Molecular genetic differentiation in earthworms inhabiting a heterogeneous Pb-polluted landscape

Article

Accepted Version

Molecular genetic differentiation in earthworms inhabiting a heterogeneous Pb-polluted landscape

J. Andre[1,2], R. A. King[1], S. R. Stürzenbaum[3], P. Kille[1], M. E. Hodson[2] and A. J. Morgan[1]

[1] Cardiff School of Biosciences, Cardiff University, Museum Avenue, Cardiff, CF10 3TL, UK.

[2] Department of Soil Science, School of Human and Environmental Sciences, University of Reading, Whiteknights, Reading, RG6 6DW, UK.

[3] King's College London, School of Biomedical & Health Sciences, Pharmaceutical Sciences Division, London, SE1 9NH, UK.

Corresponding Author: Dr Jane Andre, Cardiff School of Biosciences, Cardiff University, Museum Avenue, Cardiff, CF10 3TL, UK. Email:Andrej@cardiff.ac.uk

Abstract

Landscapes blighted by industrial activity but which accommodate seemingly resistant animal ecotypes provide an ideal opportunity to study evolution in action. We investigated the phylogenetic structure and molecular adaptation of the metal extremophile earthworm *Lumbricus rubellus* inhabiting a geochemically heterogeneous Pb/Zn mine. These worms prodigiously accumulate Pb (c.1.5% total-body dry mass),

trafficking most to Ca/PO4-rich organelles in chloragocytes. Infrared microspectroscopy (FTIR) of chloragocytes detected altered phosphate profiles in a putative tolerant population. Moreover, bioinformatic analysis of *L. rubellus* EST^{Pb} libraries indicated that constituents of Ca-signalling and sequestration pathways were aberrantly elevated. Sequencing a gene central to this pathway, sarco/endoplasmic calcium ATPase (SERCA), revealed mutations clustered in the cytosolic domain that correlated with site-specific Pb-tolerant genotypes. Our findings present a mechanism that enables locally adapted earthworm populations to tolerate a novel habitat, potentially contributing to genetic differentiation and eventual speciation. Additionally, they indicate that extreme Pb tolerance mechanisms are evolutionary appropriations of intrinsic Ca molecular machinery: Inorganic mimicry begets biomolecular adaptive mechanisms.

Capsule: Landscapes punctuated by polluted islands are inhabited by Pb-adapted invertebrate extremophiles.

Keywords: Calcium, lead, earthworms, phylogenetics, ecotoxicology

Introduction

Occupational and environmental exposures to Pb continue to cause serious human health problems in developing and industrialised countries (Tong et al., 2000) and diverse disruptive effects on both ecological processes (Rantalainen et al., 2006) and wildlife (Chapman and Wang, 2000). Intriguingly some field populations (ecotypes) of invertebrates have evolved resistance to metal stressors in their native habitats (Morgan et al., 2007), with the earthworm *Lumbricus rubellus* representing an extreme example of a complex organism that can manifest Pb tolerance in mine-associated soils contaminated to a degree exceeding by an order of magnitude the exposure level that severely compromises reproduction in spiked laboratory soils (Spurgeon et al., 1994). *L. rubellus* is an acid-tolerant species inhabiting litter on soils ranging from pH 3.8 to 8.4 (Sims and Gerard, 1985), thereby indicating that this early pioneer of mine-associated and industrial 'soils' may possess the genetic amplitude to become a locally-adapted population with the inherent ability to cope with site-specific extremes in contaminant availabilities. Although many terrestrial invertebrates reside in soils containing metal levels that far exceed effect concentrations for key life-cycle traits (Morgan and Morgan, 1990; Spurgeon and Hopkin, 1996) and despite being well documented in plants, evidence for the heritability of adaptive traits and evolution of metal-resistant ecotypes remain relatively sparse.

As sites elevated in inorganic pollutants illustrate both spatial and temporal stochasticity, with geological features alongside ancient and recent industrial activities providing an array of unique environments to which nature has adapted, cogent data sets to investigate evolution in action are presented. The Cwmystwyth valley, Wales (UK) is

a disused Pb (galena, PbS) mine that is highly heterogeneous in nature comprising of numerous micro-habitats that relate to soil metal content and pH. In combination with the acquisition of Pb-adaptive traits in *L. rubellus* ecotypes, a unique model to study earthworm phylogeography and, more specifically, whether metal contamination-associated stress and microsite heterogeneity have the potential to alter the genetic structure of populations through adaptation and speciation is provided. The ancestors of earthworms now resident at Cwmystwyth would have survived the major glaciations and climatic instability of the Devensian period in one or more of the sheltered southern refugia. With the onset of each stadial period and reformation of ice-sheets, retreating bottlenecked populations would have experienced shrinkage, dissection and extinction, whereas upon post-glacial expansion undergone adaptation and selection to new environments (Hewitt, 2000). Repeated climatic oscillations and changes in habitat range have therefore yielded increased species diversity through several genome reorganisations, manifested today by the broad environmental conditions and geographical ranges tolerated by *L. rubellus*. This includes the ability to colonise heterogeneous and potentially stressful metal-polluted habitats such as are found at Cwmystwyth.

