Ecological drivers influence the distributions of two cryptic lineages in an earthworm morphospecies

David J. Spurgeon*[1], Manuel Liebeke[2*], Craig Anderson[1,3], Peter Kille[4], Alan Lawlor[5], Jacob G. Bundy[2], Elma Lahive [1]

[1] Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Wallingford OX10 8BB, UK

[2] Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London SW7 2AZ, UK

[3] Biological and Environmental Sciences, School of Natural Sciences, University of Stirling, Stirling, FK9 4LA, UK

[4] School of Biosciences, University of Cardiff, Main Building, Museum Avenue, Cardiff CF10 3AT, UK

[5] Centre for Ecology and Hydrology, Library Avenue, Bailrigg, Lancaster, LA1 4AP, UK

*Current address: Department of Symbiosis, Max Planck Institute for Marine Microbiology, Bremen, Germany

Corresponding author full contact details: Dr David Spurgeon - Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK. Tel: 01491 772 208, dasp@ceh.ac.uk.

ABSTRACT

Substantial genetic diversity exists within earthworm morphotypes, such that traditional species designations may be incomplete. It is, however, currently not known whether these different genetic variants show ubiquity or specialty in their distribution across separated sites subject to different climatic, biotic or soil physicochemical factors. Here we report on the results of a survey in which individuals of the *Lumbricus rubellus* morphotype, a species known to comprise two deeply divergent genetic lineages in England and Wales, were sampled from 26 plots. Sequences from the mitochondrial cytochrome oxidase I gene were used to distinguish lineages for 787 individuals. In conjunction, a range of geographic, climatic, biotic and soil physiochemical variables were also collected for each locality.

Genotyping indicated that Lineage A was more common than Lineage B, comprising 58% of the collected *L. rubellus*. Six site populations comprised only Lineage A, while only a single site comprised entirely Lineage B. The remaining 20 sites containing both lineages. A multivariate ordination of site variables identified major difference between sites were associated with low pH, organic-rich soils in Western wet upland areas and pollutant levels associated with sites in the South. Earthworm genotype (as proportion of Lineage A) was not correlated with either of these major environmental axes. When individual variables of soil pH and the percentage of soil organic matter, which are known to be key driver of soil species distributions, were investigated as single variables significant relationship with lineage frequency were found. Soil organic matter content was significantly negatively correlated with Lineage A proportion, while pH was significantly positively correlated. This lineage preference may be related to lineage metabolism and/or behavioral differences.

Measurement of tissue metal concentrations in worms from 17 sites identified a significant site effect in all cases, but a lineage effect only for arsenic (higher Lineage B). Tissue arsenic concentrations varied between lineages, supporting previous observations that there are differences in the way the two lineages have adapted to manage exposure to this metalloid.

56 Keywords: Biogeography, Earthworm, Cryptic species, pH, Soil organic matter

1. INTRODUCTION

Soils contain a wealth of invertebrate biodiversity recognised for their important contributions to ecological processes (Bardgett and van der Putten, 2014; Fitter et al., 2005; Giller, 1996). One key group of species are the “ecosystem engineers”: those organisms that modify the physical state of the soil and resource availability for other species. Earthworms are known as a key group of ecosystem engineers in many habitats. They perform a range of physical (aeration, bioturbation, litter fragmentation) and biological (microbial interactions, exudate production) roles in soil (Blouin et al., 2013; Lavelle et al., 1997; Sackett et al., 2013; Umarov et al., 2008). Because of their functional importance, earthworms have emerged as a major taxon for biomonitoring and biomarker assessments of human induced pressures on soil communities (Cluzeau et al., 2012; Rutgers et al., 2009).

As soil invertebrate species, including earthworms, have been shown to be sensitive to a range of land use change and pollution impacts (Bundy et al., 2007; Cluzeau et al., 2012), different soil taxa have become a natural focus for research on the relationships between environmental pressures, biodiversity and soil functioning (Bartlett et al., 2010; Leveque et al., 2015; Rutgers et al., 2016). For community studies, a major constraint relates to current uncertainties in earthworm taxonomy. Traditionally earthworm identification has relied on morphology, but the paucity of suitable local keys and problems with application to juveniles has also recently encouraged the use of molecular methods (Dominguez et al., 2015; Emerson et al., 2011; Klarica et al., 2012). These genotyping studies have begun to challenge current understanding of diversity through the identification of genetically distinct cryptic lineages within previously established morphospecies.

Earthworm species in which cryptic lineage diversity has to date been identified include *Eisenia fetida*/*andrei* (Römbke et al., 2016), *Lumbricus terrestris* (James et al., 2010), *Aporrectodea caliginosa* (PerezLosada et al., 2009), *Allolobophora chlorotica* (King et al., 2008), *Amynthas gracilis* / *Amynthas cortici* (Novo et al., 2015) and *Lumbricus rubellus*. For

L. rubellus, genotyping studies based on mitochondrial cytochrome oxidase I and II markers have identified as many as 6 cryptic lineages across Europe (Giska et al., 2015), two of which are found in the UK (Andre et al., 2010; Kille et al., 2013). The two UK lineages have 10-15% divergence for the mitochondrial COI and COII sequences. While this implies they may actually be cryptic species, recent analysis of multiple nuclear markers using RADseq has not supported this interpretation, instead suggesting that different *L. rubellus* lineages may actually correspond to a single highly polymorphic species (Giska et al., 2015). Comparative studies of the two lineages in the UK have, nonetheless, identified physiological differences between them, including variation in pheromone production (Jones et al., 2016), maturation time (Anderson et al., 2013), metabolic profiles (Liebeke et al., 2014), mechanism of arsenic adaptation (Kille et al., 2013), trace element metabolism (Andre et al., 2010), and microbiome complement (Pass et al., 2015).

Despite known biological differences, the extent to which differences in distribution and physiology are related to different geographical, climate and soil physicochemical preferences between the two known UK lineages of *L. rubellus* is not established. The two lineages found co-occur at some, but not all, sites meaning that they have some likely niche divergence that facilitates coexistence (Andre et al., 2010; Giska et al., 2015; Kille et al., 2013). We aim to better understand the nature of the spatial and geochemical drivers of lineage relative abundance, and so here we test the hypothesis that the site distribution of the two cryptic *L. rubellus* lineages is based on one or more geographical, climatic, physiochemical or biotic drivers. We collected and genotyped morphotype *L. rubellus* at multiple well-characterized sites that differed in their properties to investigate the relationships that determine lineage distributions. Tissue metal concentrations were also measured to assess if trace metal levels could also influence distributions, as could be the case if the two lineages had different sensitivity to specific contaminants.

2. METHODS

2.1 Site selection

Twenty six sites located across England and Wales (Fig. 1) were visited between four times (for Devon Great Consols Mine and Control, Shipham Mine and Control, Cwmystwyth Mine and Control) and a single visit (for Porton Down, Parys Mountain, Castell, Clydach, Roman Gravel, Didcot) over four separate sampling events from Spring 2011 to Spring 2014. The chosen sites were selected to capture a range of the habitats and soil conditions under which morphotype *L. rubellus* can be collected. Land-uses covered included arable systems, broadleaf woodland, rough grassland and improved pasture habitats. Sites included both mineral and organic soils, although not true peats.

To allow the role of soil geochemistry and pollution status on lineage distribution to be addressed, sites of different known pollution history were sampled. Sites corresponded to three groups with respect to past land-use and associated expected contamination level. These were: 1) sites with no known pollution source (Unpolluted); 2) sites near to industrial facilities expected to be characterised by moderate pollution (Industrial polluted); and 3) sites at abandoned mining sites that can be expected to have high pollution (Mine polluted). For expected polluted sites from categories 2 and 3, a local control site was also sampled. This reference site was located outside of the area that was expected to be strongly influenced by the main pollution source and so was on soil expected to contain regional background pollutant concentrations.

2.2 Site geographical, biological and soil physiochemical characterisation

To allow the assessment of environmental drivers relating to lineage distribution, we used both publically available resources as well as our own analyses to gather data on each sites. Site geographical locations were collected as Easting and Northings from www.gridreferencefinder.com and site altitudes from www.freemaptools.com/elevation-

finder.htm. A series of site climate conditions were also assembled from www.metoffice.gov.uk/. These were: annual average maximum temperature, annual average minimum temperature, average January minimum temperature, average July minimum temperature, average annual rainfall, average annual rain days and average annual frost days. Initial visits to each site recorded main land-use (arable, broadleaf woodland, rough grassland and improved pasture) and where present the average sward height of vegetation at collection locations. The site was identified according to the level of shade (open, part shaded, shaded) and the presence of livestock was noted.

