This is a repository copy of DNA sequence variation and methylation in an arsenic tolerant earthworm population.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/75061/

Version: Submitted Version

**Article:**
Kille, Pete, Andre, Jane, Anderson, Craig et al. (10 more authors) (2013) DNA sequence variation and methylation in an arsenic tolerant earthworm population. Soil Biology and Biochemistry. pp. 524-532. ISSN 0038-0717

https://doi.org/10.1016/j.soilbio.2012.10.014

**Reuse**
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
DNA sequence variation and methylation in an arsenic tolerant earthworm population

Peter Kille¹#, Jane Andre¹,²#, Craig Anderson¹,³, Hui Na Ang²,³, Michael W. Bruford¹, Jacob G. Bundy⁴, Robert Donnelly¹, Mark E. Hodson²†, Gabriela Juma¹, Elma Lahive³, A. John Morgan¹, Stephen R. Stürzenbaum⁵, David J. Spurgeon³*

# PK and JA should be considered joint first authors.

¹ Cardiff School of Biosciences, BIOSI 1, University of Cardiff, P.O. Box 915, Cardiff, CF10 3TL, UK.

² Soil Research Centre, School of Human and Environmental Sciences, University of Reading, Reading, RG6 6DW, Berkshire, UK.

³ Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Wallingford, Oxfordshire OX10 8BB, UK.

⁴ Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, South Kensington, London.

⁵ King’s College London, Analytical and Environmental Sciences Division, 150 Stamford Street, London SE1 9NH, UK.

* Author for correspondence. Full contact details: Dr David Spurgeon, Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Wallingford, Oxfordshire, UK, OX10 8BB, Tel: + 44 1491 772 208; Fax: +44 1491 692 424; Email: dasp@ceh.ac.uk

† Current address: Environment Department, University of York, Heslington, York, YO10 5DD

Running title: DNA variation and methylation in an arsenic tolerant earthworm population
Abstract

Evidence is emerging that earthworms can evolve tolerance to trace element enriched soils. However, few studies have sought to establish whether such tolerance is determined through adaption or plasticity. Here we report results from a combined analysis of mitochondrial (cytochrome oxidase II, COII), nuclear (amplified fragment length polymorphism, AFLP) variation and DNA methylation in populations of the earthworm Lumbricus rubellus from sites across an abandoned arsenic and copper mine. Earthworms from the mine site population demonstrated clear arsenic tolerance in comparison to a naïve strain. COII and AFLP results suggest that L. rubellus from the unexposed and the adapted populations comprises two cryptic lineages (Linages A and B) each of which was present across all of the sites. AFLP analysis by lineage highlighted variations associated with soil metal/metalloid concentrations (most clearly for Lineage A) suggesting a genetic component to the observed tolerance. The methylation sensitive AFLP (Me-AFLP) identified a high genome methylation content (average 13.5%) in both lineages. For Lineage A, Me-AFLP analysis did not identify a strong association with soil arsenic levels. For Lineage B, however, a clear association of methylation patterns with soil arsenic concentrations was found. This suggests that Lineage B earthworms utilise epigenetic mechanisms to adapt to the presence of contamination. These fundamentally different genetic adjustments in the two clades indicate that the two lineages employ distinct adaptive strategies (genetic or epigenetic) in response to arsenic exposure. Mechanisms driving this variation may be founded within the colonisation histories of the lineages.

Keywords: arsenic tolerance, cryptic lineages, adaptive variation, DNA methylation, epigenetics
1. Introduction

Many invertebrate species are able to maintain viable populations in polluted soils where total and potentially bioavailable metal/metalloid concentrations greatly exceed toxicity values (e.g. LC$_{50s}$) for known naïve (and so sensitive) populations (laboratory strains). This suggests that under trace element exposure, some invertebrate populations develop metal tolerance through behaviour or physiological adaptive traits (Posthuma and Van Straalen, 1993; Van Straalen and Roelofs, 2007). Mechanisms underpinning this tolerance have in some cases been shown to involve heritable changes in coding or promoter regions of metal efflux pumps (Callaghan and Denny, 2002) and thiol-rich peptides involved in sequestration (glutathione-S-transferases, phytochelatins, and metallothioneins) (Janssens et al., 2007; Vatamaniuk et al., 2005). In other cases, however, the mechanisms underlying tolerance remain unknown and/or unstudied.

For earthworms, one of the most functionally important of soil taxa (Lavelle et al., 1997), indirect evidence for metal tolerance is provided by the fact that earthworms can be collected from soils containing residue levels that significantly exceed toxic effect concentrations (Spurgeon and Hopkin, 1999a, b). However, difficulties in extrapolating toxicity data between the laboratory and field due to, for example, contaminant aging and speciation (Arnold et al., 2003; Arnold et al., 2007; 2010). More directly in relation to tolerance, studies with successive generations of Eisenia fetida selected for tolerance to Zn over two generations found changes in the shape of concentration response relationships for survival that were indicative of tolerance development (Spurgeon and Hopkin, 2000). In the field, Langdon et al. (1999) noted that L. rubellus living in arsenic and copper polluted soil at two abandoned arsenic mines (Devon Great Consols, Carrock Fell) could survive in arsenic-spiked soil that was acutely toxic to earthworms from a clean site. This tolerance was preserved when the mine populations were reared on clean soil over two generations, suggesting a genetic basis for this phenotype (Langdon et al., 2009).

