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

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

1 **Molecular genetic differentiation in earthworms inhabiting a heterogeneous Pb-polluted**

2 **landscape**

3

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

5 Morgan[1]

6

7 [1] Cardiff School of Biosciences, Cardiff University, Museum Avenue, Cardiff, CF10

8 3TL, UK.

9

10 [2] Department of Soil Science, School of Human and Environmental Sciences,

11 University of Reading, Whiteknights, Reading, RG6 6DW, UK.

12

13 [3] King's College London, School of Biomedical & Health Sciences, Pharmaceutical

14 Sciences Division, London, SE1 9NH, UK.

15

16 Corresponding Author: Dr Jane Andre, Cardiff School of Biosciences, Cardiff

17 University, Museum Avenue, Cardiff, CF10 3TL, UK. Email:Andrej@cardiff.ac.uk

18

19 Abstract

20

21 Landscapes blighted by industrial activity but which accommodate seemingly resistant

22 animal ecotypes provide an ideal opportunity to study evolution in action. We

23 investigated the phylogenetic structure and molecular adaptation of the metal

24 extremophile earthworm *Lumbricus rubellus* inhabiting a geochemically heterogeneous

25 Pb/Zn mine. These worms prodigiously accumulate Pb (c.1.5% total-body dry mass),

26 trafficking most to Ca/PO_4-rich organelles in chloragocytes. Infrared

27 microspectroscopy (FTIR) of chloragocytes detected altered phosphate profiles in a

28 putative tolerant population. Moreover, bioinformatic analysis of *L. rubellus* EST[Pb]

29 libraries indicated that constituents of Ca-signalling and sequestration pathways were

30 aberrantly elevated. Sequencing a gene central to this pathway, sarco/endoplasmic

31 calcium ATPase (SERCA), revealed mutations clustered in the cytosolic domain that

32 correlated with site-specific Pb-tolerant genotypes. Our findings present a mechanism

33 that enables locally adapted earthworm populations to tolerate a novel habitat,

34 potentially contributing to genetic differentiation and eventual speciation. Additionally,

35 they indicate that extreme Pb tolerance mechanisms are evolutionary appropriations of

36 intrinsic Ca molecular machinery: Inorganic mimicry begets biomolecular adaptive

37 mechanisms.

38

39 Capsule: Landscapes punctuated by polluted islands are inhabited by Pb-adapted

40 invertebrate extremophiles.

41

42 Keywords: Calcium, lead, earthworms, phylogenetics, ecotoxicology

43

44

45

46

47

48

49 Introduction

50

51 Occupational and environmental exposures to Pb continue to cause serious human

52 health problems in developing and industrialised countries (Tong et al., 2000) and

53 diverse disruptive effects on both ecological processes (Rantalainen et al., 2006) and

54 wildlife (Chapman and Wang, 2000). Intriguingly some field populations (ecotypes) of

55 invertebrates have evolved resistance to metal stressors in their native habitats (Morgan

56 et al., 2007), with the earthworm *Lumbricus rubellus* representing an extreme example

57 of a complex organism that can manifest Pb tolerance in mine-associated soils

58 contaminated to a degree exceeding by an order of magnitude the exposure level that

59 severely compromises reproduction in spiked laboratory soils (Spurgeon et al., 1994). *L.*

60 *rubellus* is an acid-tolerant species inhabiting litter on soils ranging from pH 3.8 to 8.4

61 (Sims and Gerard, 1985), thereby indicating that this early pioneer of mine-associated

62 and industrial 'soils' may possess the genetic amplitude to become a locally-adapted

63 population with the inherent ability to cope with site-specific extremes in contaminant

64 availabilities. Although many terrestrial invertebrates reside in soils containing metal

65 levels that far exceed effect concentrations for key life-cycle traits (Morgan and

66 Morgan, 1990; Spurgeon and Hopkin, 1996) and despite being well documented in

67 plants, evidence for the heritability of adaptive traits and evolution of metal-resistant

68 ecotypes remain relatively sparse.

