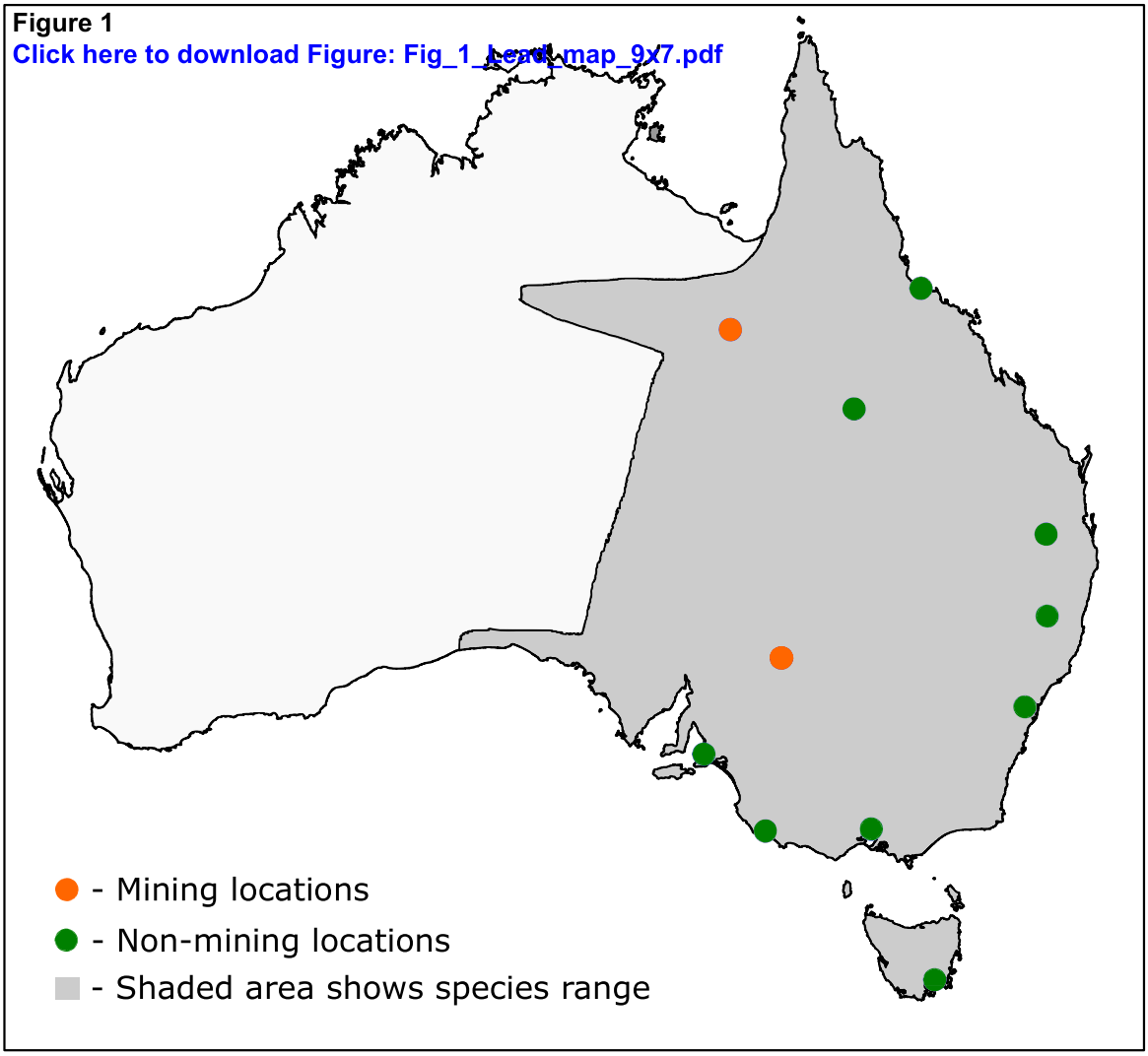

Figure 2

Click here to download Figure: Fig_2_Manhattan_NewX3_180820.pdf