Despite indications of trace metal and metalloid tolerance in earthworms, the extent to which there is a genetic and/or physiological basis of this trait has not been fully investigated. A study of isozyme specific polymorphisms within L. rubellus populations known to be adapted to combined metal and flooding stress...
failed to identify adaptive variation \cite{Simonsen2010}, although the results of this study should be treated with some caution as enzymes known to be related to metal tolerance were not targeted. The only study that has, to date, identified a potential genetic basis for tolerance to adverse soil conditions in earthworms is that for L. rubellus living at a lead/zinc mine located at Cwmystwyth, mid Wales. For this population, Andre et al. \cite{Andre2010} used mitochondrial (COII) and amplified fragment length polymorphism (AFLP) genotyping to demonstrate that a mine spoil associated population showed little genetic overlap (in AFLP profile) with individuals within populations at two less polluted sites.

While the assumption often is that individual/population survival is based on selection for increased tolerance, there is evidence emerging that the plastic responses driven by chemical influences on the epigenome may also be an important mechanism of adaptation \cite{Mirouze2011,Ren2011,Seong2011}. Among the many epigenetic mechanisms, DNA methylation represents a key response. Given that earthworms have been recorded to possess a 13% methylated cytosine content in DNA \cite{Regev1998}, the potential for mediation of adaptive tolerance through epigenetic DNA methylation should be considered. Here, we report a combined toxicological and genetic study, using mitochondrial (COII) and nuclear (AFLP) and DNA methylome analysis, for the earthworm L. rubellus sampled at sites of different metal pollution status within an As-contaminated mine complex - the Devon Great Consols (DGC) site in the UK. For the study, we sampled earthworms from a number of sites within DGC including one (Site 2 in this study) from which tolerant populations previously studied by Langdon et al. \cite{Langdon2009,Langdon1999} were collected. Adjacent and distant reference sites were also sampled. That arsenic, a major contaminant at the site, has been reported to induce epigenetic changes including hypo-and hyper-methylation of DNA \cite{Ren2011,Zhong2001} makes the site particularly suited for the analysis of DNA methylation responses in earthworms. Initially the collected populations were screened to confirm that the tolerance previously reported for populations at study Site 2 was applicable to earthworms inhabiting this and other collection sites located in the mining area. Genetic analyses were then undertaken using these populations. The hypothesis tested was that L. rubellus populations from polluted sites that show evidence of tolerance, would include individuals with mitochondrial or nuclear genotypes and/or DNA methylation patterns that were distinct from those of intolerant earthworms from (adjacent) unpolluted locations.
2. Materials and Methods

2.1 Site description, sampling and soil characterisation: This study was conducted at the abandoned Devon Great Consols mine complex located in the Tamar Valley, Devon, South-West UK (UK Ordnance Survey. Map coordinates for mine centre: SX426733 – N50:32:52 W4:13:25). This mine was worked for copper and arsenic from 1844-1900 and from 1915-1930. Across the site, the spoil from various extraction processes remain. The soils established on these wastes contain highly elevated concentrations of trace elements, including arsenic and copper. Earthworms (L. rubellus) were sampled from six locations in the region of the Devon Great Consols mine. Four locations (Sites 1-4) were situated on the mine and waste handling area (see Fig. 1). This included a location (Site 2 i.e. close to the area where arsenic was processed using the calciner method) from which the adapted population studied by Langdon et al. [2009, 1999] was collected. Two clean reference site populations were also sampled. These were at a site adjacent to the contaminated area, but which itself was not greatly enriched in arsenic and copper (Site Control - SC) and a site some 20 km distant from DGC which was outside the geological area of arsenic rich soils present in the Tamar Valley (Off-Site Control - OSC) (UK Ordnance Survey. Map coordinates SX 418901 N50:68:89 W4:24:03).

At each site, approximately 30 fully clitellate adult L. rubellus were collected by digging and hand-sorting over two consecutive days in September 2010. To ensure that genome methylation patterns were not influenced by handling stress, all earthworms were washed and blotted dry on-site and then snap frozen in liquid nitrogen. Triplicate soil samples from surface to 5 cm depth were also collected from each location. These were subsequently oven dried at 80°C and sieved through a 2 mm mesh to remove large roots and stones. Total concentrations of arsenic, barium, calcium, cadmium, chromium, copper, iron, magnesium, nickel, strontium and zinc were determined in a 1 g sample of the processed soil following an aqua regia digestion protocol [Arnold et al., 2008]. Digests were analysed on a Perkin Elmer Optima 7300 DV inductively coupled plasma optical emission spectrometry instrument. For quality control, an in house reference traceable to BCR-143R (Commission of the European Communities, Community Bureau of Reference) was included with each batch of digestions. Measured concentrations were always greater than 75% of reference values and were above 95% for As. Organic matter content of each soil sample was
measured by proxy using loss on ignition following combustion at 500°C [Rowell, 1994] and pH was quantified by electrode from a 1:5 volume soil:water mix [International Standards Organisation, 2005].