Detrimental effects of Pb exposure arise from the ability of Pb to mimic other metals, primarily Ca (Clarkson, 1993; Warren et al., 1998). Intracellular interactions between Pb and Ca are well documented, with non-sequestered Pb metal ions shown to interact and associate with proteins active in the calcium signalling pathway. This shared chemical affinity between Pb and Ca lead us to hypothesise that the network of mechanisms evolved to regulate the potentially lethal levels of intracellular free Ca^{2+} are

somehow implicated in the handling of its non-essential cationic analogue. Ordinarily the main molecular resistance mechanisms underlying metal tolerance entails either metal efflux pumps (Callaghan and Denny, 2002) or sequestration by one of three classes of thiol-rich peptides, namely glutathione, phytochelatin and metallothionein (Vatamaniuk et al., 2005; Janssens et al., 2007). However, neither of these generic resistance mechanisms has been found to underpin Pb adaptations in earthworms or any other organism. Instead, accumulated Pb is sequestered within calcium phosphate-rich chloragosomes, unique organelles with certain lysosome-like properties that are located in the chloragogenous tissue (Morgan and Morgan, 1989). It follows that specific transport mechanisms must reside in vacuolar membranes for the uptake of metals and accompanying anions, which provide the complexing negatively charged ion needed for an insoluble precipitate to be formed. With this in mind, this study aimed to provide insights into the functional mechanisms of Pb management and adaptation in chronically exposed earthworm populations. This was achieved through global transcriptomic analyses, targeted single loci experiments and in-situ biochemical fingerprinting of the main metal sequestering organ of the earthworm. Combined with characterising the entire site in terms of metal content and pH and, in parallel, measuring population divergence from whole-genome and mtDNA loci, a sophisticated means of studying earthworm speciation and evolution to ecological heterogeneity on a micro-geographic scale is presented.

Materials and Methods

Portable XRF and pH mapping of the Cwmystwyth site. A portable XRF (NITON XLiand, Thermo Scientific Inc) and GPS system (Garmin, Etrex Venture) were used in order to create a Pb profile the Cwmystwyth valley, with a total of 97 measurements taken. A series of 71 soil samples were also collected and the pH of each recorded. The mapping software SURFER© was used to convert both the metal and pH data sets into a series of 3D rendered surface maps, stacked alongside a base-map of the valley.

Mitochondrial and AFLP genotyping. *Lumbricus rubellus* earthworms were collected by digging and hand-sorting. The animals were transported back to the laboratory in their native soil and depurated as described in (Arnold and Hodson, 2007). Genomic DNA was extracted from $C1^{pH5}_{Pb*}$ (n=27), $C2^{pH4}_{Pb*}$ (n=33), $C3^{pH7}_{Pb***}$ (n=32) and $C4^{pH6}_{Pb*}$ (n=30) earthworms using DNAzol reagent (Invitrogen Ltd., Paisley, UK.). DNA was also isolated from *L. castaneus* and *L. eiseni*. Forward (5'-TAGCTCACTTAGATGCCA) and reverse (5'-GTATGCGGATTTCTAATTGT) *L. rubellus* specific cytochrome oxidase II (COII) primers were designed from mitochondrial sequences deposited in LumbriBASE (www.earthworms.org). For each PCR reaction ~100ng DNA template was amplified using 10pmole/µl forward and reverse primer, 10mM dNTP mix and 5U/µl Taq DNA polymerase buffered with 5X Mg-free Taq PCR amplification buffer and supplemented with $MgCl_2$ (1.5mM). The reaction was denatured at 95°C for 10 minutes and then cycled 35 times at 95°C for 30 seconds, 30 seconds at the required primer annealing temperature and 72° for 1 minute. This was followed by a 10 minute final extension at 70°C. The amplicon (469bp) was resolved by electrophoresis in 1X TAE buffer at 120V for approximately 30 minutes in a Pharmacia GNA-100 tank. Nucleic acid bands were then visualised on a UV gel documentation system. Prior to

sequencing PCR clean-ups were performed using Exo-SAP-IT (Amersham Pharmacia, UK) reagents. Exonuclease 1 (0.25µl) and Shrimp Alkaline Phosphatase (0.5µl) were mixed with the PCR product (10µl) and incubated at 37°C for 45 minutes followed by 80°C for 15 minutes. DNA was sequenced using ABI PRISM® BigDye v3.1 Terminator Sequencing technology (Applied Biosystems, Foster City, USA) on the ABI PRISM® 3100 DNA Sequencer run by the Cardiff University Molecular Biology Support Unit. Raw sequence traces were confirmed using Finch TV before being imported into Mega v3.1 (Kumar et al., 2004) for alignment and tree construction. The distance-based neighbour joining (NJ) algorithm (Saitou and Nei, 1987), using p-distances, was used to estimate tree topology and calculate branch lengths. Relationships between phylogenetic haplotypes were determined by maximum parsimony (MP), maximum likelihood (ML) and Bayesian methods using PAUP v3.1 and MRBAYES respectively (Huelsenbeck and Crandall, 1997; Huelsenbeck and Ronquist, 2001), MRMODELTEST v2.2 (Nylander, 2004) and the Akaike Information Criterion (AIC) were used to select the optimum model (HKY+G) of sequence evolution that best fitted the data (base frequencies of A=0.3638, C=0.2244, G=0.1277, T=0.2842 and T-ratio=3.5050 and Gamma distribution parameter= 0.2005). Node support for MP and ML analyses was determined using a non-parametric bootstrap, with 500 and 1000 replicates respectively (Holmes, 2003). For the analysis 3x106 generations were run, with one tree retained every 300th generation and the first 2500 trees discarded as burn-in. Genetic distances were calculated using p-distance in Mega and median-joining networks were drawn using NETWORK and dnasp4.