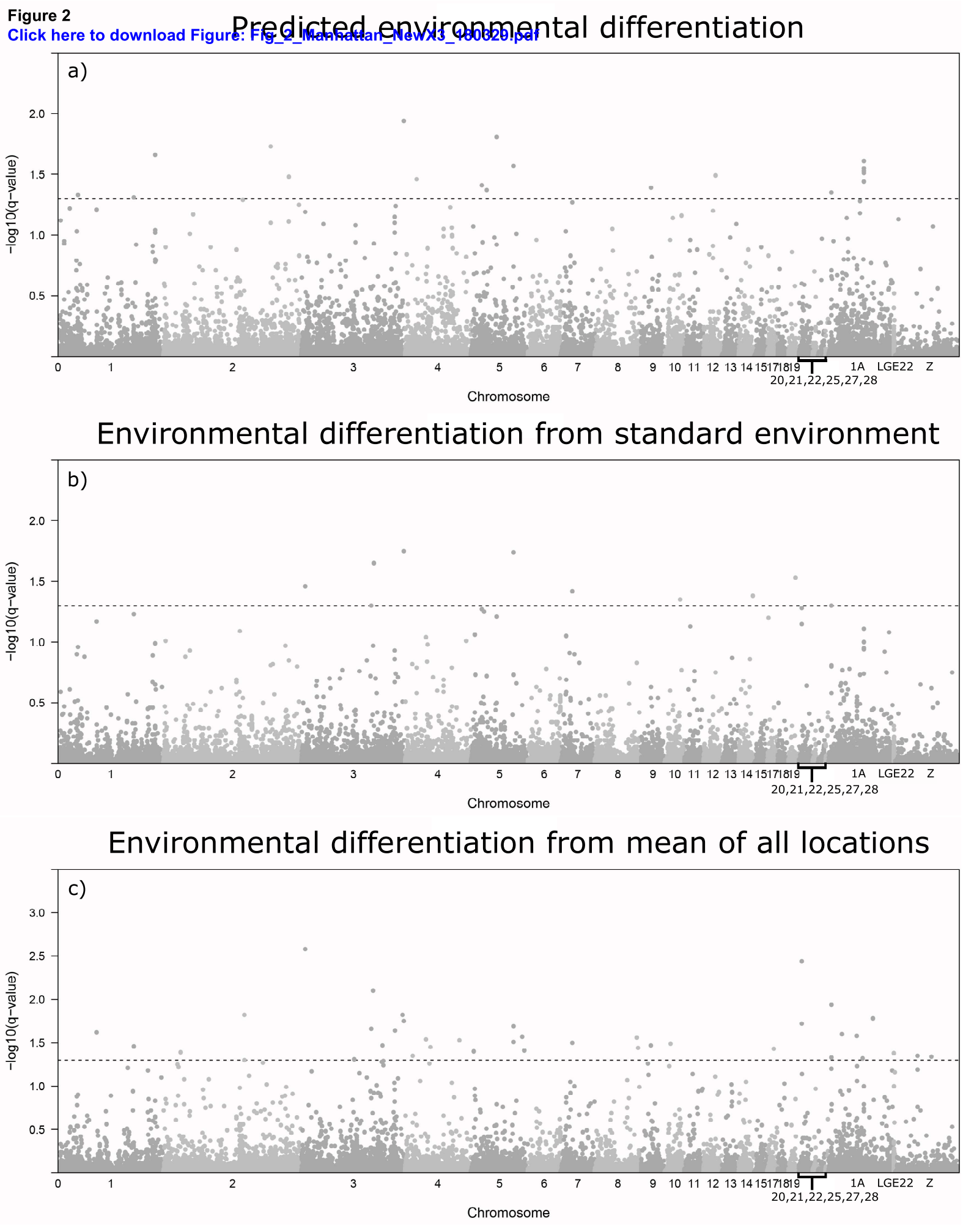

Figure 3

Click here to download Figure: Fig_3_Venn_3_NEW_180326.pdf