An initial site survey identified points on the site where morphospecies *L. rubellus* could be found. Thereafter all collections were focussed on these locations. For any one sampling event at each site, between 6 and 25 fully clitellate *L. rubellus* were collected by digging and hand-sorting from the soil to 20 cm depth. Generally the required number of worms could be collected within a reasonable search period (approximately 2 h duration). There were, however, some locations where this was not possible for particular sampling events. Climate factors (notably dry soils), low frequency of adults in the population or the requirement to limit site damage caused by digging were the major constraints. During collection, the presence of other earthworm morphospecies was noted. Only common species were recorded (>5 individuals observed). In total 10 other species were found: *Aporrectodea caliginosa*, *Aporrectodea rosea*, *Aporrectodea longa*, *Allolobophora chlorotica*, *Lumbricus castaneus*, *Dendrobaena rubida*, *Lumbricus terrestris*, *Lumbricus festivus*, *Octolasion cyaneum*, and *Octolasion tyrtaeum tyrtaeum*. At the end of sampling, the *L. rubellus* collected were washed and blotted dry on-site and then snap frozen in liquid nitrogen before being transfered to the laboratory under dry ice storage.

Triplicate soil samples from surface to 5 cm depth were collected from each site collection location. All soil samples were oven dried at 80°C to constant weight and then sieved

through a 2 mm mesh to remove large roots and stones. Total concentrations of aluminium, arsenic, barium, cadmium, cobalt, chromium, copper, iron, lead, manganese, mercury, molybdenum, nickel, selenium, titanium, vanadium, zinc, calcium and total phosphorous were determined in a 1 g sample of this processed soil following an aqua regia digestion protocol (Arnold et al., 2008; Emmett et al., 2010; Spurgeon et al., 2008). Digests were subsequently analysed on a Perkin Elmer Optima 7300 DV inductively coupled plasma optical emission spectrometry instrument. For quality control, an in house reference traceable to BCR-143R (Commission of the European Communities, Community Bureau of Reference) was included with each batch of digestions. Measured concentrations were within 10% of certified values for all measured elements with the exception of Al where the value was 55%. Organic matter content of each soil sample was measured by proxy using loss on ignition following combustion at 500°C (Rowell, 1994) and soil pH was quantified by electrode from a 1:2.5 volume soil:water mix (i.e. 1 volume soil with 2.5 volumes water added)(International Organisation for Standards, 2005).

2.3 Lineage assignment by mitochondrial cytochrome oxidase I (COI) sequencing

DNA was extracted from ~10 mg of frozen tissue (taken from the tail of each individual using a scalpel) by automated DNA extraction using a Nucleoplex Plants Tissues DNA Extraction Kit (Nucleoplex, Manchester, UK). After DNA quantification using Nanodrop (Thermo Scientific, Willmington, DE), polymerase chain reaction amplification of the COI gene was conducted using a set of established forward (GGTCAACAAATCATAAAGATATTGG) and reverse (TAAACTTCAGGGTGACCAAAAAATCA) primers (Folmer et al., 1994) amplified after 5 minutes at 95°C over 40 cycles of 30 sec 95°C, 30 seconds 48°C and 60 seconds 48°C. A sub-set of all PCR products were checked by gel electrophoresis to ensure successful amplification and purified for sequencing using 0.25 U each of Exonuclease I and Shrimp Alkaline Phosphatase (NEB, Hitchin, UK), incubated at 37°C for 45 minutes and 80 °C for 15 minutes. Purified PCR products were then sequenced as in Andre et al. (2010),

192 using ABI PRISM® BigDye v3.1 Terminator Sequencing technology (Applied Biosystems,
193 USA).
194
195 Sequences were aligned and trimmed for tree construction using the Maximum Likelihood
196 method and General Time Reversible substitution model with a gamma distribution in Mega
197 v5.01. Sequences for *L. rubellus* associated with specific mitochondrial lineages already
198 documented in the UK were incorporated into the analysis as anchor sequences (Anderson
199 et al., 2013), with sequences for *L. terrestris*, *L, festivus* and *L. castaneus* included as an
200 out-group. Tree topology was supported by bootstrap analyses over 1000 iterations.
201 Individuals that showed a close relationship with one of the two previously identified UK *L.*
202 *rubellus* lineages were identified from the analysis. Any individuals showing intermediate
203 status resulting from probable sequencing errors were excluded from further analysis.
204
205 *2.4 Earthworm tissue trace element concentrations*
206 Earthworm tissues from 494 individuals taken from a sub-set of 17 sites (Alice Holt ECN
207 Control, Avonmouth Control, Avonmouth Incinerator, Avonmouth Savalco, Cwmystwyth
208 control, Cwmystwyth mine, Devon Great Consols Control, Devon Great Consols Mine,
209 Drayton ECN Control, Port Talbot Control, Port Talbot blast furnace, Porton Down ECN,
210 Scunthorpe blast furnace, Scunthorpe Control, Shipham control, Shipham mine, Snowdown
211 ECN control) were prepared for analysis (nb samples from remaining sites were lost due to
212 storage issues). These samples were analysed for tissue Al, As, Cd, Co, Cr, Cu, Fe, Mn, Mo,
213 Ni, Pb, Se, Sr and Zn concentrations. Whole earthworms, after tail removal for DNA
214 extraction, were initially ground to powder under liquid nitrogen in a cryogenic mill. The
215 powder was freeze-dried and a 100 mg sample digested with 10 ml of 70% HNO_3 (Ultrapure)
216 at 200^{o}C for 15 minutes within a microwave vessel. Samples were run as two batches on a
217 Perkin Elmer DRCII ICP-MS. Each batch included multiple certified reference material
218 samples for TORT-2 and DOLT-4 (National Research Council, Canada). Certified values for
219 reference materials corroborated well with measured values. Average recovery was 91%

220 (range 85% for Se to 110% for Pb) in the first batch of samples and 94.8% (range 53.9 for Al
221 to 129% for Se) in the second batch. Recoveries of only two metals, Al and Se, were outside
222 80% of certified values for any run, with 19 of 27 determinations within 10%. With systematic
223 bias absent, acquired data can be used for statistical processing without requirement for
224 recovery correction.

225

226 *2.5 Data handling and statistical analysis*

227 The number of *L. rubellus* returning COI sequences that were closely related to reference
228 sequences from previously collected lineage A and Lineage B individuals were counted for
229 each study site. These were calculated as proportions before being logit transformed as the
230 most appropriate transformation for biological proportion data (Warton and Hui, 2011), with
231 value of zero and one modified by addition and subtraction of half of the lowest proportion
232 respectively. Environmental drivers were established as either categorical (e.g. site type, site
233 shading, livestock presence/absence, earthworm species presence/absence) or as
234 continuous measured variables. The values for soil metal concentrations were log
235 transformed to obtain a Gaussian distribution in accordance with established practice
236 (Davies, 1989).

237

238 Relationships amongst site geographical, climate and soil variables (after appropriate
239 transformation) were initially investigated using principal component analysis in Minitab 14
240 (Minitab, PA, USA). This reduces the dimensionality of the original complex dataset in an
241 unsupervised fashion, by successively generating new axes (principal components), that are
242 the linear combinations of the original data that explain greatest overall variance. The
243 principal components (PCs) arising can be interpreted as ordinations representing a
244 summary of environmental factors. Pearson correlations between (logit transformed)
245 proportions of Lineage A individuals at each site and the individual principal component
246 scores were then calculated to investigate the relationships between ordinated site

247 characteristics and genotype. Based on this analysis and prior knowledge, possible
248 individual primary driver variables were identified that were in turn assessed for Pearson
249 correlation with genotype. Because of this key role, percentage organic matter content and
250 soil pH were selected as focus variables. Site and lineage effects on earthworm tissue log
251 transformed trace element concentrations were analysed using a mixed model based
252 general linear model in Minitab 14. Within the model, site and lineage were included as fixed
253 variables, with sampling campaign (1-4) as a random factor.