69

70 As sites elevated in inorganic pollutants illustrate both spatial and temporal

71 stochasticity, with geological features alongside ancient and recent industrial activities

72 providing an array of unique environments to which nature has adapted, cogent data sets

73 to investigate evolution in action are presented. The Cwmystwyth valley, Wales (UK) is

74   a disused Pb (galena, PbS) mine that is highly heterogeneous in nature comprising of

75   numerous micro-habitats that relate to soil metal content and pH. In combination with

76   the acquisition of Pb-adaptive traits in *L. rubellus* ecotypes, a unique model to study

77   earthworm phylogeography and, more specifically, whether metal contamination-

78   associated stress and microsite heterogeneity have the potential to alter the genetic

79   structure of populations through adaptation and speciation is provided. The ancestors of

80   earthworms now resident at Cwmystwyth would have survived the major glaciations

81   and climatic instability of the Devensian period in one or more of the sheltered southern

82   refugia. With the onset of each stadial period and reformation of ice-sheets, retreating

83   bottlenecked populations would have experienced shrinkage, dissection and extinction,

84   whereas upon post-glacial expansion undergone adaptation and selection to new

85   environments (Hewitt, 2000). Repeated climatic oscillations and changes in habitat

86   range have therefore yielded increased species diversity through several genome

87   reorganisations, manifested today by the broad environmental conditions and

88   geographical ranges tolerated by *L. rubellus*. This includes the ability to colonise

89   heterogeneous and potentially stressful metal-polluted habitats such as are found at

90   Cwmystwyth.

91

92   Detrimental effects of Pb exposure arise from the ability of Pb to mimic other metals,

93   primarily Ca (Clarkson, 1993; Warren et al., 1998). Intracellular interactions between

94   Pb and Ca are well documented, with non-sequestered Pb metal ions shown to interact

95   and associate with proteins active in the calcium signalling pathway. This shared

96   chemical affinity between Pb and Ca lead us to hypothesise that the network of

97   mechanisms evolved to regulate the potentially lethal levels of intracellular free Ca^{2+} are

98   somehow implicated in the handling of its non-essential cationic analogue. Ordinarily

99   the main molecular resistance mechanisms underlying metal tolerance entails either

100   metal efflux pumps (Callaghan and Denny, 2002) or sequestration by one of three

101   classes of thiol-rich peptides, namely glutathione, phytochelatin and metallothionein

102   (Vatamaniuk et al., 2005; Janssens et al., 2007). However, neither of these generic

103   resistance mechanisms has been found to underpin Pb adaptations in earthworms or any

104   other organism. Instead, accumulated Pb is sequestered within calcium phosphate-rich

105   chloragosomes, unique organelles with certain lysosome-like properties that are located

106   in the chloragogenous tissue (Morgan and Morgan, 1989). It follows that specific

107   transport mechanisms must reside in vacuolar membranes for the uptake of metals and

108   accompanying anions, which provide the complexing negatively charged ion needed for

109   an insoluble precipitate to be formed. With this in mind, this study aimed to provide

110   insights into the functional mechanisms of Pb management and adaptation in

111   chronically exposed earthworm populations. This was achieved through global

112   transcriptomic analyses, targeted single loci experiments and in-situ biochemical

113   fingerprinting of the main metal sequestering organ of the earthworm. Combined with

114   characterising the entire site in terms of metal content and pH and, in parallel,

115   measuring population divergence from whole-genome and mtDNA loci, a sophisticated

116   means of studying earthworm speciation and evolution to ecological heterogeneity on a

117   micro-geographic scale is presented.

118

119   Materials and Methods

120

121   <u>Portable XRF and pH mapping of the Cwmystwyth site.</u> A portable XRF (NITON

122   XLiand, Thermo Scientific Inc) and GPS system (Garmin, Etrex Venture) were used in

123   order to create a Pb profile the Cwmystwyth valley, with a total of 97 measurements

124   taken. A series of 71 soil samples were also collected and the pH of each recorded. The

125   mapping software SURFER$^{©}$ was used to convert both the metal and pH data sets into a

126   series of 3D rendered surface maps, stacked alongside a base-map of the valley.