AFLP analysis was adapted from (Ajmone-Marsan et al., 1997) with approximately 200ng of genomic DNA extracted from $C1^{pH5}_{Pb*}$ (n=24), $C2^{pH4}_{Pb*}$ (n=30), $C3^{pH7}_{Pb***}$ (n=23) and $C4^{pH6}_{Pb*}$ (n=18) individuals. Pre-selective EcoR1 (GACTGCGTACCAATTCA) and Taq1 (GACTGCGTACCAATTCC) primers were used and for the selective amplifications two primer combinations (E32 (5'-GACTGCGTACCAATTCAAC-3') /T32 (5'-GATGAGTCCTGACCGAAAC-3') and E32/T38 5'-GATGAGTCCTGAC CGAACT-3') were employed. Reactions were run by the Cardiff University Molecular Biology Support Unit and analysed on an Applied Biosystems 3130x1 fragment analyser. Bands between 70 and 325 base pairs (bp) and with a minimum peak height of 70 units were scored using GeneMapper analysis software. Microsoft Access, Excel and the Excel macro GenAlEx6 were used to create a cumulative table of all loci from each individual and transform the data into a binary form. Principal coordinates (PCO) analysis and phylogenetic tree construction, supported by bootstrap analysis (1000 replicates), was performed using the neighbour joining algorithm (based upon the Nei's distance) to in PAUP v4.0b10 to estimate tree topology and calculate branch length. The software Structure v2.2 was used to delineate clusters of individuals on a multi-locus, genotype basis using a Bayesian algorithm. The number of inferred populations ran from 1 to 5, with 8 replicate runs, a burn-in of 75 000 cycles followed by 300 000 for data collection. L(K), the modal choice criterion, is calculated in Structure and the true number of populations (K) can be deferred from its maximal value. ΔK, the rate of change in the log probability of data between successive K-values, provides a visual means to easily identify the number of clusters in a sample of individuals (Evanno et al., 2005).

192 EST libraries and informatics. Upon sampling $C3^{pH7}_{Pb***}$ (n=5) earthworms were
193 immediately immersed and maintained in liquid nitrogen and stored at -80°C until
194 required. R1 (n=5) earthworms were transported back to the laboratory in their native
195 soil and individuals exposed to 500, 750, 1250, 1750 and 2250 mg/kg^{-1} Pb in the form
196 of $Pb(NO_3)_2$ spiked Kettering loam soil and maintained at a WHC of 75% for three
197 weeks at 15°C. Spiked soil was left to equilibrate for three days prior to earthworm
198 addition. Following exposure, earthworms were snap-frozen in liquid nitrogen and
199 stored at -80°C. Earthworm total RNA (~1.25mg) was extracted using the TRI-reagent
200 method (Sigma-Aldrich, UK) and mRNA isolated using an mRNA Purification Kit
201 (Amersham, UK). cDNA libraries were constructed using the pBluescript® II XR
202 cDNA Library Construction Kit (Stratagene Europe, Amsterdam, The Netherlands).
203 PCR was used to screen the libraries and quantify insert size. Each PCR contained 5µl
204 neat culture, 10X Triton free PCR Buffer (10µl), MgCl2 (0.25µl, 1mM), universal
205 M13F and M13R primers (0.2µl, 10mM), dATP, dCTP, dTTP, dGTP (0.2µl, 100mM)
206 and Taq polymerase (0.16µl, 50U/ml) in a 95µl reaction. The reaction was denatured at
207 95°C for 10 minutes and then cycled 35 times at 95°C for 30 seconds, 30 seconds at the
208 primer annealing temperature of 56°C and 72° for 1 minute. This was followed by a 10
209 minute final extension at 70°C. Products were resolved by electrophoresis using E-Gel®
210 technology (Invitrogen Ltd., Paisley, UK) and associated editing software. High quality
211 clones were cherry-picked using the MultiPROBE® II HT EX liquid handling system
212 (PerkinElmer, Bucks., UK) and associated WinPrep® software. The composite plate
213 products were purified using Montage® Multiscreen PCRµ96 cleanup plates by vacuum
214 filtration and the DNA was resuspended in sterile water (30µl). Sequencing reactions
215 were completed by the SBSSS facility at Edinburgh University and sequences named

according to the NERC Environmental Genomics scheme to allow for bioinformatics analysis. The raw trace chromatograms from the sequencing reaction were processed using trace2dbEST (Sturzenbaum et al., 2003) which contains a base calling component (phred) and a sequence trimming component (cross_match). This software produces good quality EST sequences, formatted for submitting to NCBI dbEST (http://www.ncbi.nlm.nih.gov/dbEST). The EST sequences were clustered using CLOBB (Sturzenbaum, et al., 2003) to derive a consensus putative gene sequence contig and then processed by the software package PartiGene (Sturzenbaum, et al., 2003) (http://www.nematodes.org/PartiGene). Cluster information can be retrieved by LumbriBASE (http://www.earthworms.org) through simple text queries, identification of sequence similarity and library specific searches. The biological process and molecular function of gene sequences were described by defining their Gene Ontology (GO) classification using blast2go (http://www.blast2GO.de).

Fourier-transform Infrared Spectroscopy. Soil and adult (fully clitellate) *L. rubellus* earthworms were collected from $C2^{pH4}_{Pb*}$ and $C3^{pH7}_{Pb***}$ and the posterior segments immediately excised and quench-frozen in liquid nitrogen. The frozen tissue was transported to the laboratory under liquid nitrogen and stored at -20°C until required. Tissues were mounted in CryoEmbed and sectioned longitudinally at a nominal thickness of 50µm in a Bright cryostat. Sections were mounted on Kevley slides and air-dried overnight in the cold chamber of the cryostat. Infra-red spectra were collected in transmission mode from station 11.1 at the CLRC Daresbury Synchrotron Radiation Source. The chloragogenous tissue was visually identified and each section imaged and

analysed, with five spectra from five different regions of the tissue (i.e. x25 spectra per individual earthworm) collected.

SERCA. Plasmid preparations of individual LumbriBASE clones (Genbank accession numbers CF416761 and CO048347) were prepared using a Wizard® Plus SV Miniprep kit (Promega Ltd., Southampton, UK). Preparations were sequenced in their entirety by "walking" along the gene, after each step re-designing a specific reverse primer to complement the universal M13 forward. Primers were designed using the software Primer3 (Rozen and Skaletsky, 2000) and Oligo® (MBI Inc, USA) and sequencing performed as described above. These full-length library sequences were used to design *L. rubellus* specific SERCA primers in order to amplify the gene transcribed in individuals of each genealogical lineage. Reactions were denatured at 95°C for 10 minutes and then cycled 35 times at 95°C for 30 seconds, 30 seconds at the required primer annealing temperature and 72° for 1 minute. This was followed by a 10 minute final extension at 72°C. DNA was sequenced as described above by the Cardiff University Molecular Biology Support Unit.