254 *3. RESULTS*

255 High quality *L. rubellus* COI sequences were obtained from DNA samples taken from 787
256 earthworms for assignment as either Lineage A or Lineage B individuals. The maximum
257 number of sequences from any one site was 73, from Cwmystwyth Mine, and the minimum
258 3, from Avonmouth Control (Fig. 1). In total, 457 individuals were assigned as Lineage A,
259 58% of the number collected. The remaining 330 (42%) were assigned as Lineage B. Eight
260 sites (Avonmouth Savalco, Avonmouth Incinerator, Clydach Smelter, Didcot Power Station,
261 Dinas Powys, Parys Mountain, Scunthorpe Control and Scunthorpe) had populations
262 comprising only Lineage A individuals. Seven sites (Avonmouth Control, Castell Mine,
263 Cwmystwyth Control, Cwmystwyth Mine, Drayton ECN Control, Shipham Control and
264 Shipham Mine) contained populations comprised largely Lineage B, although only Castell
265 Mine was exclusively B. All remaining sites had mixed lineage populations, although with
266 more Lineage A than B individuals.
267
268 The site geographical, physical and soil characteristics were analysed using principal
269 component analysis. The first PC explained 21.4% of total variance. A number of
270 parameters were positively correlated with this axis, including soil % loss on ignition (LOI);
271 soil log Fe, log Co and log Al concentrations; some earthworm species; and site altitude and
272 climate variables including average rainfall and number of rain days. Negatively correlated
273 variables included soil pH; average July max temperature and average temperature; log Ca
274 and log P concentrations; and Easting (Fig 2). This first PC axis could therefore be
275 interpreted as representing a set of variables characterised by the presence of high organic
276 matter, low pH soils, associated with wetter and colder upland regions located mainly in the
277 West of England and Wales. The second PC axis explained a further 17% of variation, and
278 was positively associated with Northing and the weather variable of average frost days and
279 average rain days. Variables negatively associated with this axis included pollutant metal
280 concentrations such as log Pb, log Zn and log Cd concentrations (Fig 2). This axis can be

281 interpreted as representing a gradient of metal contamination of sites located primarily in the
282 South of England and Wales.
283
284 To assess if the site characteristics summarised by the two first PCs potentially act as
285 drivers of lineage distribution, the PC1 and PC2 scores were correlated with the (logit
286 transformed) Lineage A proportion at each location. This did not identify significantly
287 relationships between Lineage A proportion with site PC1 or PC2 score (p=0.25 and
288 p=0.074 respectively). As sites PC scores were not significant, we next went on to
289 investigate if individual variables measured are related to the relative frequency of *L.*
290 *rubellus* lineages. Specifically we selected soil pH and LOI for initial assessment, as these
291 are established as drivers of patterns of diversity (Griffiths et al., 2011; Raty and Huhta,
292 2003). Both variables were significantly correlated with logit Lineage A proportion (soil % OM
293 -0.529, p=0.005; pH -0.392, p=0.048). The nature of these two relationships were
294 summarised by locally weighted scatterplot smoother model fits. These indicate a decline in
295 proportion of Lineage A (higher logit transformed values) as site soil % LOI increases from
296 0-20%, thereafter remaining constant. The model fits for pH indicated an initial decline in the
297 proportion of Lineage A individuals (higher logit transformed values) as pH increases from
298 4.5 to 5, with, thereafter, an increase in frequency (lower logit transformed values) where
299 site pH increases from 5 to 7.5 (Fig. 4 a,b). Amongst other measured variables, only log soil
300 Ca concentrations (-0.487, P=0.012) and the average annual number of rain days (0.433,
301 p=0.027) were also significantly correlated with Linage A proportion. Both of these variables
302 are, however, also significantly correlated with soil pH (Annual rain days: -0.584, p=0.002;
303 log soil Ca concentration:-0.751, p<0.001) making precise attribution of cause challenging.
304
305 Separate univariate models were generated to analyse tissue metal concentrations in
306 relation to collection site and lineage. The collection site had a highly significant (p<0.001)
307 influence on tissue concentrations for all analysed trace elements. This is only to be
308 expected, given that the sites include locations with no history of local pollution, to highly

309 contaminated industrial and mine sites. Lineage was also a significant factor in the model for
310 As ($p<0.02$). This difference is, however, based on only a relatively small difference in
311 average tissue concentrations between lineages across all sites. Thus, average
312 concentrations in Linage A of 10.69 (n=300) was slightly lower than the average tissue
313 arsenic concentrations of 11.7 mg/kg (n=195) for Lineage B, Hence although statistically
314 significant, the absolute magnitude of difference in tissue As concentrations between
315 lineages is small. For all other analysed metals, there was no significant effect of lineage on
316 tissue concentration ($p>0.05$).

317 *4. DISCUSSION*

318 Species distributions can be affected by a range of environmental drivers, including
319 physiological tolerances, dispersal constraints, biotic interactions and anthropogenic
320 influences (Dennis and Hellberg, 2010; Gaston, 2003). Among earthworms, species show
321 preference for certain habitats, for example common compost earthworm species such as
322 *Eisenia fetida*, *Perionyx excavatus* and *Eudrilus eugeniae* preferentially occupy organic
323 matter rich habitats associated with animal manure or composting vegetation (Edwards,
324 2004). Further, some species also have preference for different soil physiochemical
325 properties. For example, Jaensch et al (2013) found differences in morphospecies
326 preference across different soil pH classes, with species such as *Allolobophora chlorotica*,
327 *Aporrectodea rosea*, *Aporrectodea longa* and *Lumbricus terrestris* preferring soils with pH
328 >5.6, and *Dendrodrilus rubida*, *Dendrobaena octaedra* and *L. rubellus* soils with pH <5.6.
329
330 The UK earthworm fauna is notably denuded, comprising only around 20-25 native species,
331 compared to about 180 species that are found in neighboring France (Bouché, 1972; Sims
332 and Gerard, 1985). The reduced earthworm fauna of the UK can be linked to its recent
333 history of glaciation and the severing of the land bridge to Europe that restricted earthworm
334 colonization after glacial retreat. This influence of quaternary glaciation is consistent with
335 what is known about the current distribution and genetic structure of a range of species
336 across Europe and the UK (Hewitt, 2000). Among UK earthworm species, the majority show
337 a widespread and cosmopolitan distribution (Boag et al., 1997; Carpenter et al., 2012;
338 Rutgers et al., 2016; Sims and Gerard, 1985). Habitat preferences are known, such as those
339 for pH and for organic rich habitats as discussed previously, however the spatial
340 heterogeneity of terrestrial habitats means that at coarse recording scales (e.g. 10 km^2 or
341 even 1 km^2), a significant proportion of UK earthworm species may be present in any given
342 sampling area (e.g. a mixed land-use area subject to comprehensive earthworm sampling
343 within different vegetation stands and habitats).
344

345 Genetic marker studies have identified deeply divergent cryptic lineages within many
346 common UK earthworm morphospecies based on mitochondrial or nuclear genetic marker
347 analysis. An active debate currently surrounds the question of whether these cryptic
348 lineages correspond to cryptic species or highly polymorphic species variants (Blakemore et
349 al., 2010; Giska et al., 2015; King et al., 2008). In the specific case of *L. rubellus*, the
350 presence of cryptic lineages is established from studies conducted from measurement of
351 highly divergent (13-15%) sequences for both of the cytochrome oxidase I and II
352 mitochondrial genes (Andre et al., 2010; Donnelly et al., 2014; Kille et al., 2013). Pan-
353 European studies have shown that at continental scale, morphotype *L. rubellus* may
354 comprise of 5 or more such deeply divergent lineages (Giska et al., 2015), two of which were
355 here found across sites in England and Wales (Fig 1). Recently RADseq analysis suggests
356 that cryptic *L. rubellus* lineages may represent a case of a highly polymorphic single species
357 rather than true cryptic species (Giska et al., 2015). Nonetheless, previous studies of the
358 lineage physiology have identified a number of differential responses between lineages (as
359 previously outlined notably for the two UK lineages). For example, Jones et al. (2016) found
360 that the two lineage were favorably attracted to soils that had previously been worked by
361 earthworm of their own rather than the alternative lineage. These results suggests that
362 pheromone attractants may allow mate selection in mix populations,. such as those that are
363 found at the majority of our sampled site. Such selection has the potential to underpin
364 lineage differences in habitat preference and, as a consequence, different spatial
365 distributions at local scale.
366
367 Earthworms are key ecosystem engineers for the role that play an important role in the
368 creating of the spatial structure and chemistry of the soil habitat through bioturbation, litter
369 degradation and nutrient cycling (Edwards, 2004; Lavelle et al., 1997; Liebeke et al., 2015).
370 The extent to which the divergent lineages of common earthworm species overlap in respect
371 of habitat preference will be an important determinant of morphospecies contributions to
372 different ecosystem processes across space and time. The analysis here suggests that, in