127

128   <u>Mitochondrial and AFLP genotyping</u>. *Lumbricus rubellus* earthworms were collected

129   by digging and hand-sorting. The animals were transported back to the laboratory in

130   their native soil and depurated as described in (Arnold and Hodson, 2007). Genomic

131   DNA was extracted from Cl_{Pb*}^{pH5} (n=27), $C2_{Pb*}^{pH4}$ (n=33), $C3_{Pb***}^{pH7}$ (n=32) and $C4_{Pb*}^{pH6}$ (n=30)

132   earthworms using DNAzol reagent (Invitrogen Ltd., Paisley, UK.). DNA was also

133   isolated from *L. castaneus* and *L. eiseni*. Forward (5'-TAGCTCACTTAGATGCCA)

134   and reverse (5'-GTATGCGGATTTCTAATTGT) *L. rubellus* specific cytochrome

135   oxidase II (COII) primers were designed from mitochondrial sequences deposited in

136   LumbriBASE (www.earthworms.org). For each PCR reaction ~100ng DNA template

137   was amplified using 10pmole/µl forward and reverse primer, 10mM dNTP mix and

138   5U/µl Taq DNA polymerase buffered with 5X Mg-free Taq PCR amplification buffer

139   and supplemented with $MgCl_2$ (1.5mM). The reaction was denatured at 95°C for 10

140   minutes and then cycled 35 times at 95°C for 30 seconds, 30 seconds at the required

141   primer annealing temperature and 72° for 1 minute. This was followed by a 10 minute

142   final extension at 70°C. The amplicon (469bp) was resolved by electrophoresis in 1X

143   TAE buffer at 120V for approximately 30 minutes in a Pharmacia GNA-100 tank.

144   Nucleic acid bands were then visualised on a UV gel documentation system. Prior to

145   sequencing PCR clean-ups were performed using Exo-SAP-IT (Amersham Pharmacia,

146   UK) reagents. Exonuclease 1 (0.25µl) and Shrimp Alkaline Phosphatase (0.5µl) were

147   mixed with the PCR product (10µl) and incubated at 37°C for 45 minutes followed by

148   80°C for 15 minutes. DNA was sequenced using ABI PRISM® BigDye v3.1 Terminator

149   Sequencing technology (Applied Biosystems, Foster City, USA) on the ABI PRISM®

150   3100 DNA Sequencer run by the Cardiff University Molecular Biology Support Unit.

151   Raw sequence traces were confirmed using Finch TV before being imported into Mega

152   v3.1 (Kumar et al., 2004) for alignment and tree construction. The distance-based

153   neighbour joining (NJ) algorithm (Saitou and Nei, 1987), using p-distances, was used to

154   estimate tree topology and calculate branch lengths. Relationships between phylogenetic

155   haplotypes were determined by maximum parsimony (MP), maximum likelihood (ML)

156   and Bayesian methods using PAUP v3.1 and MRBAYES respectively (Huelsenbeck

157   and Crandall, 1997; Huelsenbeck and Ronquist, 2001), MRMODELTEST v2.2

158   (Nylander, 2004) and the Akaike Information Criterion (AIC) were used to select the

159   optimum model (HKY+G) of sequence evolution that best fitted the data (base

160   frequencies of A=0.3638, C=0.2244, G=0.1277, T=0.2842 and T-ratio=3.5050 and

161   Gamma distribution parameter= 0.2005). Node support for MP and ML analyses was

162   determined using a non-parametric bootstrap, with 500 and 1000 replicates respectively

163   (Holmes, 2003). For the analysis 3x106 generations were run, with one tree retained

164   every 300th generation and the first 2500 trees discarded as burn-in. Genetic distances

165   were calculated using p-distance in Mega and median-joining networks were drawn

166   using NETWORK and dnasp4.