2.2. Toxicity tests to identify putative arsenic tolerance

To identify potential tolerance, a 14 day exposure to a single pre-determined arsenic concentration was undertaken to compare survival patterns of earthworms from sites located within and adjacent to the DGC complex to those for a known naïve population. The soil concentration used for this assay was derived from a preliminary study conducted to assess survival of the naïve population at 150 and 300 mg/kg arsenic. The earthworms used were taken from a culture established from a field collected population (Lasebo BV, Nijkerkerveen, The Netherlands). At each tested concentration, 15 replicate containers, each including 200 g dry weight of a clay loam soil (Broughton Loams, Kettering, UK) [see Spurgeon et al., 2003], were spiked with sodium arsenate solution (Santa Cruz Biotechnology Inc., Santa Cruz, California, US) to give the required metalloid concentration and a soil moisture content of approximately 50% of field capacity. After a one week stabilisation period, one adult L. rubellus was added to each replicate and kept at 13 ± 1 °C under constant light for seven days. Earthworms were observed daily and mortality recorded. Based on these findings, a screening concentration of 300 mg/kg arsenic was selected for the definitive tolerance assay, since this concentration resulted in progressive mortality of the naïve earthworms over the exposure period. Thus the definitive assay was conducted using the 300 mg/kg concentration with 15 earthworms from each of the DGC sites (Sites 1-4) and the SC reference population. The exposure was extended to 14 days to allow the potential to identify survival patterns in more tolerant populations.

2.3 Mitochondrial cytochrome oxidase II (mtCOII)) sequencing:

DNA was purified from ~10 mg of tissue from the anterior of each individual using the DNAzol reagent (Life Technologies, Paisley, UK). PCR amplification of the cytochrome oxidase II (COII) mitochondrial gene made use of forward (TAGCTCACCTTAGATGCCA) and reverse (GTATGCGGATTTCTAATTGT)
primers and was conducted following Andre et al. (2010). PCR products were assessed electrophoretically prior to purification and sequencing using ABI PRISM® BigDye v3.1 Terminator technology (Applied Biosystems, USA). Obtained sequences were aligned by ClustalW prior to tree construction using the Maximum Likelihood (ML) and Bayesian methods in Mega v5.01 and MRBAYES v3.2, respectively. ML estimation incorporated the Tajima-Nei model, supported by bootstrap analyses over 1000 iterations. Bayesian analyses were conducted using a General Time Reversible model with a proportion of invariable sites and a gamma-shaped distribution over 2 independent runs. Four Markov Chains were run over 2 million iterations and sampled every 1000 generations, with the first 500 trees discarded as burn-in. Both phylogenetic estimates incorporated outlier sequences from Lumbricus castaneus and Lumbricus terrestris as well as sequences that represent previously recognised L. rubellus clades (Andre et al., 2010).

2.4 AFLP and methylation sensitive AFLP profiling: A combined AFLP and Me-AFLP protocol was optimised in a pilot methylation study and was based on parallel use of methylation- and non-methylation-sensitive restriction enzymes (HapII and MspI) to treat DNA samples prior to primer ligation and amplification (Xiong et al., 1999). Both HapII and MspI recognize a CCGG sequence; however, while MspI is able to cut methylated recognition sites (as well as unmethylated ones), HapII is unable to cut at such locations when they are methylated (i.e. only unmethylated recognition sites are cut). The extent of methylation of restriction sites can therefore be ascertained by recording bands amplified by MspI but not HapII. Such bands can be used to compare individual methylation patterns. AFLP analysis was conducted for individuals from the six collection locations using pre-selective primers and analysis on an Applied Biosystems 3130 x 1 fragment analyser (Andre et al., 2010). Cumulative AFLP fragment profiles were transformed to a binary form and principal coordinates (PCO) analysis used to visualise the genetic relationship between individuals using GenAlEx 6.4.1.
3. Results

Soil analyses highlighted the extent and severity of the arsenic (and copper) contamination at DGC. Arsenic and copper levels were greatly elevated in soils from all sites on the mining area (Sites 1-4), with cobalt also higher than SC and OSC soils by at least a factor of two at the sites (Table 1). The most polluted arsenic soil (Site 4) contained almost 20,000 mg/kg of arsenic, over 900 mg/kg copper and also elevated cobalt, cadmium and lead levels. The remaining three mine spoil contaminated sites each contained over 4000 mg/kg As and over 500 mg/kg copper. As expected, the lowest concentrations of arsenic and copper and other trace metals were found at the SC and OSC reference sites. Levels at SC were in the 300 mg/kg As range, a concentration still elevated above background arsenic levels for British soils [Emmett et al., 2010]. OSC soils contained arsenic levels consistent with these background concentrations.