373 the case of the two UK lineages of *L. rubellus*, there are ecological drivers of distribution.
374 Individually, soil pH and % OM were both significant correlated with the proportions of *L.*
375 *rubellus* Lineage A (and conversely Lineage B) collected across the 26 sample sites. These
376 two measurement parameters were selected for particular focus because they are
377 recognized as important environmental drivers of the distribution of a number of soil taxa
378 (Cassagne et al., 2003; Griffiths et al., 2011; Raty and Huhta, 2003). Additionally, there are
379 also correlations with other climate and soil variable that are themselves know to influence
380 soil pH through soil geochemistry and leaching.
381
382 Different soil pH preferences have direct effects on earthworm traits including reproduction,
383 growth and survival (Baker and Whitby, 2003; Spurgeon et al., 2006; Van Gestel et al.,
384 1992). *L. rubellus* is tolerant of relatively low pH, being commonly (and even preferentially)
385 found in moderately acidic soils (Jaensch et al., 2013). Results here suggest that this
386 cosmopolitan nature could partly arise from different lineage pH preferences, with Lineage B
387 found in more acid habitats from pH 4.5 to 5.5 and Lineage A preferentially in nearer neutral
388 pHs of 5.5 and above. Thus, within the current study, Lineage B was absent from 6 of 26
389 sampled sites, while Lineage A was found at all except one of the sampled sites. The
390 detailed genetics of the two cryptic lineages may provide some clues to the basis of such
391 differences. Studies of mitochondrial and genetic marker genes have established that
392 Lineage B has lower genetic diversity of measured traits than Lineage A (Donnelly et al.,
393 2014; Kille et al., 2013). This suggests that Lineage B may have undergone a population
394 bottleneck that restricted the genetic diversity, and possibly, the colonization capacity of this
395 lineage.
396
397 The strongest correlate of lineage frequencies was soil organic matter (% loss on ignition).
398 The fresh and partially degraded soil organic component provides earthworms with food. It
399 is, therefore, possible that this association is driven by different dietary requirements of the
400 two lineages, as has been recognized for different earthworm species (Piearce, 1978).

401 However, in addition to acting as food, soil organic matter also contributes to soil structure
402 and moisture retention. Earthworms are known to be sensitive to soil texture, with regional
403 studies linking species distributions to soil sand, clay and organic matter content (Joschko et
404 al., 2006; Salome et al., 2011). Soils lacking in organic matter are also vulnerable to
405 prolonged periods of high soil moisture deficit. This can be challenging for earthworms given
406 their critical need to retain water balance. The significant correlation with site average rain
407 days also points to a possible influence of soil hydrology on distributions. Metabolomic
408 analyses have identified that many earthworm species contain a high number of betaines
409 which likely act as osmolytes that help to retain soil water balance (Liebeke and Bundy,
410 2013). Any differences in the extent of such protection between lineages may influence
411 colonization ability for more drought susceptible soils.
412
413 Although there is correlation of lineage frequency with both soil pH and soil organic matter,
414 the fact that these two soil variables are co-correlated to other environmental variables
415 makes it hard to unequivocally assign them as the major drivers of lineage distribution. For
416 example, high organic matter/low pH soils are more common in the West of England and
417 Wales than in the East. This geographic relationship could potentially be associated with
418 different recolonization histories for the two lineages, e.g. perhaps recolonization from
419 different glacial refugia (Hewitt, 2000). However, as there is no significant correlation of
420 Easting to lineage proportion, this seems less likely than direct effects of soil organic
421 matter/pH. Ultimately, to tease apart the drivers of lineage preference, higher resolution
422 collection and mapping and experimental manipulation of habitats would be required.
423
424 Differences in physiology that separate species in relation to habitat preference could also
425 affect the way that the two lineages handle and accumulate different trace elements. For the
426 site-level analysis, the soil concentrations of major pollutant metals were correlated with
427 PC2, which was not associated with lineage. For the individual analysis of tissue metals,
428 arsenic was the only one found to vary with lineage (significantly higher in Lineage B

429 individuals). Previous work has indicated that the two lineages differ in the genetic
430 mechanisms underlying the development of arsenic tolerance. Analysis of amplified
431 fragment length polymorphisms indicated that Lineage A showed differences in patterns of
432 nuclear markers indicating genetic tolerance, while Lineage B showed a difference in DNA
433 methylation patterning, but not genetic differences (Kille et al., 2013). The observed
434 difference here in As accumulation between lineages across sites suggests that these
435 genetic differences lead to phenotypic differences in the handling of As.
436
437 *5. CONCLUSIONS*
438 Earthworms represent 'super-sentinels' exploited for environmental monitoring and
439 ecotoxicology, as well as being keystone soil engineers essential for soil quality. The
440 identification of possible drivers of species and lineage distributions has potential
441 implications for their use in environmental assessment as well as in studies of ecosystem
442 service delivery. For example, when assessing biodiversity effects of pollution and land-use
443 change it may be valuable to consider the occurrence of different lineages to understand
444 how populations may adapt to change through changes in lineage frequency. This analysis
445 may be required because the two widespread cryptic lineages of *L. rubellus* differ in their
446 habitat preferences with frequencies changing as conditions change. Given that bacterial
447 communities are also known to differ in relation to soil pH, then difference in the nature and
448 strengths of earthworm and microbial interactions can be expected between lineages. These
449 relationships between soil macrofauna and microbes are key to soil carbon turnover, nutrient
450 cycling and soil structural characteristics and this aspect warrants further investigation.
451 Earthworms are also valuable for metal biomonitoring. Our results suggest that the lineages
452 behave identically with respect to metal bioaccumulation, with the exception of As. Thus,
453 selection of morphotype *L. rubellus* will provide a coherent picture of metal accumulation
454 independent of lineage, unless As is a specific focus of any assessment.

455 *6. ACKNOWLEDGMENTS*

456 This study was supported by the Natural Environment Research Council (NERC), UK, under
457 grant number NE/H00973/1. We thank Dr Rachael Madison and Dr Judith Garforth for help
458 with earthworm and soil collection and the Forestry Commission, ADAS and Countryside
459 Commission for Wales for allowing access to Environmental Change Network (ECN) sites
460 and Alice Holt, Snowdon and Drayton Experimental Frame respectively.

461 *7. REFERENCES*

462 Anderson, C.J., Kille, P., Lawlor, A.J., Spurgeon, D.J., 2013. Life-history effects of arsenic
463 toxicity in clades of the earthworm *Lumbricus rubellus*. Environ Pollut 172, 200-207.
464 Andre, J., King, R.A., Stürzenbaum, S.R., Kille, P., Hodson, M.E., Morgan, A.J., 2010.
465 Molecular genetic differentiation in earthworms inhabiting a heterogeneous Pb-polluted
466 landscape. Environ Pollut 158, 883-890.
467 Arnold, B.E., Hodson, M.E., Charnock, J., Peijnenburg, W., 2008. Comparison of subcellular
468 partitioning, distribution, and internal speciation of Cu between Cu-tolerant and naive
469 populations of *Dendrodrilus rubidus* Savigny. Environ Sci Technol 42, 3900-3905.
470 Baker, G.H., Whitby, W.A., 2003. Soil pH preferences and the influences of soil type and
471 temperature on the survival and growth of *Aporrectodea longa* (Lumbricidae). Pedobiologia
472 47, 745-753.
473 Bardgett, R.D., van der Putten, W.H., 2014. Belowground biodiversity and ecosystem
474 functioning. Nature 515, 505-511.
475 Bartlett, M.D., Briones, M.J.I., Neilson, R., Schmidt, O., Spurgeon, D., Creamer, R.E., 2010.
476 A critical review of current methods in earthworm ecology: From individuals to populations.
477 European Journal of Soil Biology 46, 67-73.
478 Blakemore, R.J., Kupriyanova, E.K., Grygier, M.J., 2010. Neotypification of *Drawida*
479 *hattamimizu* Hatai, 1930 (Annelida, Oligochaeta, Megadrili, Moniligastridae) as a model
480 linking mtDNA (COI) sequences to an earthworm type, with a response to the 'Can of
481 Worms' theory of cryptic species. Zookeys, 1-29.
482 Blouin, M., Hodson, M.E., Delgado, E.A., Baker, G., Brussaard, L., Butt, K.R., Dai, J.,
483 Dendooven, L., Peres, G., Tondoh, J.E., Cluzeau, D., Brun, J.J., 2013. A review of
484 earthworm impact on soil function and ecosystem services. Eur. J. Soil Sci. 64, 161-182.