Total RNA was extracted from tail-clips of an adult individual sampled from $C3^{pH7}_{Pb***}$ and $C2^{pH4}_{Pb*}$ using the Tri-reagent method (Sigma-Aldrich, UK). Complementary DNA (cDNA) was synthesised from messenger RNA (mRNA) using reverse transcriptase. Total RNA (7-20µg) was heated at 65°C for 3 minutes and combined with anchored oligo d(T) (1µl, 100mM) and random hexamers (2µl, 100mM) and incubated at 70°C for ten minutes. A reaction mix of 5X RT buffer (6µl), DTT (3µl, 0.1M) and dNTP mix (1.2µl, 10mM) was prepared and added to the RNA mix and incubated at 25°C for two

minutes. Superscript (1µl) was added and the reaction incubated at 42°C for 3 hours. A series of three PCR reactions were optimised and performed in order to obtain the full-length SERCA sequence of each individual; between sequenced sections of the gene there was a large overlap to ensure the same SERCA isoform was being amplified in each instance. PCRs were performed as described above. Reactions that yielded products >2000bp were modified. For these reactions DNA (~100ng) template was amplified using 10µM forward and reverse primer, 25mM dNTP mix and 1µl Herculase® II Fusion DNA Polymerase buffered with 5X Herculase® II PCR reaction buffer (Stratagene Europe, Amsterdam, The Netherlands). Each reaction was supplemented with an optimised quantity of $MgCl_2$ (25mM). The reaction was denatured at 95°C for 10 minutes and then cycled 35 times at 95°C for 20 seconds, 20 seconds at the required primer annealing temperature and 68° for 4 minutes. This was followed by a 4 minute final extension at 68°C. Protein sequences were aligned using bioinformatic software tool Mega v3.1 (Kumar, et al., 2004) and modelled using SWISS-MODEL (http://swissmodel.expasy.org//SWISS MODEL.html), Swiss-PdbViewer and Pymol.

Total RNA was extracted from nine, previously genotyped (mtDNA) adult individuals sampled from $C1^{pH5}_{Pb*}$, $C2^{pH4}_{Pb*}$, $C3^{pH7}_{Pb***}$ and $C4^{pH6}_{Pb*}$. This was followed by cDNA synthesis, as described above. A PCR was designed to enable easy identification by gel electrophoresis of the expressed isoform, with a combination of three primers used in each reaction (F1 5'-CTGGCCGGAATTCGTGTTATC-3', F2 5'-ATACTCTTCG CTGTCTTGCGT-3', R1 5'-CCGCTGGCTCTTCTTCCG-3'). The two forward primers were designed so that each one isolated one of the two isoforms. The resulting products

were of different sizes to enable simple identification on an agarose gel following resolution by electrophoresis.

Results and Discussion

The 'field laboratory': the metalliferous site and its resident earthworms. Mine sites are notoriously heterogeneous in geochemical nature. Mineral (galena, PbS) extraction from the Cwmystwyth Mine stopped in 1921. Today, the contaminated surface spoils are situated on base-poor underlying geology, punctuated by calcareous 'islands' around derelict buildings (Figure 1). In the decades since its abandonment, the spatially chequered and hostile site has developed micro-habitats colonised by a limited variety of naturally occurring plants and invertebrates. Thus, it serves as an ideal evolutionary field laboratory. *L. rubellus* was sampled from four sites across the mine: $C1^{pH5}_{Pb*}$, $C2^{pH4}_{Pb*}$, $C3^{pH7}_{Pb***}$, and $C4^{pH6}_{Pb*}$ (the number of asterisks denotes the level of contamination as classified by the Kelly index (ICRCL 59/83): * contaminated (1000-2000ppm), ** heavily contaminated (2000-10 000ppm) and *** unusually heavily contaminated (>10 000ppm) (Figure 1). Population divergence was measured using amplified fragment length polymorphism (AFLP) analysis and mitochondrial cytochrome oxidase II (mtDNA COII) gene sequence data of individuals split between the four sites. Two distinct lineages, differentiated at both the mitochondrial and nuclear level, were revealed with a mean inter-lineage mtDNA sequence divergence of approximately 13%, indicative of a cryptic species complex (Figures 2A and B). Moreover, from the distinct clustering of $C3^{pH7}_{Pb***}$ individuals a true genetic archipelago is inferred that can be related to the calcareous, circum-neutral (pH6.5) and heavily

311 polluted nature of this island in an otherwise acidic, moderately polluted environment.
312 Such cryptic complexes are typical in taxa that thrive in specialised environments and it
313 perhaps explains why islands of anthropogenic contamination result in the loss or
314 "erosion" of genetic diversity. Potential adaptive and sympatric speciation processes
315 may be occurring, with the concurrent, contaminant-driven acquisition of adaptive gene
316 complexes in response to the unique nature of the site.
317
318 As phylogenetic population structure is shaped by ongoing processes of genetic drift
319 and gene flow, combined with past historical events, unravelling the *L. rubellus* species
320 complex requires inferences on both the structure of the phylogeny and demographic
321 tendencies. Accurately inferring the population-level dynamics of evolutionary
322 mechanisms that involve adaptive, and also sympatric, speciation is complex, especially
323 as the process is not clearly defined and involves both temporal and spatial
324 stochasticity. The timeline of divergence leading to sustained differentiation is neither
325 rapid nor definable and, due to the combined effects of gene flow and selection of
326 adaptively important genes, the genomes of incompletely isolated populations will
327 contain an assortment of variable and undifferentiated regions (Supporting data).
328 Fluctuations in the global climate have led to major ice ages during the Quaternary
329 period, with the Pleistocene epoch (1 808 000 to 11 500 before present (BP)) covering
330 the most recent period of repeated glaciations. Glaciation evidence can be related to the
331 profile of mtDNA haplotypes in both lineage A and B, the shape of the corresponding
332 mismatch distributions (Figures 2C and D), and estimated time since population
333 expansion. Lineage A comprises nine haplotypes that contain two or more individuals.
334 This, combined with a ragged multimodal mismatch distribution, is suggestive of a