Measured site soil LOI and pH values are presented in Table 1. Whilst the pH of all four sites located on the mine area and the OSC was similar acidic (pH 4.1 - 4.8), the SC site had a pH of 5.6. Overall, there was minimal pH variation between sites, and no correlation with soil arsenic or copper concentration (Pearson correlation, p > 0.05). The absence of a correlation indicates that soil pH influences are unlikely to confound attempts to link genetic variation to soil contaminant levels. For LOI, the lowest values in the mining site soils (4.2 - 17.2%) were found at Sites 1 and 4, while the remaining two soils had higher LOI values (29.7 – 49.6%). This may reflect the vegetation of the sites: open in the case of Sites 1 and 4, wooded at Sites 2 and 3. The two control site soils had %LOI levels intermediate within the range of the two pairs of mine sampling locations.

Exposure of the naïve population to 300 mg/kg of arsenic in soil resulted in a progressive mortality, culminating in only 7% survival after 14 days of exposure (Fig. 2). In the SC population progressive mortality over time was also seen. This, however, proceeded at a slower rate than for the naïve earthworms, culminating in 46% survival after 14 days. These variations in mortality rates resulted in different LT₅₀ estimates from Weibull models fits (SigmaPlot 12.0) for the naïve and SC populations; these being 5.3 (95% Confidence Intervals 4.9-5.6) and 12.4 (95% Confidence Intervals 11.6-13.3) days respectively. In the four DGC mine site populations there was observable mortality in the Site 1 population, although 73% survival
after 14 days was higher than for either the naïve or SC earthworms. Populations from the remaining three
DGC mine sites show low mortality, with 100% survival for Site 2 and 4 earthworms and 85% survival for
Site 3 earthworms.

The mtDNA COII analysis indicated the presence of two distinct lineages (A and B) within the sampled L.
rubellus (Fig. 3a). The two cryptic lineages show a 18% and 14% genetic divergence from L. castaneus and
L. terrestris respectively. Average difference between lineages was 10.3%. Internal within the lineages, the
Lineage A earthworms have a maximum 1.4% genetic difference, while for Lineage B earthworms this was
0.06%. This high level of divergence between the two major lineage branches identifies L. rubellus as a
complex of cryptic lineages as found previously [Andre et al., 2010]. A comparison of the frequency of
lineage occurrence at each sampled site identified differences in population COII haplogroup composition.

Populations at two sites, Site 4 and Site OSC, included 90% or more of individuals from Lineage B; while in
contrast the Site 3 populations included 76% of Lineage A individuals. The remaining three sites each had an
approximately equal proportion of each lineage, with Lineage A slightly dominant at Site SC (57%) and Site
1 (54%) and Lineage B at Site 2 (64%). That both lineages were found at all sites, often in similar
proportions, and also that the two sites showing greatest lineage selection (dominance of Lineage B at both
Site 4 and Site OSC) included both the most and least arsenic contaminated soils, is indicative of an absence
of a mitochondrial lineage association with soil contamination status.

Standard AFLP analysis conducted using MspI (which cuts at all recognition sites independent of
methylation status) reemphasised the presence of two L. rubellus lineages as indicated by the mitochondrial
markers. All mine site earthworms fell clearly into one of the two major lineages, but apparent inter-lineage
individuals were observed among SC and OSC earthworms. These hybrids show AFLP genotypes
intermediate between the two lineages on PC1 and divergent on PC2 (Fig. 3b). The presence of hybrids is in
agreement with previous observation of AFLP profiles in L. rubellus [Andre et al., 2010]. The dominance of
the lineage effects within a PCO analysis of the AFLP data meant it was not possible to visualise site effects
within the complete data-set. Consequently independent lineage-based analyses were conducted (n.b.
putative hybrid individuals were excluded from these analyses).
Within Lineage A, PCO highlighted a site dependent effect on AFLP marker patterns. Within the PC1 and PC2 scores plot, SC earthworms were clearly separated from earthworms collected from Sites 2 and 3, with the Site 1 individuals intermediate and closer to the SC earthworms (Fig. 4a). Scores for PC2 (and also PC3), but not PC1 within the PCO were significantly correlated with site soil arsenic concentration (Pearson correlation p < 0.01). This significant association suggests that soil arsenic concentrations, as well as possibly the concentration of other co-correlated metals such as Cu, are an important driver of genome structure in Lineage A L. rubellus across the DGC site and surrounding area.

For Lineage B PCO analysis did not identify a separation of populations within a PC1 and PC2 scores plot, although a partial separation of Site 1 and 4 was evident (data not shown). Both of these populations, however, overlap with profiles from the SC and OSC earthworms within this plot. Correlation of PC1 and PC2 scores with soil arsenic concentrations were not significant. Only for the PC3 score was a significant correlation found (Pearson correlation p < 0.02) indicating a weak separation underpinned by the distribution particularly of the Site 2 and Site 4 individuals on this component (Fig. 4b). These results identify that while soil metals such as arsenic and correlated elements are a driver for genome structure in Lineage B, these factors are less important than for Lineage A with effects only observed for the lower contribution PCs.