485 Boag, B., Palmer, L.F., Neilson, R., Legg, R., Chambers, S.J., 1997. Distribution, prevalence
486 and intensity of earthworm populations in arable land and grassland in Scotland. Ann. Appl.
487 Biol. 130, 153-165.
488 Bouché, M.B., 1972. Lombriciens de france. Ecologie et systématique. Institut national de la
489 Recherche Agronomique, Paris.
490 Bundy, J.G., Keun, H., Sidhu, J.K., Spurgeon, D.J., Svendsen, C., Kille, P., Morgan, A.J.,
491 2007. Metabolic profile biomarkers of metal contamination in a sentinel terrestrial species are
492 applicable across multiple sites. Environ Sci Technol 41, 4458-4464.
493 Carpenter, D., Sherlock, E., Jones, D.T., Chiminoides, J., Writer, T., Neilson, R., Boag, B.,
494 Keith, A.M., Eggleton, P., 2012. Mapping of earthworm distribution for the British Isles and
495 Eire highlights the under-recording of an ecologically important group. Biodivers. Conserv.
496 21, 475-485.
497 Cassagne, N., Gers, C., Gauquelin, T., 2003. Relationships between Collembola, soil
498 chemistry and humus types in forest stands (France). Biol. Fertil. Soils 37, 355-361.
499 Cluzeau, D., Guernion, M., Chaussod, R., Martin-Laurent, F., Villenave, C., Cortet, J., Ruiz-
500 Camacho, N., Pernin, C., Mateille, T., Philippot, L., Bellido, A., Rouge, L., Arrouays, D.,
501 Bispo, A., Peres, G., 2012. Integration of biodiversity in soil quality monitoring: Baselines
502 for microbial and soil fauna parameters for different land-use types. European Journal of Soil
503 Biology 49, 63-72.
504 Davies, B.E., 1989. Data handling and pattern recognition for contaminated soils. Environ.
505 Geochem. Health 11, 137-143.
506 Dennis, A.B., Hellberg, M.E., 2010. Ecological partitioning among parapatric cryptic species.
507 Mol. Ecol. 19, 3206-3225.

508 Dominguez, J., Aira, M., Breinholt, J.W., Stojanovic, M., James, S.W., Perez-Losada, M.,
509 2015. Underground evolution: New roots for the old tree of lumbricid earthworms. Mol.
510 Phylogenet. Evol. 83, 7-19.
511 Donnelly, R.K., Harper, G.L., Morgan, A.J., Pinto-Juma, G.A., Bruford, M.W., 2014.
512 Mitochondrial DNA and morphological variation in the sentinel earthworm species
513 *Lumbricus rubellus*. European Journal of Soil Biology 64, 23-29.
514 Edwards, C.A., 2004. Earthworm Ecology, 2 ed. CRC Press, Boca Raton, Florida, USA, p.
515 441.
516 Emerson, B.C., Cicconardi, F., Fanciulli, P.P., Shaw, P.J.A., 2011. Phylogeny,
517 phylogeography, phylobetadiversity and the molecular analysis of biological communities.
518 Philosophical Transactions of the Royal Society B-Biological Sciences 366, 2391-2402.
519 Emmett, B.A., Reynolds, B., Chamberlain, P.M., Rowe, E., Spurgeon, D., Brittain, S.A.,
520 Frogbrook, Z., Hughes, S., Lawlor, A.J., Poskitt, J., Potter, E., Robinson, D.A., Scott, A.,
521 Wood, C., Woods, C., 2010. Countryside Survey: Soils Report from 2007. NERC Centre for
522 Ecology and Hydrology, Wallingford, UK.
523 Fitter, A.H., Gilligan, C.A., Hollingworth, K., Kleczkowski, A., Twyman, R.M., Pitchford,
524 J.W., Programme, N.S.B., 2005. Biodiversity and ecosystem function in soil. Funct. Ecol. 19,
525 369-377.
526 Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for
527 amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan
528 invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294-299.
529 Gaston, K.J., 2003. The Structure and Dynamics of Geographic Ranges. Oxford University
530 Presss, Oxford, UK.
531 Giller, P.S., 1996. The diversity of soil communities, the 'poor man's tropical rainforest'.
532 Biodivers. Conserv. 5, 135-168.

Giska, I., Sechi, P., Babik, W., 2015. Deeply divergent sympatric mitochondrial lineages of the earthworm *Lumbricus rubellus* are not reproductively isolated. BMC Evol. Biol. 15.

Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M., Whiteley, A.S., 2011. The bacterial biogeography of British soils. Environ. Microbiol. 13, 1642-1654.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907-913.

International Organisation for Standards, 2005. Determination of pH., Geneva, Switzerland.

Jaensch, S., Steffens, L., Hoefer, H., Horak, F., Ross-Nickoll, M., Russell, D., Toschki, A., Roembke, J., 2013. State of knowledge of earthworm communities in German soils as a basis for biological soil quality assessment. Soil Organisms 85, 215-233.

James, S.W., Porco, D., Decaens, T., Richard, B., Rougerie, R., Erseus, C.A.F.N.J.S.W., Porco, D., Decaens, T., Richard, B., Rougerie, R., Erseus, C., 2010. DNA Barcoding Reveals Cryptic Diversity in *Lumbricus terrestris* L., 1758 (Clitellata): Resurrection of *L. herculeus* (Savigny, 1826). PLoS One 5, e15629.

Jones, G.L., Wills, A., Morgan, A.J., Thiomas, R.J., Kille, P., Novo, M., 2016. The worm has turned: Behavioural drivers of reproductive isolation between cryptic lineages. Soil Biol. Biochem. 98, 11-17.

Joschko, M., Fox, C.A., Lentzsch, P., Kiesel, J., Hierold, W., Kruck, S., Timmer, J., 2006. Spatial analysis of earthworm biodiversity at the regional scale. Agric. Ecosys. Environ. 112, 367-380.

Kille, P., Andre, J., Anderson, C., Ang, H.N., Bruford, M.W., Bundy, J.G., Donnelly, R., Hodson, M.E., Juma, G., Lahive, E., Morgan, A.J., Sturzenbaum, S.R., Spurgeon, D.J., 2013. DNA sequence variation and methylation in an arsenic tolerant earthworm population. Soil Biol. Biochem. 57, 524-532.

King, R.A., Tibble, A.L., Symondson, W.O.C., 2008. Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. Mol. Ecol. 17, 4684-4698.

Klarica, J., Kloss-Brandstatter, A., Traugott, M., Juen, A., 2012. Comparing four mitochondrial genes in earthworms - Implications for identification, phylogenetics, and discovery of cryptic species. Soil Biol. Biochem. 45, 23-30.

Lavelle, P., Bignell, D., Lepage, M., Wolters, V., Roger, P., Ineson, P., Heal, O.W., Dhillon, S., 1997. Soil function in a changing world: the role of invertebrate ecosystem engineers. European Journal of Soil Biology 33, 159-193.

Leveque, T., Capowiez, Y., Schreck, E., Mombo, S., Mazzia, C., Foucault, Y., Dumat, C., 2015. Effects of historic metal(loid) pollution on earthworm communities. Sci Total Environ 511, 738-746.

Liebeke, M., Bruford, M.W., Donnelly, R.K., Ebbels, T.M.D., Hao, J., Kille, P., Lahive, E., Madison, R.M., Morgan, A.J., Pinto-Juma, G.A., Spurgeon, D.J., Svendsen, C., Bundy, J.G., 2014. Identifying biochemical phenotypic differences between cryptic species. Biol. Lett. 10.

Liebeke, M., Bundy, J.G., 2013. Biochemical diversity of betaines in earthworms. Biochem. Biophys. Res. Commun. 430, 1306-1311.

Liebeke, M., Strittmatter, N., Fearn, S., Morgan, A.J., Kille, P., Fuchser, J., Wallis, D., Palchykov, V., Robertson, J., Lahive, E., Spurgeon, D.J., McPhail, D., Takats, Z., Bundy, J.G., 2015. Unique metabolites protect earthworms against plant polyphenols. Nature Comms. 6.

Novo, M., Cunha, L., Maceda-Veiga, A., Talavera, J.A., Hodson, M.E., Spurgeon, D., Bruford, M.W., Morgan, A.J., Kille, P., 2015. Multiple introductions and environmental factors affecting the establishment of invasive species on a volcanic island. Soil Biol. Biochem. 85, 89-100.