167

168 AFLP analysis was adapted from (Ajmone-Marsan et al., 1997) with approximately

169 200ng of genomic DNA extracted from $C1_{Pb*}^{pH5}$ (n=24), $C2_{Pb*}^{pH4}$ (n=30), $C3_{Pb***}^{pH7}$ (n=23) and

170 $C4_{Pb*}^{pH6}$ (n=18) individuals. Pre-selective EcoR1 (GACTGCGTACCAATTCA) and Taq1

171 (GACTGCGTACCAATTCC) primers were used and for the selective amplifications

172 two primer combinations (E32 (5'-GACTGCGTACCAATTCAAC-3') /T32 (5'-

173 GATGAGTCCTGACCGAAAC-3') and E32/T38 5'-GATGAGTCCTGAC CGAACT-

174 3') were employed. Reactions were run by the Cardiff University Molecular Biology

175 Support Unit and analysed on an Applied Biosystems 3130x1 fragment analyser. Bands

176 between 70 and 325 base pairs (bp) and with a minimum peak height of 70 units were

177 scored using GeneMapper analysis software. Microsoft Access, Excel and the Excel

178 macro GenAlEx6 were used to create a cumulative table of all loci from each individual

179 and transform the data into a binary form. Principal coordinates (PCO) analysis and

180 phylogenetic tree construction, supported by bootstrap analysis (1000 replicates), was

181 performed using the neighbour joining algorithm (based upon the Nei's distance) to in

182 PAUP v4.0b10 to estimate tree topology and calculate branch length. The software

183 Structure v2.2 was used to delineate clusters of individuals on a multi-locus, genotype

184 basis using a Bayesian algorithm. The number of inferred populations ran from 1 to 5,

185 with 8 replicate runs, a burn-in of 75 000 cycles followed by 300 000 for data

186 collection. L(K), the modal choice criterion, is calculated in Structure and the true

187 number of populations (K) can be deferred from its maximal value. ΔK, the rate of

188 change in the log probability of data between successive K-values, provides a visual

189 means to easily identify the number of clusters in a sample of individuals (Evanno et al.,

190 2005).

191

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

216   according to the NERC Environmental Genomics scheme to allow for bioinformatics

217   analysis. The raw trace chromatograms from the sequencing reaction were processed

218   using trace2dbEST (Sturzenbaum et al., 2003) which contains a base calling component

219   (phred) and a sequence trimming component (cross_match). This software produces

220   good quality EST sequences, formatted for submitting to NCBI dbEST

221   (http://www.ncbi.nlm.nih.gov/dbEST). The EST sequences were clustered using

222   CLOBB (Sturzenbaum, et al., 2003) to derive a consensus putative gene sequence

223   contig and then processed by the software package PartiGene (Sturzenbaum, et al.,

224   2003) (http://www.nematodes.org/PartiGene). Cluster information can be retrieved by

225   LumbriBASE (http://www.earthworms.org) through simple text queries, identification

226   of sequence similarity and library specific searches. The biological process and

227   molecular function of gene sequences were described by defining their Gene Ontology

228   (GO) classification using blast2go (http://www.blast2GO.de).

229

230   <u>Fourier-transform Infrared Spectroscopy</u>. Soil and adult (fully clitellate) *L. rubellus*

231   earthworms were collected from $C2_{Pb*}^{pH4}$ and $C3_{Pb***}^{pH7}$ and the posterior segments

232   immediately excised and quench-frozen in liquid nitrogen. The frozen tissue was

233   transported to the laboratory under liquid nitrogen and stored at -20°C until required.

234   Tissues were mounted in CryoEmbed and sectioned longitudinally at a nominal

235   thickness of 50μm in a Bright cryostat. Sections were mounted on Kevley slides and air-

236   dried overnight in the cold chamber of the cryostat. Infra-red spectra were collected in

237   transmission mode from station 11.1 at the CLRC Daresbury Synchrotron Radiation

238   Source. The chloragogenous tissue was visually identified and each section imaged and

239   analysed, with five spectra from five different regions of the tissue (i.e. x25 spectra per

240   individual earthworm) collected.