To assess the patterns of genome methylation in individual earthworms, a second AFLP analysis was conducted using the MspI methylation sensitive restriction enzyme. Me-AFLP indicated that the genome of L. rubellus had an approximate 13.5% methylated cytosine (m5C) residue content. Across the mine sampling locations, the average extent of genome methylation ranged from 10.6% in earthworms at Site 1 to 22.1% for Site 4. Even though the highest average genome methylation content was at the most arsenic polluted site, the fact that earthworms from the two reference sites had intermediate average methylation levels (SC 19.4%, OSC 13.2%) meant there was no clear correlation (Pearson correlation p > 0.05) between methylation level and soil arsenic concentration. This suggests that in the mine soils, arsenic does not have a strong global hyper- or hypo-methylation effect for the resident earthworms.
Within the Me-AFLP analysis the presence of two distinct L. rubellus lineages was reconfirmed. Consequently, lineage-specific Me-AFLP profiles were analysed, with the hybrid individuals excluded. For the Lineage A PCO analysis, a segregation of individuals collected from Site 2 and Site 4 was identified within the PC1 and PC2 score plot (Fig. 4c). The remaining sites showed substantial overlap between individual profiles. Correlations of PC1, PC2 and PC3 scores with measured soil arsenic concentration were non-significant in all cases (Pearson correlation $p > 0.05$). This suggests that soil arsenic was not the principal driver of methylation pattern difference between individuals. For Lineage B, there was a partial separation of profiles of earthworms from the SC and OSC locations from individuals collected from each sampled mine site population (Fig. 4d). Correlation of PC1, PC2 and PC3 scores with measured site soil arsenic concentration indicated a significant correlation for the first principle component (Pearson correlation $p < 0.02$). This indicates that soil arsenic (and co-correlated trace metals) represents a potentially significant driver of earthworm genome methylation status for lineage B earthworms.
4. Discussion

Soil contamination by mineral extraction, fossil fuel consumption, waste disposal and pesticide use is a common problem [Hall et al., 2006]. Among trace elements, arsenic represents one of the greatest hazards because of its widespread distribution and toxicity to humans and wildlife [Chen et al., 2009; Thomas et al., 2001]. The toxicity of arsenic has been established for earthworms. Meharg et al. [1998] determined an LC$_{50}$ of approximately 100 mg/kg As for Lumbricus terrestris after 8 days and Fischer and Koszorus [1992] found that a 25 mg/kg potassium arsenate exposure reduced growth and cocoon production in Eisenia fetida. For L. rubellus, Langdon et al. [2001] found an LC$_{50}$ of 96 mg/kg As for a clean site population, although populations from mine sites (including DGC) gave higher values (up to 1,510 mg/kg) suggesting tolerance.

Building on this work, Langdon et al. [2009] revealed that the adaptation in the mine site earthworms was maintained when earthworms were bred for two generations on clean soil. Cross-tolerance to copper was also found [Langdon et al., 2001].

In the test to assess the presence of potential tolerance in L. rubellus collected from the DGC line complex area sites, there was a clear indication that the populations inhabiting the DGC site locations substantially enriched in arsenic display a tolerance phenotype. Earthworms sampled from the populations at Site 1-4 all showed low mortality on exposure to a soil arsenic concentration that induced acute toxicity in earthworms from a naïve population and also in the SC reference population. Interestingly the different rates of mortality in naïve and SC earthworms, as highlighted by differences in LT50s for these populations suggest that SC earthworms possess a partial arsenic tolerant phenotype. This may be related to the presence of arsenic concentrations that greatly exceed accepted background concentrations in SC soil [Emmet et al., 2010].

Tolerance to chemical exposure can classically take two forms. Most simply, it can be the result of phenotypic plasticity. In this case, exposure to a substance upregulates biochemical pathways (e.g. metal binding proteins, mono-oxygenases and multi-drug resistance transporters), which work to detoxify or eliminate the substance. If the exposure is removed, upregulation of detoxification systems can persist, predisposing individuals to deal with a future chemical challenge. This plasticity has been widely reported in human subjects subjected to long-term drug exposure [Stewart and Badiani, 1993] and also in species...
exposed to toxicants in the field (Rajamohan and Sinclair, 2009; Romach et al., 2000). Maintenance of elevated protein levels and the widely reported effects of stressor exposure on the epigenome (Martinez-Zamudio and Ha, 2011), including arsenic (Ren et al., 2011), can provide a mechanism through which such tolerance may be temporally conserved.

A second mechanism of tolerance development exploits adaptive variation within populations. There is good evidence that this kind of adaptive selection can occur in response to long term chemical exposure. One example is driven by the selection of alleles coding for amino acids associated with active sites of detoxification enzymes. Pesticide resistance is frequently underpinned by this mechanism, with polymorphic cytochrome P450 genes often the selection target (Karunker et al., 2008; Miyo and Oguma, 2010). For metals, selection for metallothionein promoter alleles and other trans-acting genetic factors has been found to underpin cadmium tolerance in the collembolan Orchesella cincta (Janssens et al., 2007; Roelofs et al., 2006; van Straalen et al., 2011).