335 stationary population that has undergone multiple introductions and bottleneck episodes
336 (Harpending, 1994). Additionally, from the parameters Tau and date of growth in
337 mutational units, expansion is estimated to have occurred approximately 250 000 years
338 BP (assuming one generation per year) and may have corresponded with a non-glacial
339 environment such as the Hoxnian interstadial (~250 000 BP) (Brown, 1979) (Figure
340 2E). In contrast, lineage B consists of three haplotypes that contain two or more
341 individuals, and displays a unimodal mismatch distribution, and a post-glacial
342 population expansion time of approximately 17 000 years BP was calculated. This
343 combined evidence suggests that the population experienced a single burst of growth
344 and expanded after the height of the last glaciation period (~25 000 BP) with adaptation
345 or selection occurring in response to the warmer climate experienced towards the end of
346 the Devensian glaciation and onset of the Windermere interstadial (Brown, 1979;
347 Harpending, 1994).
348
349 'In-situ' biochemical fingerprinting and EST libraries from Pb-mine and laboratory
350 exposed naïve worms. These field earthworm populations prodigiously accumulate up
351 to 1.5% of total body dry mass Pb (Morgan, 2001), with Ca/PO_4-rich earthworm
352 chloragocyte cells constituting the main metal sequestering organ (Cotter-Howells et al.,
353 2005). Fourier transform infrared (FT-IR) microspectroscopy on a high energy
354 synchrotron source was used to determine the chemical composition of cryo-sectioned
355 chloragocytes in earthworms belonging to each lineage at the two heavily polluted,
356 albeit one acidic ($C2^{pH4}_{Pb*}$) and one calcareous ($C3^{pH7}_{Pb***}$), mine sites. The chlorogogenous
357 tissue was found to have a distinctive FTIR spectrum (Supporting data) and site-specific
358 disparities in the composition of chlorogogenous tissue (in the $1100cm^{-3}$ region of the

spectrum) were apparent, which correlated with phosphorous-containing functional groups (Figure 3A and B). As the earthworm chloragocyte is thought to be involved in haem biosynthesis (Jamieson and Molyneux, 1981), a conserved pathway that is inhibited by Pb at several junctures (Warren, et al., 1998), Pb trafficking into and across earthworm chloragocytes must be tightly regulated in these animals that are continuously exposed to high concentrations of metal in their native environments and whose strategy for dealing with it involves intracellular accumulative immobilization. Indeed, inferences on the mechanisms of adaptive evolution to environmental heterogeneity require not only abstract genotype- to phenotype associations but more meaningful molecular genetic interpretations regarding the nature of induced phenotypic variation. The transcriptomic profile of an organism provides a snapshot of gene expression to provide information regarding developmental stage, life-history or responses in relation to particular environmental stressors. EST libraries are also the substrate for comparative genomic studies, through investigating differential expression between cDNA populations. Two libraries were constructed from earthworm populations with contrasting histories of Pb exposure; $C3^{pH7}_{Pb***}$ earthworms, chronically exposed in the field, and earthworms sampled from a clean reference site, R1, acutely exposed to lead in the form of $Pb(NO_3)_2$ under laboratory conditions. In combination with the plethora of EST cluster information already available in LumbriBASE (www.earthworms.org), originating from control (adult, head-enriched), life-stage (late cocoon, juvenile) and other exposure (Cu, Cd, fluoranthene, atrazine) libraries, a metal tolerant genotype may be related to phenotype and the functional systems that underlie lead handling within these earthworm populations defined. Both libraries comprised high quality sequences with an average length of between 500 and 600 base pairs. This

ensured the maximum numbers of sequences were annotated to enable accurate downstream analysis and interpretation using the software LumbriBASE, Blast2GO and associated KEGG resource, which generates pathway maps that highlight gene ontology relationships between annotated sequences. Of interest was the significant number of $C3^{pH7}_{Pb***}$ gene products (when compared to other libraries) associated with intracellular Ca^{2+} sensing (Calmodulin, Troponin C) and buffering (Sarcoplasmic calcium binding protein (SCP), Parvalbumin) (Gao et al., 2006; Ishida and Vogel, 2006). Both sets of proteins belong to the EF-hand super-family of proteins implicated in calcium binding and central to the Ca-signalling pathway. These observations indicate that components of the Ca signalling pathway are central to Pb sequestration within chloragocytes which, in turn, may be associated with adjustments in the metabolism of their common complexing PO_4^- anion. This yields a number of candidate loci that may contribute to a Pb-tolerance phenotype by modifying molecules involved in the cellular physiology of an essential cation (Ca^{2+}) to accommodate its non-essential cationic mimic (Pb^{2+}).

Sarcoplasmic/ endoplasmic reticulum calcium ATPase (SERCA).

Sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA) is a central transport carrier protein of the Ca signalling pathway that resides in membranes of intracellular storage sites for the uptake of excess Ca and, conceivably, Pb (Tsien et al., 1987). Three isoforms have been described in vertebrates (MacLennan et al., 1985; Campbell et al., 1991; Vilsen and Andersen, 1992) and one in invertebrates (Palmero and Sastre, 1989; Escalante and Sastre, 1994; Shi et al., 1998a; b), fungi (Ghislain et al., 1990) and plants (Wimmers et al., 1992). All isoforms are similar in structure and have a 75-85% identical amino acid sequence (Toyofuku et al., 1992). Despite the identification of

several vertebrate isoforms, *L. rubellus* is the first invertebrate found to harbour multiple SERCA proteins. Two structurally different forms were identified in populations, sampled from different field sites, and expression was found to be co-incident with the mitochondrial lineage marker (COII), even where nuclear hybridisation was observed (Figure 4A and B). Their structure differed in amino acids located in the cytosolic nucleotide-binding domain (the flap) of the protein, a region thought to have a critical role in determining calcium affinity and turnover (Figure 4C). This observation indicates that not only are the intracellular trans-membrane Ca and Pb pathways confluent at the molecular (SERCA) level and are associated with adjustments in the metabolism of their common complexing PO_4^- anion, but the entire machinery is demonstrably genotype-specific (Figure 4C). It is important to point out, however, that whist the SERCA molecule is an important connector between the vulnerable cytosol and the intra-vesicular depository of immobilized Pb, other components of the Ca pathway warrant study to determine whether they are structurally or functionally modified.

Conclusion

Findings presented here demonstrate that tolerance in earthworms with a protracted history of Pb exposure is not merely an expression of genetic divergence, rather that field populations may evolve Pb-adaptation traits by modifying molecular regulators of Ca physiology. Whilst adaptive changes in enzyme structure are, for good reason, less probable than changes in the promoters that regulate enzyme expression (Crawford et al., 1999b; a), they are clearly not molecular traits that can be ignored, as the similarity

in the ionic radii of Ca^{2+} (1.00Å) and Pb^{2+} (1,19Å) (Bridges and Zalups, 2005) may, for example, facilitate structural mutations. Together, these observations describe a molecular mechanism for Pb resistance in earthworms that may underpin their phylogenetic differentiation within fairly discrete microhabitats across the study site. Furthermore, the concept serves as a paradigm for invertebrates generally because the ability for incorporating Pb into 'calcospherites' is ubiquitous, demonstrating that evolution innovates by modifying existing structures or pathways.

Acknowledgments

This work was performed during the tenure of a Natural Environmental Research Council studentship (JA) registered at Reading University (NER/S/A/2004/12418). The FTIR was performed at the Daresbury Synchrotron, Station 11.1, managed by Dr F. Bahrami (CCLRC beamtime award 45211). We would like to thank Dr Peter Brabham (School of Earth, Ocean and Planetary Sciences, Cardiff University) for his geomapping support.

References

Ajmone-Marsan, P., Valentini, A., Cassandro, M., Vecchiotti-Antaldi, G., Bertoni, G., Kuiper, M., 1997. AFLP markers for DNA fingerprinting in cattle. Animal Genetics 28, 418-426.

Arnold, R.E., Hodson, M.E., 2007. Effect of time and mode of depuration on tissue copper concentrations of the earthworms *Eisenia andrei*, *Lumbricus rubellus* and *Lumbricus terrestris*. Environmental Pollution 148, 21-30.

454 Bridges, C.C., Zalups, R.K., 2005. Molecular and ionic mimicry and the transport of
455 toxic metals. Toxicology and Applied Pharmacology 204, 274-308.
456 Brown, E.H., 1979. The Shape of Britain. Transactions of the Institute of British
457 Geographers, 449–462.
458 Callaghan, A., Denny, N., 2002. Evidence for an Interaction between p-Glycoprotein
459 and Cadmium Toxicity in Cadmium-Resistant and-Susceptible Strains of Drosophila
460 melanogaster. Ecotoxicology and Environmental Safety 52, 211-213.
461 Campbell, A.M., Kessler, P.D., Sagara, Y., Inesi, G., Fambrough, D.M., 1991.
462 Nucleotide sequences of avian cardiac and brain SR/ER $Ca^{(2+)}$-ATPases and functional
463 comparisons with fast twitch $Ca^{(2+)}$-ATPase. Calcium affinities and inhibitor effects.
464 Journal of Biological Chemistry 266, 16050-16055.
465 Chapman, P.M., Wang, F., 2000. Issues in Ecological Risk Assessment of Inorganic
466 Metals and Metalloids. Human and Ecological Risk Assessment 6, 965-988.
467 Clarkson, T.W., 1993. Molecular and Ionic Mimicry of Toxic Metals. Annual Review
468 of Pharmacology and Toxicology 33, 545-571.
469 Cotter-Howells, J., Charnock, J.M., Winters, C., Kille, P., Fry, J.C., Morgan, A.J., 2005.
470 Metal compartmentation and speciation in a soil sentinel: The earthworm, *Dendrodrilus*
471 *rubidus*. Environmental Science & Technology 39, 7731-7740.
472 Crawford, D.L., Segal, J.A., Barnett, J.L., 1999a. Evolutionary analysis of TATA-less
473 proximal promoter function. Mol Biol Evol 16, 194-207.
474 Crawford, D.L., Segal, J.A., Barnett, J.L., 1999b. Evolutionary analysis of TATA-less
475 proximal promoter function. Molecular Biology and Evolution 16, 194-207.
476 Escalante, R., Sastre, L., 1994. Structure of Artemia franciscana sarco/endoplasmic
477 reticulum Ca-ATPase gene. Journal of Biological Chemistry 269, 13005-13012.