581 Pass, D.A., Morgan, A.J., Read, D.S., Field, D., Weightman, A.J., Kille, P., 2015. The effect
582 of anthropogenic arsenic contamination on the earthworm microbiome. Environ. Microbiol.
583 17, 1884-1896.
584 PerezLosada, M., Ricoy, M., Marshall, J.C., Dominguez, J., 2009. Phylogenetic assessment
585 of the earthworm Aporrectodea caliginosa species complex (Oligochaeta: Lumbricidae)
586 based on mitochondrial and nuclear DNA sequences. Mol. Phylogenet. Evol. 52, 293-302.
587 Piearce, T.G., 1978. Gut contents of some Lumbricid earthworms. Pedobiologia 18, 153-157.
588 Raty, M., Huhta, V., 2003. Earthworms and pH affect communities of nematodes and
589 enchytraeids in forest soil. Biol. Fertil. Soils 38, 52-58.
590 Römbke, J., Aira, M., Backeljau, T., Breugelmans, K., Domínguez, J., Funke, E., Graf, N.,
591 Hajibabaei, M., Pérez-Losada, M., Porto, P.G., Schmelz, R.M., Vierna, J., Vizcaíno, A.,
592 Pfenninger, M., 2016. DNA barcoding of earthworms (*Eisenia fetida*/*andrei* complex) from
593 28 ecotoxicological test laboratories. Appl. Soil Ecol. doi:10.1016/j.apsoil.2015.02.010.
594 Rowell, D.L., 1994. Soil Science: Methods and Applications. Longman Scientific and
595 Technical, Harlow, UK.
596 Rutgers, M., Orgiazzi, A., Gardi, C., Roembke, J., Jaensch, S., Keith, A.M., Neilson, R.,
597 Boag, B., Schmidt, O., Murchie, A.K., Blackshaw, R.P., Peres, G., Cluzeau, D., Guernion,
598 M., Briones, M.J.I., Rodeiro, J., Pineiro, R., Diaz Cosin, D.J., Paulo Sousa, J., Suhadolc, M.,
599 Kos, I., Krogh, P.-H., Faber, J.H., Mulder, C., Bogte, J.J., van Wijnen, H.J., Schouten, A.J.,
600 de Zwart, D., 2016. Mapping earthworm communities in Europe. Appl. Soil Ecol. 97, 98-
601 111.
602 Rutgers, M., Schouten, A.J., Bloem, J., vanEekeren, N., deGoede, R.G.M., Akkerhuis, G.,
603 vanderWal, A., Mulder, C., Brussaard, L., Breure, A.M., 2009. Biological measurements in a
604 nationwide soil monitoring network. Eur. J. Soil Sci. 60, 820-832.

605 Sackett, T.E., Smith, S.M., Basiliko, N., 2013. Indirect and direct effects of exotic
606 earthworms on soil nutrient and carbon pools in North American temperate forests. Soil Biol.
607 Biochem. 57, 459-467.
608 Salome, C., Guenat, C., Bullinger-Weber, G., Gobat, J.M., Le Bayon, R.C., 2011. Earthworm
609 communities in alluvial forests: Influence of altitude, vegetation stages and soil parameters.
610 Pedobiologia 54, S89-S98.
611 Sims, R.W., Gerard, B.M., 1985. Earthworms. Linnean Society and the Estuarine and
612 Brackish-Water Sciences Association, London.
613 Spurgeon, D.J., Lofts, S., Hankard, P.K., Toal, M., McLellan, D., Fishwick, S., Svendsen, C.,
614 2006. Effect of pH on metal speciation and resulting metal uptake and toxicity for
615 earthworms. Environ Toxicol Chem 25, 788-796.
616 Spurgeon, D.J., Rowland, P., Ainsworth, G., Rothery, P., Long, S., Black, H.I.J., 2008.
617 Geographical and pedological drivers of distribution and risks to soil fauna of seven metals
618 (Cd, Cu, Cr, Ni, Pb, V and Zn) in British soils. Environ Pollut 153, 273-283.
619 Umarov, M.M., Striganova, B.R., Kostin, N.V., 2008. Specific features of nitrogen
620 transformation in the gut and coprolites of earthworms. Biol. Bulletin 35, 643-652.
621 Van Gestel, C.A.M., Dirven-Van Breemen, E.M., Baerselman, R., 1992. Influence of
622 environmental conditions on the growth and reproduction of the earthworm *Eisenia andrei* in
623 an artificial soil substrate. Pedobiologia 36, 109-120.
624 Warton, D.I., Hui, F.K.C., 2011. The arcsine is asinine: the analysis of proportions in
625 ecology. Ecology 92, 3-10.

626

627 LEGENDS TO FIGURES

628

629 Figure 1. Location of collection sites and the proportion of Lineage A (dark blue shading) and
630 Lineage B (light yellow shading) *L. rubellus* based on the total number of collected and
631 assigned genotyped individual (given in brackets) for the 26 sites visit over four separate
632 collection campaigns

633

634 Figure 2. Principal component analysis results show the ordination of site geographical,
635 climatic, biotic and soil chemical variables of sample sites showing the major related site
636 characteristic variables.

637

638 Figure 3. Boxplots showing median (centre line), upper and lower quartile (box limits) and
639 upper and 95% confidence intervals (whiskers) of trace metal concentrations measured
640 across 17 samples site for assigned Lineage A and Lineage B *L. rubellus*.

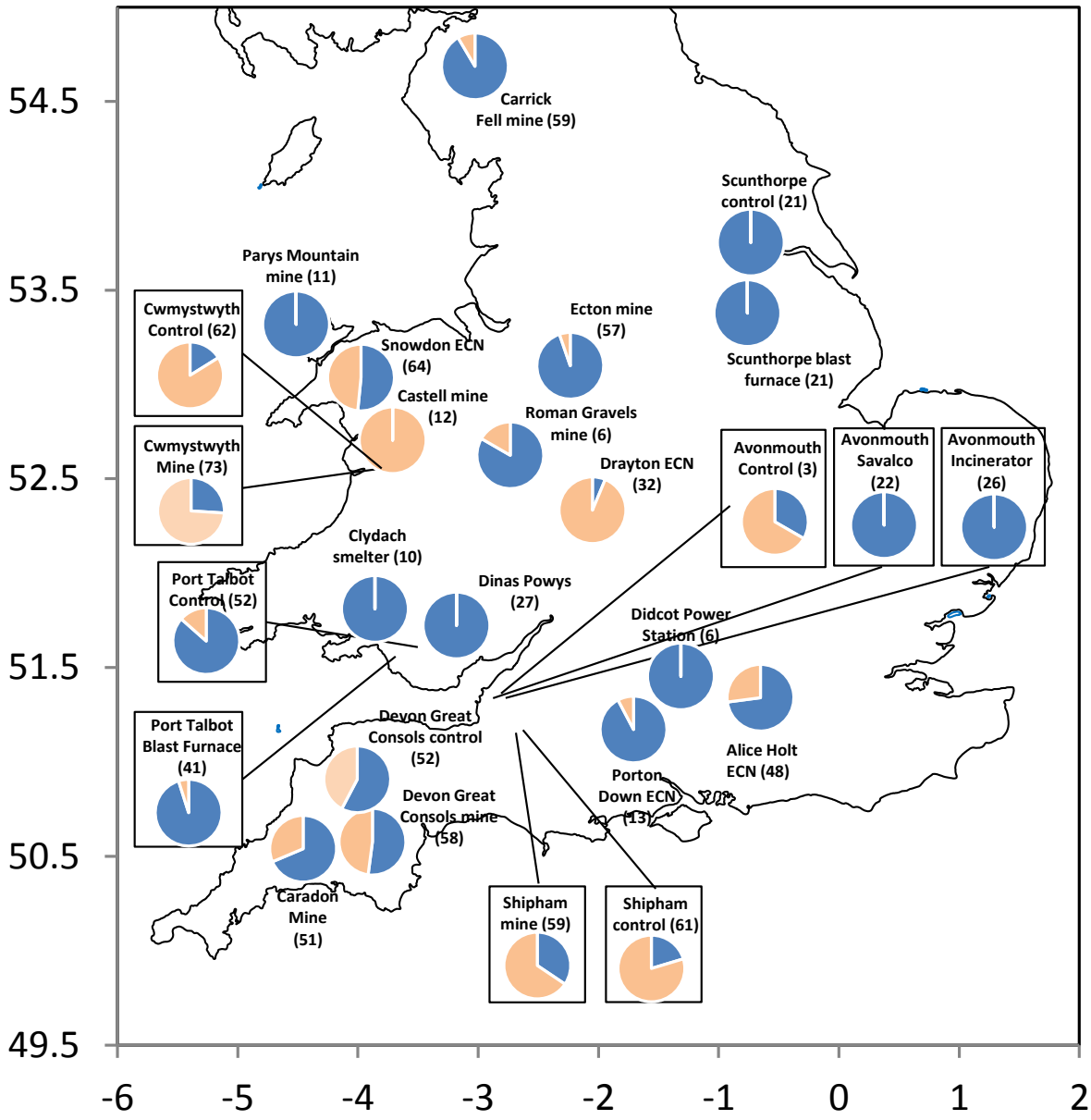
641

642 Figure 4. Scatterplots with fitted locally weighted scatterplot smoother line of proportion of
643 Lineage A *L. rubellus* in relation to (a) Soil % OM and (b) soil pH.