Characterisation of metallothionein promoter alleles of earthworms collected from metalliferous and unpolluted soils has so far failed to detect adaptive variation (Stürzenbaum et al., 2004). With evidence for targeted selection absent, a logical next step is to move to genome wide analysis (Baird et al., 2008; Hohenlohe et al., 2010). Using a combined approach applying mitochondrial genotyping and conventional and methylation-sensitive AFLPs, an analysis of both genotypic and epigenetic associations of the confirmed adapted and putative reference populations of L. rubellus with different metal/metalloid exposure histories was conducted. The aim was to assess the basis of the arsenic tolerance observed in the toxicity test. The mitochondrial genotyping and AFLP profiling (using both methylation sensitive and insensitive enzymes) all indicated that L. rubellus comprises two distinct lineages that differ by over 10% in their mitochondrial COII sequence. This reflects the presence of two cryptic lineages within the morphospecies (Andre et al., 2010). Hybrid individuals were found although only at the two reference sites. This prevalence in uncontaminated soils does not support a role of pollution in the breakdown of species boundaries as found by Vonlanthen et al. (2012).
As in a previous study with L. rubellus from polluted landscapes (Andre et al., 2010), there was no evidence of a lineage or intra-lineage haplotype association with either polluted or unpolluted sites. This supports the decision to move to a more detailed analysis of population structure. The AFLP analysis conducted for Lineage A indicated a clear separation of earthworms between sites, with the most important principal components associated with soil pollution status. For Lineage B, an influence of soil arsenic on AFLP profile was also found, albeit in this case within one of the more minor principle components (PC3). Such associations that link genetic distance to pollution status have been observed in previous field studies of aquatic invertebrates (Martins et al., 2009) and for both genetic units of the phylogeographically divergent metallophyte Arabidopsis halleri (Pauwels et al., 2012). Such relationships point to a genetic component that may underpin the previous observations of arsenic tolerance in L. rubellus collected at Site 3 by Langdon et al. (2009; 1999), especially given the high frequency of Lineage A individuals at this site.

Although sequence driven differentiation between populations can clearly be important, there is emerging evidence that epigenetic effects can also play a role in adaptation to local environmental conditions. Known epigenetic mechanisms include DNA methylation, histone modifications, and small interfering (siRNA), and micro RNAs (miRNA). Of these, DNA methylation has so far been most widely studied in animals (Suzuki and Bird, 2008). Studies have identified that metals and metalloids can perturb DNA methylation including hypomethylation by Cd (Takiguchi et al., 2003) and targeted gene silencing via hypermethylation by Ni (Lee et al., 1995). For arsenic, the potential competition with DNA for methyl groups for respectively methyl metabolites and DNA modification can create an interplay between hypomethylation (Arita and Costa, 2009; Zhao et al., 1997) and hypermethylation (Jensen et al., 2008) in arsenic toxicology (Ren et al., 2011).

To date relatively little is known about the role of DNA methylation as a component of adaptive variation in invertebrate organisms. Studies on a range of invertebrate species have highlighted extensive variation in the 5-methyl cytosine content of the genome (Regev et al., 1998). Thus, while some species, including the nematode Caenorhabditis elegans and fruitfly Drosophila melanogaster, have low to negligible 5-methyl cytosine levels (Bird, 2002; Regev et al., 1998), some taxa possess methylation levels in the 10-15% range. Me-AFLP indicated an approximate 13.5% methylated cytosine (m5C) residue content in the L. rubellus
genome. This represents a high level of DNA methylation for an invertebrate species, but is consistent with previous results for the earthworm Aporrectodea caliginosa [Regev et al., 1998]. This suggests that DNA methylation may have an important role in annelids, although to date relatively little is known about how such methylation is controlled. For example, a study on the marine annelid species Chaetopterus variopedatus was able to identify a protein that had a high homology to known invertebrate methyltransferases, but could not confirm a role of this protein in ‘de-novo’ methylation of double stranded DNA [del Gaudio et al., 1999].

On exposure to arsenic (and co-contaminant metals), an analysis of methylation patterns using the MeAFLP approach showed a site-specific influence. For Lineage A earthworms, separation between sites for the MeAFLP profiles was seen. This separation could not, however, be significantly associated with soil arsenic concentration as was the case for the standard AFLP analysis for this Lineage. This may indicate that other soil, biotic and local scale climatic factors may instead be modifying the epigenome. For Lineage B earthworms, pattern of DNA methylation could be significantly related to soil arsenic levels, suggesting a potential role of trace element exposure, although it is also feasible that environmental factors (e.g. soil texture, soil moisture, food availability), that are co-correlated to soil pollutant levels, could also be important. Evidence from detailed analysis of stress associated genes, such as metallothionein in the snail Helix pomatia, has already identified the presence of genomic regions that confer a high potential for epigenetic regulation indicating a potential role for epigenetic mechanisms in metal responses [Egg et al., 2009]. Further, in D. melanogaster stress exposure has been shown to result in epigenetic heterochromatic disruption that is transmissible in a non-Mendelian fashion [Seong et al., 2011]. The association of DNA methylation patterns with arsenic exposure observed here suggests a potential role of epigenetic mechanisms in stress adaptation in earthworms that concur with the evidence available for other taxa.