478 Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of
479 individuals using the software STRUCTURE: a simulation study. Molecular Ecology
480 14, 2611-2620.
481 Gao, Y., Gillen, C.M., Wheatly, M.G., 2006. Molecular characterization of the
482 sarcoplasmic calcium-binding protein (SCP) from crayfish Procambarus clarkii.
483 Comparative Biochemistry and Physiology Part B Biochemistry & Molecular Biology
484 144, 478-487.
485 Ghislain, M., Goffeau, A., Halachmi, D., Eilam, Y., 1990. Calcium homeostasis and
486 transport are affected by disruption of cta3, a novel gene encoding $Ca^{2(+)}$-ATPase in
487 Schizosaccharomyces pombe. Journal of Biological Chemistry 265, 18400-18407.
488 Harpending, H.C., 1994. Signature of Ancient Population-Growth in a Low-Resolution
489 Mitochondrial-DNA Mismatch Distribution. Human Biology 66, 591-600.
490 Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907-913.
491 Holmes, S., 2003. Bootstrapping phylogenetic trees: Theory and methods. Statistical
492 Science 18, 241-255.
493 Huelsenbeck, J.P., Crandall, K.A., 1997. Phylogeny estimation and hypothesis testing
494 using maximum likelihood. Annual Review of Ecology and Systematics 28, 437-466.
495 Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic
496 trees. Bioinformatics 17, 754-755.
497 Ishida, H., Vogel, H.J., 2006. Protein-peptide interaction studies demonstrate the
498 versatility of calmodulin target protein binding. Protein and Peptide Letters 13, 455-
499 465.
500 Jamieson, B.G.M., Molyneux, B.G., 1981. The Ultrastructure of the Oligochaeta.
501 Academic Press.

Janssens, T.K.S., Marien, J., Cenijn, P., Legler, J., van Straalen, N.M., Roelofs, D., 2007. Recombinational micro-evolution of functionally different metallothionein promoter alleles from Orchesella cincta. BMC Evolutionary Biology 7, 88.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5, 150-163.

MacLennan, D.H., Brandl, C.J., Korczak, B., Green, N.M., 1985. Amino-acid sequence of a Ca^{2+} Mg^{2+}-dependent ATPase from rabbit muscle sarcoplasmic reticulum, deduced from its complementary DNA sequence. Nature 316, 696-700.

Morgan, A.J., Kille, P., Sturzenbaum, S.R., 2007. Microevolution and ecotoxicology of metals in invertebrates. Environmental Science & Technology 41, 1085-1096.

Morgan, J.E., Morgan, A.J., 1989. The Effect of Lead Incorporation on the Elemental Composition of Earthworm (Annelida, Oligochaeta) Chloragosome Granules. Histochemistry 92, 237-241.

Morgan, J.E., Morgan, A.J., 1990. The distribution of cadmium, copper, lead, zinc and calcium in the tissues of the earthworm *Lumbricus rubellus* sampled from one uncontaminated and four polluted soils. Oecologia 84, 559-566.

Morgan, J.E., Richards, S. P., Morgan, A. J., 2001. Stable strontium accumulation by earthworms; a paradigm for radiostrntium interactions with its cationic analogue, calcium. Environmental Toxicology and Chemistry 20, 1236-1243.

Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

524 Palmero, I., Sastre, L., 1989. Complementary DNA cloning of a protein highly
525 homologous to mammalian sarcoplasmic reticulum Ca-ATPase from the crustacean
526 Artemia. Journal of Molecular Biology 210, 737-748.
527 Rantalainen, M.L., Torkkeli, M., Strommer, R., Setala, H., 2006. Lead contamination of
528 an old shooting range affecting the local ecosystem--A case study with a holistic
529 approach. Science of the Total Environment 369, 99-108.
530 Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for
531 biologist programmers. Methods in Molecular Biology 132, 365-386.
532 Saitou, N., Nei, M., 1987. The Neighbor-Joining Method - a New Method for
533 Reconstructing Phylogenetic Trees. Molecular Biology and Evolution 4, 406-425.
534 Shi, X., Chen, M., Pe, H., Hardwicke, P.M.D., 1998a. Amino acid sequence of a Ca 2-
535 transporting ATPase from the sarcoplasmic reticulum of the cross-striated part of the
536 adductor muscle of the deep sea scallop: comparison to serca enzymes of other animals.
537 Comparative Biochemistry and Physiology Part B Biochemistry & Molecular Biology
538 120, 359-374.
539 Shi, X., Chen, M., Pe, H., Hardwicke, P.M.D., 1998b. Amino acid sequence of a Ca 2-
540 transporting ATPase from the sarcoplasmic reticulum of the cross-striated part of the
541 adductor muscle of the deep sea scallop: comparison to serca enzymes of other animals.
542 Comparative Biochemistry and Physiology Part B 120, 359-374.
543 Sims, R.W., Gerard, B.M., 1985. Earthworms: keys and notes for the identification and
544 study of the species. Brill Archive.
545 Spurgeon, D.J., Hopkin, S.P., 1996. Effects of metal-contaminated soils on the growth
546 and, sexual development, and early cocoon production of the earthworm *Eisenia fetida*,
547 with particular reference to zinc. Ecotoxicology and Environmental Safety 35, 86-95.