644

645 FIG. 1

646



647

FIG. 2

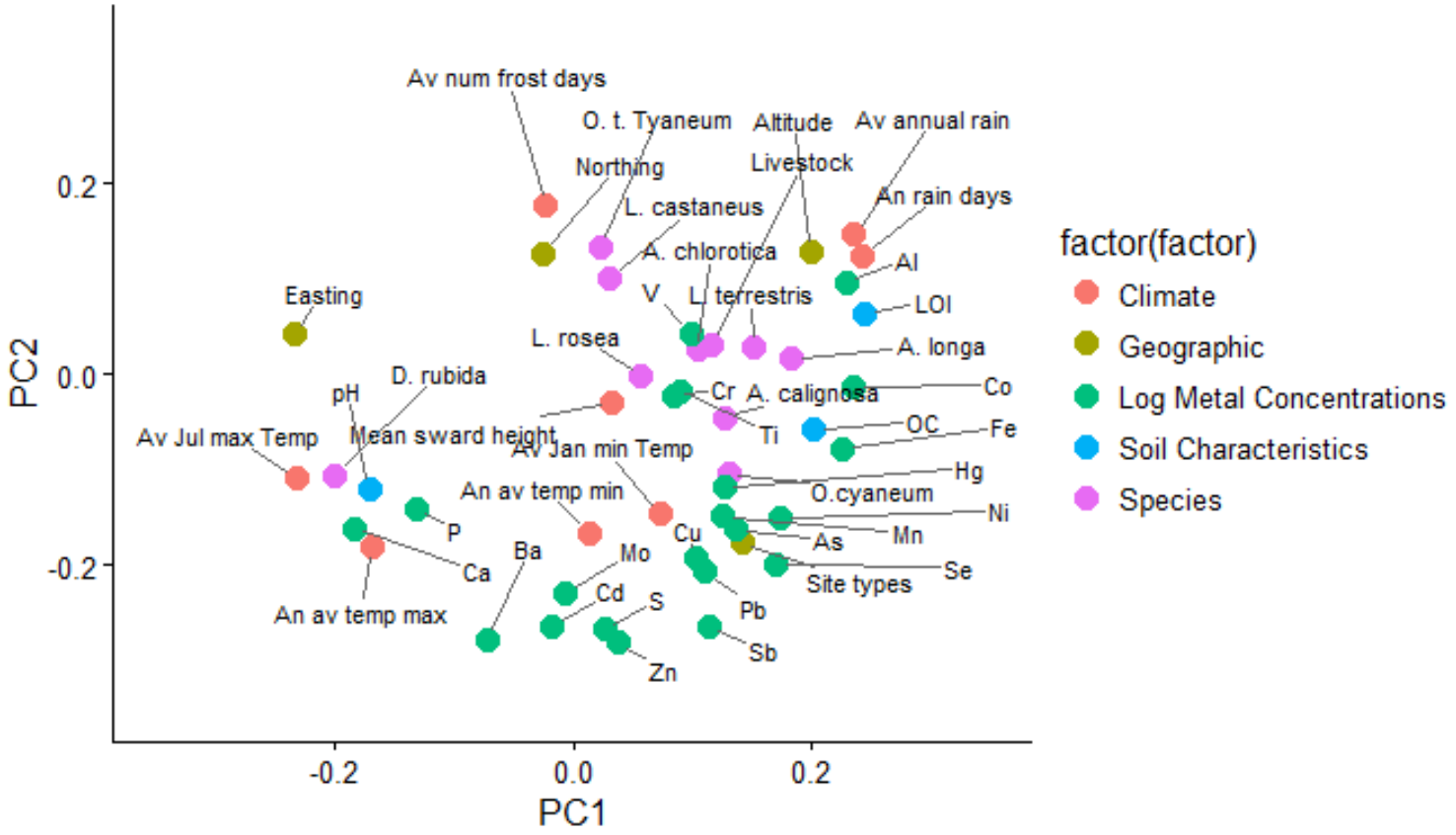

Av num frost days
O. t. Tyaneum
Altitude
Av annual rain
Northing
Livestock
An rain days
L. castaneus
A. chlorotica
Al
Easting
V
L. terrestris
LOI
L. rosea
A. longa
D. rubida
pH
Cr
A. calignosa
Co
Av Jul max Temp
Mean sward height
Av Jan min Temp
Ti
OC
Fe
Hg
An av temp min
O.cyaneum
Ni
P
Cu
As
Mn
Ba
Mo
Ca
Site types
Se
Cd
S
An av temp max
Pb
Sb
Zn
0.2
0.0
-0.2
PC2
PC1
factor(factor)
Climate
Geographic
Log Metal Concentrations
Soil Characteristics
Species