To extend the understanding of the role of genetic and epigenetic modification, a fruitful avenue for extending this novel study from a strong associative appreciation to a mechanistic understanding of arsenic-mediated molecular-genetic adaptations would entail assaying the transcription levels of specific genes known to be involved in metal/metalloid trafficking and metabolism. Moreover, establishing whether
Epigenetic marks are preferentially targeted to such genes and their regulatory regions in earthworms exposed to elevated levels of methylation-modifying arsenic in their native field soils is also a matter of priority. Such work clearly has the potential to link genotype to phenotype in adapted populations, so providing insight into the functional basis of adaptive traits in a key soil dwelling taxon.

The variation in lineage-specific responses observed across the genome and epigenome raises the intriguing prospect that the two L. rubellus cryptic lineages may employ different strategies to response to long-term arsenic exposure. The strong AFLP based separation of Lineage A earthworms in relation to soil arsenic concentrations across major PCs suggests that in this Lineage substantial genome modification has occurred as a result of long-term exposure. In contrast, the evidence for sequence modification is somewhat less compelling in Lineage B and therefore changes in genome methylation status seem to play a role in facilitating plasticity in response to soil arsenic concentration as indicated by the Me-AFLPs. Previous studies have identified differences in sensitivity between closely related lineages or species to chemical exposure. An example is the role of biotransformation capacity for determining the sensitivity of Capitella capitata “species” to PAH exposure. However, to date we are not aware of any studies that have identified such divergent genome responses to chemical exposure within two genetic lineages of a known morphospecies. The detailed basis for the evolution of distinct genetic and/or epigenetic mechanisms that drive arsenic adaptation in the two L. rubellus lineages, thus, emerge as potential models that could be further exploited to understand species plasticity in response to long-term chemical stress.

The genetic structure evident within the putative L. rubellus lineages is consistent with expectations in relation to survival within glacial refugia and subsequent recolonisation, as has been demonstrated for other species. Patterns of recolonisation during the Holocene, including recent human-mediated dispersal, may have resulted in different lineages reaching the DGC area over different timeframes. Andre et al. inferred that the two L. rubellus lineages have very different evolutionary histories with Lineage A representing a stationary population that has experienced multiple introductions and bottleneck episodes with expansion estimated to have occurred about 250,000 years BP, while Lineage B comprises an unimodal mismatch distribution with an estimated post-glacial population
expansion time of approximately 17,000 years BP. It is perhaps this differential in the timescale for adaptation to local arsenic contamination that has determined the lineage specific balance between adaptive variation and plasticity for the two lineages at sites across the DGC mine.
Acknowledgements

This work was partly supported by the Natural Environmental Research Council grant (NE/H009973/1). We thank Helen Hooper and Huw Ricketts and others for help with sample collection and Dr Claus Svendsen for discussion of the results. We also recognise the contribution of Mr and Mrs Morgan (Snr) for the production of AJM in 1948 and at a time of the year appropriate for earthworm sampling.
References


19


Spurgeon, D.J., Hopkin, S.P., 1999b. Tolerance to zinc in populations of the earthworm Lumbricus rubellus from uncontaminated and metal-contaminated ecosystems. Archives of Environmental Contamination and Toxicology 37, 332-337.


Table 1. Summary data for trace element concentrations, soil pH and wt% loss on ignition (%LOI) for soil samples collected from sites across the Devon Great Consols mine complex located in south-west England. For site locations see Fig. 1. Values are means of triplicate subsamples, standard deviations are given in brackets.