241

242   <u>SERCA</u>. Plasmid preparations of individual LumbriBASE clones (Genbank accession

243   numbers CF416761 and CO048347) were prepared using a Wizard® Plus SV Miniprep

244   kit (Promega Ltd., Southampton, UK). Preparations were sequenced in their entirety by

245   "walking" along the gene, after each step re-designing a specific reverse primer to

246   complement the universal M13 forward. Primers were designed using the software

247   Primer3 (Rozen and Skaletsky, 2000) and Oligo® (MBI Inc, USA) and sequencing

248   performed as described above. These full-length library sequences were used to design

249   *L. rubellus* specific SERCA primers in order to amplify the gene transcribed in

250   individuals of each genealogical lineage. Reactions were denatured at 95°C for 10

251   minutes and then cycled 35 times at 95°C for 30 seconds, 30 seconds at the required

252   primer annealing temperature and 72° for 1 minute. This was followed by a 10 minute

253   final extension at 72°C. DNA was sequenced as described above by the Cardiff

254   University Molecular Biology Support Unit.

255

256   Total RNA was extracted from tail-clips of an adult individual sampled from $C3_{Pb***}^{pH7}$ and

257    $C2_{Pb*}^{pH4}$ using the Tri-reagent method (Sigma-Aldrich, UK). Complementary DNA

258   (cDNA) was synthesised from messenger RNA (mRNA) using reverse transcriptase.

259   Total RNA (7-20μg) was heated at 65°C for 3 minutes and combined with anchored

260   oligo d(T) (1μl, 100mM) and random hexamers (2μl, 100mM) and incubated at 70°C

261   for ten minutes. A reaction mix of 5X RT buffer (6μl), DTT (3μl, 0.1M) and dNTP mix

262   (1.2μl, 10mM) was prepared and added to the RNA mix and incubated at 25°C for two

263   minutes. Superscript (1µl) was added and the reaction incubated at 42°C for 3 hours. A

264   series of three PCR reactions were optimised and performed in order to obtain the full-

265   length SERCA sequence of each individual; between sequenced sections of the gene

266   there was a large overlap to ensure the same SERCA isoform was being amplified in

267   each instance. PCRs were performed as described above. Reactions that yielded

268   products >2000bp were modified. For these reactions DNA (~100ng) template was

269   amplified using 10µM forward and reverse primer, 25mM dNTP mix and 1µl

270   Herculase® II Fusion DNA Polymerase buffered with 5X Herculase® II PCR reaction

271   buffer (Stratagene Europe, Amsterdam, The Netherlands). Each reaction was

272   supplemented with an optimised quantity of $MgCl_2$ (25mM). The reaction was

273   denatured at 95°C for 10 minutes and then cycled 35 times at 95°C for 20 seconds, 20

274   seconds at the required primer annealing temperature and 68° for 4 minutes. This was

275   followed by a 4 minute final extension at 68°C. Protein sequences were aligned using

276   bioinformatic software tool Mega v3.1 (Kumar, et al., 2004) and modelled using

277   SWISS-MODEL (http://swissmodel.expasy.org//SWISS MODEL.html), Swiss-

278   PdbViewer and Pymol.

279

280   Total RNA was extracted from nine, previously genotyped (mtDNA) adult individuals

281   sampled from $C1_{Pb*}^{pH5}$, $C2_{Pb*}^{pH4}$, $C3_{Pb***}^{pH7}$ and $C4_{Pb*}^{pH6}$. This was followed by cDNA synthesis,

282   as described above. A PCR was designed to enable easy identification by gel

283   electrophoresis of the expressed isoform, with a combination of three primers used in

284   each reaction (F1 5'-CTGGCCGGAATTCGTGTTATC-3', F2 5'-ATACTCTTCG

285   CTGTCTTGCGT-3', R1 5'-CCGCTGGCTCTTCTTCCG-3'). The two forward primers

286   were designed so that each one isolated one of the two isoforms. The resulting products

287 were of different sizes to enable simple identification on an agarose gel following

288 resolution by electrophoresis.