548 Spurgeon, D.J., Hopkin, S.P., Jones, D.T., 1994. Effects of cadmium, copper, lead and
549 zinc on growth, reproduction and survival of the earthworm Eisenia fetida (Savigny):
550 Assessing the environmental impact of point-source metal contamination in terrestrial
551 ecosystems. Environmental Pollution 84, 123-130.
552 Sturzenbaum, S.R., Parkinson, J., Blaxter, M., Morgan, A.J., Kille, P., Georgiev, O.,
553 2003. The earthworm Expressed Sequence Tag project. Pedobiologia 47, 447-451.
554 Tong, S., von Schirnding, Y.E., Prapamontol, T., 2000. Environmental lead exposure: a
555 public health problem of global dimensions. Bulletin of the World Health Organization
556 78, 1068-1077.
557 Toyofuku, T., Kurzydlowski, K., Lytton, J., MacLennan, D.H., 1992. The nucleotide
558 binding/hinge domain plays a crucial role in determining isoform-specific Ca^{2+}
559 dependence of organellar $Ca^{(2+)}$-ATPases. Journal of Biological Chemistry 267, 14490-
560 14496.
561 Tsien, R.W., Hess, P., McCleskey, E.W., Rosenberg, R.L., 1987. Calcium Channels -
562 Mechanisms of Selectivity, Permeation, and Block. Annual Review of Biophysics and
563 Biomolecular Structure 16, 265-290.
564 Vatamaniuk, O.K., Bucher, E.A., Sundaram, M.V., Rea, P.A., 2005. CeHMT-1, a
565 Putative Phytochelatin Transporter, Is Required for Cadmium Tolerance in
566 Caenorhabditis elegans. Journal of Biological Chemistry 280, 23684.
567 Vilsen, B., Andersen, J.P., 1992. Deduced amino acid sequence and E1-E2 equilibrium
568 of the sarcoplasmic reticulum $Ca^{(2+)}$-ATPase of frog skeletal muscle. Comparison with
569 the $Ca^{(2+)}$-ATPase of rabbit fast twitch muscle. FEBS Letters 306, 213-218.

570 Warren, M.J., Cooper, J.B., Wood, S.P., Shoolingin-Jordan, P.M., 1998. Lead
571 poisoning, haem synthesis and 5-aminolaevulinic acid dehydratase. Trends in
572 Biochemical Sciences 23, 217-221.
573 Wimmers, L.E., Ewing, N.N., Bennett, A.B., 1992. Higher plant $Ca^{(2+)}$-ATPase:
574 primary structure and regulation of mRNA abundance by salt. Proceedings of the
575 National Academy of Sciences of the United States of America 89, 9205-9209.
576
577
578 Figure legends
579
580 Figure 1. Earthworm population structure superimposed on a geochemical map of the
581 Cwmystwyth Pb mine. Surface maps depicting the pH (B) and Pb (C) levels are
582 overlayed on a topographical map of the Cwmystwyth valley (D), derived from 71 and
583 97 independent measurements respectively, and generated using SURFER©. The four
584 earthworm sampling sites are indicated by vertical black guide lines together with the
585 median joining networks depicting the phylogenetic structure, based upon COII
586 sequence data, of each population studied (A). The size of each haplotype group within
587 the network is proportional to the total number of individuals attributed to the genotype.
588 Mitochondrial lineage A individuals are shown as open circles whilst lineage B are
589 filled circles.
590
591 Figure 2. Mitochondrial and nuclear analysis of the earthworm, *L. rubellus*, population
592 structure and corresponding mitochondrial mismatch distributions. Earthworms
593 collected at four equally specially distributed sites with contrasting geo-chemical

properties were analysed for their mitochondrial (panel A) and nuclear (panel B) genotype. Sites included $C4^{pH6}_{Pb*}$ (blue triangles) and $C1^{pH5}_{Pb*}$ (purple diamonds), at the boundary of mine, together with $C3^{pH7}_{Pb***}$ (grey circles) and $C2^{pH4}_{Pb*}$ (red squares). A. Median-joining network analysis based upon 440 bp sequence of the COII mitochondrial gene of 119 *L. rubellus* individuals. The size of each haplotype group within the network is proportion to the total number of individuals attributed to the genotype whilst the earthworm source is indicated by fill colour. The left and right hand branches of the network are denoted lineage A & B respectively. B. AFLP multi-locus profiling principal component analysis showing individuals from the four sample stations. Those individuals exhibiting mitochondrial lineage B genotype are circled in red whilst those attributed to lineage A are circled in grey. The clustering of $C3^{pH7}_{Pb***}$ earthworms distinct from other lineage A earthworms is indicated with a dotted grey line. Hybrid individuals are shown by a lack of fill colour. C and D. Arlequin simulated mitochondrial mismatch distributions, using the model of demographic expansion, of lineage A (C) and lineage B (D) haplotypes respectively. The solid lines are the observed mismatch distribution and the dotted line shows the distribution simulated under the expansion model. E. Associated sum of squared deviation (SSD), Raggedness (Rg) and p-value statistics, based on 1000 data bootstraps.

Figure 3. Metabolomic fingerprinting of earthworm chloragogenous tissue using Fourier Transform Infrared spectroscopy. A. The fingerprint region of averaged infra-red spectra of earthworm chloragogenous tissue collected from $C3^{pH7}_{Pb***}$ (grey) and $C2^{pH4}_{Pb*}$ (red). Individual spectra were processed by the software package OPUS©). B. The main

617 difference in $C3^{pH7}_{Pb***}$ and $C2^{pH4}_{Pb*}$ averaged spectra (~1080cm^{-1}), corresponded to

618 phosphorus-containing functional groups, C. XLSTAT simulated dendrogram

619 illustrating the clustering of $C3^{pH7}_{Pb***}$ and $C2^{pH4}_{Pb*}$ earthworms according to their infra-red

620 spectral patterns (1096-1123cm^{-1})

621

622 Figure 4. Analysis of earthworm SERCA variants. A. phylogenetic analysis of

623 genotyped individuals, based upon the cytochrome oxidase II (COII) gene, from sites

624 including the control sites $C4^{pH6}_{Pb*}$ (blue triangles) and $C1^{pH5}_{Pb*}$ (purple diamonds) at the

625 boundary of the mine and contaminated sites $C3^{pH7}_{Pb***}$ (grey circles) and $C2^{pH4}_{Pb*}$ (red

626 squares) alongside *L. castaneus* and *L. eiseni* (white triangles). B. Discriminatory PCR

627 illustrating the lineage-specific expression of the SERCA variants and C. PyMol

628 simulated model of SERCA. The conserved calcium binding sites are indicated in

629 yellow and amino acid differences in the two *L. rubellus* isoforms in red. The

630 phosphorylation (P) and nucleotide binding (N) domains are shown

631