<table>
<thead>
<tr>
<th>Site</th>
<th>Al (mg/kg)</th>
<th>As (mg/kg)</th>
<th>Ba (mg/kg)</th>
<th>Cd (mg/kg)</th>
<th>Co (mg/kg)</th>
<th>Cr (mg/kg)</th>
<th>Cu (mg/kg)</th>
<th>Fe (mg/kg)</th>
<th>Mn (mg/kg)</th>
<th>Ni (mg/kg)</th>
<th>Pb (mg/kg)</th>
<th>Zn (mg/kg)</th>
<th>pH</th>
<th>%LOI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>9430</td>
<td>4620</td>
<td>47.2</td>
<td>0.7</td>
<td>19.5</td>
<td>10.8</td>
<td>529</td>
<td>48000</td>
<td>430</td>
<td>16.4</td>
<td>61</td>
<td>134</td>
<td>4.5</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>(6720)</td>
<td>(2020)</td>
<td>(32.4)</td>
<td>(0.4)</td>
<td>(11.5)</td>
<td>(9.8)</td>
<td>(266)</td>
<td>(24600)</td>
<td>(335)</td>
<td>(15.8)</td>
<td>(35)</td>
<td>(74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>6060</td>
<td>5220</td>
<td>70.0</td>
<td>&lt; 0.2</td>
<td>52.7</td>
<td>6.9</td>
<td>606</td>
<td>99200</td>
<td>802</td>
<td>28.4</td>
<td>191</td>
<td>164</td>
<td>4.1</td>
<td>49.6</td>
</tr>
<tr>
<td></td>
<td>(720)</td>
<td>(470)</td>
<td>(2.5)</td>
<td>(4.8)</td>
<td>(0.3)</td>
<td>(27)</td>
<td>(9300)</td>
<td>(96)</td>
<td>(2.8)</td>
<td>(21)</td>
<td>(22)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 3</td>
<td>17300</td>
<td>6270</td>
<td>45.9</td>
<td>0.18</td>
<td>25.7</td>
<td>17.8</td>
<td>2647</td>
<td>79600</td>
<td>630</td>
<td>22.3</td>
<td>225</td>
<td>277</td>
<td>4.8</td>
<td>29.7</td>
</tr>
<tr>
<td></td>
<td>(3900)</td>
<td>(1010)</td>
<td>(8.4)</td>
<td>(0.09)</td>
<td>(5.4)</td>
<td>(3.6)</td>
<td>(606)</td>
<td>(2600)</td>
<td>(135)</td>
<td>(4.0)</td>
<td>(53)</td>
<td>(43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 4</td>
<td>13600</td>
<td>19200</td>
<td>71.2</td>
<td>10.2</td>
<td>&lt; 3.6</td>
<td>20.0</td>
<td>910</td>
<td>65900</td>
<td>262</td>
<td>9.5</td>
<td>148</td>
<td>63</td>
<td>4.6</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>(2060)</td>
<td>(3470)</td>
<td>(8.9)</td>
<td>(1)</td>
<td>(3.3)</td>
<td>(3.0)</td>
<td>(120)</td>
<td>(10600)</td>
<td>(33)</td>
<td>(1.8)</td>
<td>(15)</td>
<td>(8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>21500</td>
<td>310</td>
<td>45.5</td>
<td>0.41</td>
<td>&lt; 3.6</td>
<td>31.7</td>
<td>107</td>
<td>45800</td>
<td>585</td>
<td>27.5</td>
<td>68</td>
<td>140</td>
<td>5.6</td>
<td>17.2</td>
</tr>
<tr>
<td></td>
<td>(1800)</td>
<td>(70)</td>
<td>(2.8)</td>
<td>(0.4)</td>
<td>(4.2)</td>
<td>(16)</td>
<td>(6350)</td>
<td>(103)</td>
<td>(9.1)</td>
<td>(8)</td>
<td>(34)</td>
<td>(34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSC</td>
<td>7840</td>
<td>&lt;50</td>
<td>37.0</td>
<td>&lt; 0.2</td>
<td>&lt; 3.6</td>
<td>10</td>
<td>14</td>
<td>1420</td>
<td>427</td>
<td>3.7</td>
<td>21</td>
<td>69</td>
<td>4.4</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>(4770)</td>
<td>(20.7)</td>
<td>(6.0)</td>
<td>(8)</td>
<td>(2.8)</td>
<td>(12)</td>
<td>(35)</td>
<td>(8650)</td>
<td>(234)</td>
<td>(2.8)</td>
<td>(35)</td>
<td>(35)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1: Aerial images showing the location of the Devon Great Consols mine site in the South West UK (top right insert panel) and locations of the 5 sampling locations (Site 1-4 and Site SC) situated in the area on, and immediately adjacent to, the Devon Great Consols mine workings.

Figure 2: Temporal patterns of survival of L. rubellus collected at five locations of contrasting geochemistry (4 polluted and 1 site reference) within the Devon Great Consols mine complex and surrounding area and a known naïve population following exposure to 300 mg/kg of arsenic in a spiked clay loam soil over 14 days.

Figure 3: Mitochondrial and nuclear analysis of L. rubellus population structure and corresponding mitochondrial mismatch distributions of collected L. rubellus. Panel A: shows a phylogenetic tree of mitochondrial COII genotype showing branching of major lineage (Left and right hand branches of the network are denoted Lineage B & A respectively) and the numbers of individuals from each site within the lineages. Panel B: AFLP multi-locus profiling PCO analysis showing individuals from the six sample stations. Lineage A individuals cluster to the right on PC1, Lineage B to the left. Hybrids (found at SC and OSC only) lie between and above the two Lineage Groups.

Figure 4: Nuclear genome analysis of L. rubellus collected at six sites (4 polluted and 2 reference) of contrasting geochemistry within the Devon Great Consols mine complex and surrounding area. Panel (i) shows the result of a PCO of AFLP profiles for L. rubellus unambiguously ascribed to Lineage A, Panel (ii) shows the result of a PCO of AFLP profiles for L. rubellus unambiguously ascribed to Lineage B, Panel (iii) shows the result of a PCO of methylation sensitive AFLP analysis of L. rubellus unambiguously ascribed to Lineage A, Panel (iv) shows the result of a PCO of methylation sensitive AFLP analysis of L. rubellus unambiguously ascribed to Lineage B.
Fig. 1.
Fig. 2.

Percentage surviving vs Exposure time (days)

- Site 1
- Site 2
- Site 3
- Site 4
- On-site control
- Naive

Site 1 to Naive represent different conditions or treatments.
Fig. 3.