289

290 Results and Discussion

291

292 The 'field laboratory': the metalliferous site and its resident earthworms. Mine sites are

293 notoriously heterogeneous in geochemical nature. Mineral (galena, PbS) extraction from

294 the Cwmystwyth Mine stopped in 1921. Today, the contaminated surface spoils are

295 situated on base-poor underlying geology, punctuated by calcareous 'islands' around

296 derelict buildings (Figure 1). In the decades since its abandonment, the spatially

297 chequered and hostile site has developed micro-habitats colonised by a limited variety

298 of naturally occurring plants and invertebrates. Thus, it serves as an ideal evolutionary

299 field laboratory. *L. rubellus* was sampled from four sites across the

300 mine: $C1_{Pb*}^{pH5}$, $C2_{Pb*}^{pH4}$, $C3_{Pb***}^{pH7}$, and $C4_{Pb*}^{pH6}$ (the number of asterisks denotes the level of

301 contamination as classified by the Kelly index (ICRCL 59/83): * contaminated (1000-

302 2000ppm), ** heavily contaminated (2000-10 000ppm) and *** unusually heavily

303 contaminated (>10 000ppm) (Figure 1). Population divergence was measured using

304 amplified fragment length polymorphism (AFLP) analysis and mitochondrial

305 cytochrome oxidase II (mtDNA COII) gene sequence data of individuals split between

306 the four sites. Two distinct lineages, differentiated at both the mitochondrial and nuclear

307 level, were revealed with a mean inter-lineage mtDNA sequence divergence of

308 approximately 13%, indicative of a cryptic species complex (Figures 2A and B).

309 Moreover, from the distinct clustering of $C3_{Pb***}^{pH7}$ individuals a true genetic archipelago

310 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 ($C2_{Pb*}^{pH4}$) and one calcareous ($C3_{Pb***}^{pH7}$), 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

359 spectrum) were apparent, which correlated with phosphorous-containing functional

360 groups (Figure 3A and B). As the earthworm chloragocyte is thought to be involved in

361 haem biosynthesis (Jamieson and Molyneux, 1981), a conserved pathway that is

362 inhibited by Pb at several junctures (Warren, et al., 1998), Pb trafficking into and across

363 earthworm chloragocytes must be tightly regulated in these animals that are

364 continuously exposed to high concentrations of metal in their native environments and

365 whose strategy for dealing with it involves intracellular accumulative immobilization.

366 Indeed, inferences on the mechanisms of adaptive evolution to environmental

367 heterogeneity require not only abstract genotype- to phenotype associations but more

368 meaningful molecular genetic interpretations regarding the nature of induced

369 phenotypic variation. The transcriptomic profile of an organism provides a snapshot of

370 gene expression to provide information regarding developmental stage, life-history or

371 responses in relation to particular environmental stressors. EST libraries are also the

372 substrate for comparative genomic studies, through investigating differential expression

373 between cDNA populations. Two libraries were constructed from earthworm

374 populations with contrasting histories of Pb exposure; $C3_{Pb***}^{pH7}$ earthworms, chronically

375 exposed in the field, and earthworms sampled from a clean reference site, R1, acutely

376 exposed to lead in the form of $Pb(NO_3)_2$ under laboratory conditions. In combination

377 with the plethora of EST cluster information already available in LumbriBASE

378 (www.earthworms.org), originating from control (adult, head-enriched), life-stage (late

379 cocoon, juvenile) and other exposure (Cu, Cd, fluoranthene, atrazine) libraries, a metal

380 tolerant genotype may be related to phenotype and the functional systems that underlie

381 lead handling within these earthworm populations defined. Both libraries comprised

382 high quality sequences with an average length of between 500 and 600 base pairs. This

383 ensured the maximum numbers of sequences were annotated to enable accurate

384 downstream analysis and interpretation using the software LumbriBASE, Blast2GO and

385 associated KEGG resource, which generates pathway maps that highlight gene ontology

386 relationships between annotated sequences. Of interest was the significant number of

387 $C3_{Pb****}^{pH7}$ gene products (when compared to other libraries) associated with intracellular

388 Ca^{2+} sensing (Calmodulin, Troponin C) and buffering (Sarcoplasmic calcium binding

389 protein (SCP), Parvalbumin) (Gao et al., 2006; Ishida and Vogel, 2006). Both sets of

390 proteins belong to the EF-hand super-family of proteins implicated in calcium binding

391 and central to the Ca-signalling pathway. These observations indicate that components

392 of the Ca signalling pathway are central to Pb sequestration within chloragocytes which,

393 in turn, may be associated with adjustments in the metabolism of their common

394 complexing PO_4^- anion. This yields a number of candidate loci that may contribute to a

395 Pb-tolerance phenotype by modifying molecules involved in the cellular physiology of

396 an essential cation (Ca^{2+}) to accommodate its non-essential cationic mimic (Pb^{2+}).

397

398 <u>Sarcoplasmic/ endoplasmic reticulum calcium ATPase (SERCA).</u>

399 Sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA) is a central transport

400 carrier protein of the Ca signalling pathway that resides in membranes of intracellular

401 storage sites for the uptake of excess Ca and, conceivably, Pb (Tsien et al., 1987). Three

402 isoforms have been described in vertebrates (MacLennan et al., 1985; Campbell et al.,

403 1991; Vilsen and Andersen, 1992) and one in invertebrates (Palmero and Sastre, 1989;

404 Escalante and Sastre, 1994; Shi et al., 1998a; b), fungi (Ghislain et al., 1990) and plants

405 (Wimmers et al., 1992). All isoforms are similar in structure and have a 75-85%

406 identical amino acid sequence (Toyofuku et al., 1992). Despite the identification of

407   several vertebrate isoforms, *L. rubellus* is the first invertebrate found to harbour

408   multiple SERCA proteins. Two structurally different forms were identified in

409   populations, sampled from different field sites, and expression was found to be co-

410   incident with the mitochondrial lineage marker (COII), even where nuclear

411   hybridisation was observed (Figure 4A and B). Their structure differed in amino acids

412   located in the cytosolic nucleotide-binding domain (the flap) of the protein, a region

413   thought to have a critical role in determining calcium affinity and turnover (Figure 4C).

414   This observation indicates that not only are the intracellular trans-membrane Ca and Pb

415   pathways confluent at the molecular (SERCA) level and are associated with adjustments

416   in the metabolism of their common complexing PO_4^- anion, but the entire machinery is

417   demonstrably genotype-specific (Figure 4C). It is important to point out, however, that

418   whist the SERCA molecule is an important connector between the vulnerable cytosol

419   and the intra-vesicular depository of immobilized Pb, other components of the Ca

420   pathway warrant study to determine whether they are structurally or functionally

421   modified.

422

423   Conclusion

424

425   Findings presented here demonstrate that tolerance in earthworms with a protracted

426   history of Pb exposure is not merely an expression of genetic divergence, rather that

427   field populations may evolve Pb-adaptation traits by modifying molecular regulators of

428   Ca physiology. Whilst adaptive changes in enzyme structure are, for good reason, less

429   probable than changes in the promoters that regulate enzyme expression (Crawford et

430   al., 1999b; a), they are clearly not molecular traits that can be ignored, as the similarity

431   in the ionic radii of Ca^{2+} (1.00Å) and Pb^{2+} (1,19Å) (Bridges and Zalups, 2005) may, for

432   example, facilitate structural mutations. Together, these observations describe a

433   molecular mechanism for Pb resistance in earthworms that may underpin their

434   phylogenetic differentiation within fairly discrete microhabitats across the study site.

435   Furthermore, the concept serves as a paradigm for invertebrates generally because the

436   ability for incorporating Pb into 'calcospherites' is ubiquitous, demonstrating that

437   evolution innovates by modifying existing structures or pathways.

438

446

447   References

448   Ajmone-Marsan, P., Valentini, A., Cassandro, M., Vecchiotti-Antaldi, G., Bertoni, G.,

449   Kuiper, M., 1997. AFLP markers for DNA fingerprinting in cattle. Animal Genetics 28,

450   418-426.

451   Arnold, R.E., Hodson, M.E., 2007. Effect of time and mode of depuration on tissue

452   copper concentrations of the earthworms *Eisenia andrei*, *Lumbricus rubellus* and

453   *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.

502 Janssens, T.K.S., Marien, J., Cenijn, P., Legler, J., van Straalen, N.M., Roelofs, D.,

503 2007. Recombinational micro-evolution of functionally different metallothionein

504 promoter alleles from Orchesella cincta. BMC Evolutionary Biology 7, 88.

505 Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular

506 Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5,

507 150-163.

508 MacLennan, D.H., Brandl, C.J., Korczak, B., Green, N.M., 1985. Amino-acid sequence

509 of a $Ca^{2+} Mg^{2+}$-dependent ATPase from rabbit muscle sarcoplasmic reticulum, deduced

510 from its complementary DNA sequence. Nature 316, 696-700.

511 Morgan, A.J., Kille, P., Sturzenbaum, S.R., 2007. Microevolution and ecotoxicology of

512 metals in invertebrates. Environmental Science & Technology 41, 1085-1096.

513 Morgan, J.E., Morgan, A.J., 1989. The Effect of Lead Incorporation on the Elemental

514 Composition of Earthworm (Annelida, Oligochaeta) Chloragosome Granules.

515 Histochemistry 92, 237-241.

516 Morgan, J.E., Morgan, A.J., 1990. The distribution of cadmium, copper, lead, zinc and

517 calcium in the tissues of the earthworm *Lumbricus rubellus* sampled from one

518 uncontaminated and four polluted soils. Oecologia 84, 559-566.

519 Morgan, J.E., Richards, S. P., Morgan, A. J., 2001. Stable strontium accumulation by

520 earthworms; a paradigm for radiostrntium interactions with its cationic analogue,

521 calcium. Environmental Toxicology and Chemistry 20, 1236-1243.

522 Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author.

523 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

594 properties were analysed for their mitochondrial (panel A) and nuclear (panel B)

595 genotype. Sites included $C4_{Pb*}^{pH6}$ (blue triangles) and $C1_{Pb*}^{pH5}$ (purple diamonds), at the

596 boundary of mine, together with $C3_{Pb***}^{pH7}$ (grey circles) and $C2_{Pb*}^{pH4}$ (red squares).

597 A. Median-joining network analysis based upon 440 bp sequence of the COII

598 mitochondrial gene of 119 *L. rubellus* individuals. The size of each haplotype group

599 within the network is proportion to the total number of individuals attributed to the

600 genotype whilst the earthworm source is indicated by fill colour. The left and right hand

601 branches of the network are denoted lineage A & B respectively. B. AFLP multi-locus

602 profiling principal component analysis showing individuals from the four sample

603 stations. Those individuals exhibiting mitochondrial lineage B genotype are circled in

604 red whilst those attributed to lineage A are circled in grey. The clustering of $C3_{Pb***}^{pH7}$

605 earthworms distinct from other lineage A earthworms is indicated with a dotted grey

606 line. Hybrid individuals are shown by a lack of fill colour. C and D. Arlequin simulated

607 mitochondrial mismatch distributions, using the model of demographic expansion, of

608 lineage A (C) and lineage B (D) haplotypes respectively. The solid lines are the

609 observed mismatch distribution and the dotted line shows the distribution simulated

610 under the expansion model. E. Associated sum of squared deviation (SSD), Raggedness

611 (Rg) and p-value statistics, based on 1000 data bootstraps.

612

613 Figure 3. Metabolomic fingerprinting of earthworm chloragogenous tissue using Fourier

614 Transform Infrared spectroscopy. A. The fingerprint region of averaged infra-red

615 spectra of earthworm chloragogenous tissue collected from $C3_{Pb***}^{pH7}$ (grey) and $C2_{Pb*}^{pH4}$

616 (red). Individual spectra were processed by the software package OPUS$^{©}$). B. The main

617 difference in $C3_{Pb^{***}}^{pH7}$ and $C2_{Pb^{*}}^{pH4}$ averaged spectra (~1080cm^{-1}), corresponded to

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

619 illustrating the clustering of $C3_{Pb^{***}}^{pH7}$ and $C2_{Pb^{*}}^{pH4}$ 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 $C4_{Pb^{*}}^{pH6}$ (blue triangles) and $C1_{Pb^{*}}^{pH5}$ (purple diamonds) at the

625 boundary of the mine and contaminated sites $C3_{Pb^{***}}^{pH7}$ (grey circles) and $C2_{Pb^{*}}^{pH4}$ (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