FIG. 3

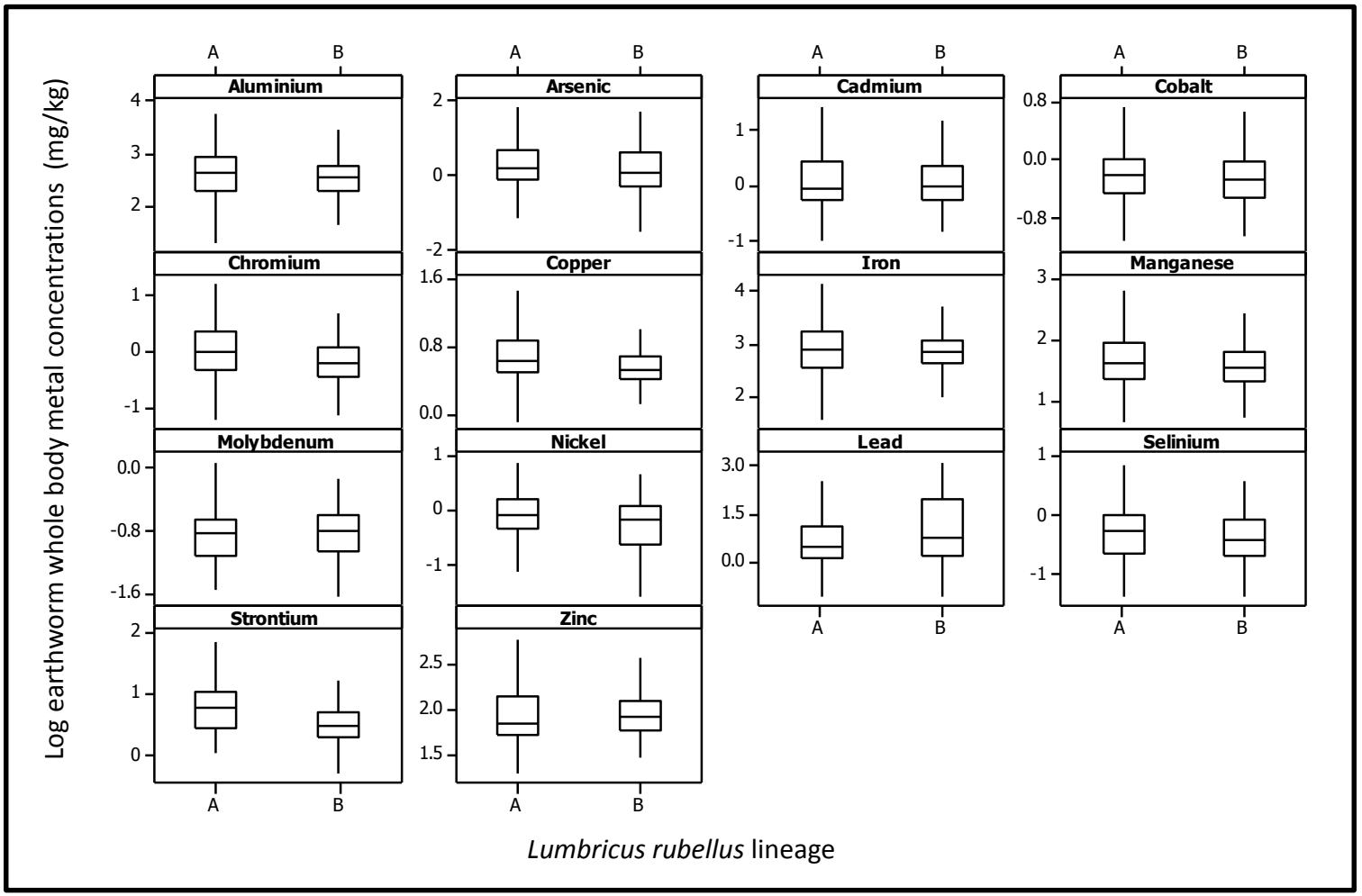
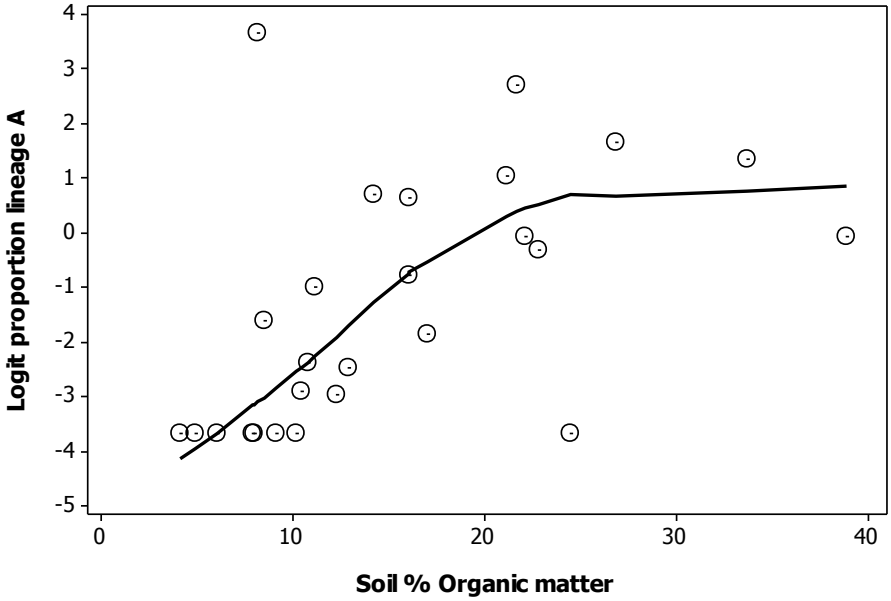
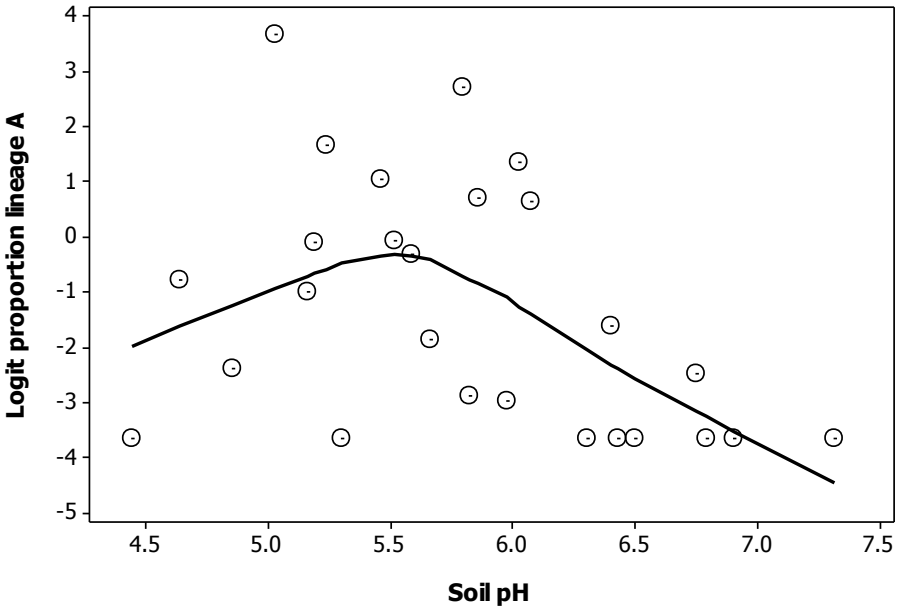

Log earthworm whole body metal concentrations (mg/kg)
Aluminium
Arsenic
Cadmium
Cobalt
Chromium
Copper
Iron
Manganese
Molybdenum
Nickel
Lead
Selinium
Strontium
Zinc
A
B
Lumbricus rubellus lineage

FIG. 4


Logit proportion lineage A
4
3
2
1
0
-1
-2
-3
-4
-5
0
10
20
30
40
Soil % Organic matter


Logit proportion lineage A
4
3
2
1
0
-1
-2
-3
-4
-5
4.5
5.0
5.5
6.0
6.5
7.0
7.5
Soil pH

SUPPLEMENTARY TABLES

Supplementary Table 1. Geographical locations and reported climatic conditions of the 23 sites used for the collection of morphotype *L. rubellus*.

Site name	Site types	Land use	Ordnance Survey Grid reference	Easting	Northing	Altitude	Annual Average temp Max	Annual average temp min	Average Jan Min Temp	Average Jul Max Temp	Average Annual Rain	Annual rain days	Average No Forst Days
Alice Holt ECN Control	Unpolluted	Broadleaf woodland	SU 80060 39821	480060	139821	88.6	14.1	6.4	1.6	21.9	755	121	45.9
Avonmouth Control	Unpolluted	Improved pasture	ST 57006 82149	357006	182149	7	14.2	7	2.2	21.5	802	126	34.9
Avonmouth Incinerator	Industrial polluted	Rough grassland	ST 54099 81659	354099	181652	6	14.2	7	2.2	21.5	802	126	34.9
Avonmouth Savalco	Industrial polluted	Rough grassland	ST 53859 79411	353859	179411	6	14.2	7	2.2	21.5	802	126	34.9
Caradon Mine	Mining polluted	Rough grassland	SX 25624 69792	225624	69792	226	13.2	7	3	19.1	1385	172	30.6
Carrick Fell Mine	Mining polluted	Rough grassland	NY 32211 32982	332211	532982	661	13	5.8	1.6	19.7	1521	176	56.5
Castell Mine	Mining polluted	Rough grassland	SN 77415 81254	277415	281254	297	11.9	5.2	1	18.2	186	191	58.4
Clydach Smelter	Industrial polluted	Broadleaf woodland	SN 69587 01409	269587	201409	25	13.5	8.5	4	19.6	999	148	9.7
Cwmystwyth control	Unpolluted	Rough grassland	SN 79598 74222	279598	274222	198	11.9	5.2	1	18.2	1856	191	58.4
Cwmystwyth mine	Mining polluted	Rough grassland	SN 80852 75166	280852	275166	177	11.9	5.2	1	18.2	1856	191	58.4
Devon Great Consouls Control	Unpolluted	Improved pasture	SX 42560 74019	242560	74019	133	14	8.1	4	19.9	1007.4	142	16.3
Devon Great Consouls Mine	Mining polluted	Broadleaf woodland	SX 42385 73152	242385	73152	133	14	8.1	4	19.9	1007.4	142	16.3
Didcot Power Station	Industrial polluted	Broadleaf woodland	SU 51645 91402	451645	191402	53	14.4	5.9	1.2	22.6	661	112	57.7
Drayton ECN Control	Unpolluted	Improved pasture	SP 16391 55061	416391	255061	66	14.5	5.9	1.3	22.8	614	114	52.2
Dinas Powys	Unpolluted	Broadleaf woodland	ST 15868 70431	315868	170431	57	14.7	7	2.3	21.7	1151.9	149	35.7
Ecton Mine	Mining polluted	Broadleaf woodland	SK 09698 58263	409698	358263	103	13.9	6	1.2	22.1	598	112	49.1
Parys Mountain	Mining polluted	Rough grassland	SH 43829 89971	243829	389971	117	13.2	7.7	3.6	18.8	841	143	20.3
Port Talbot Control	Unpolluted	Rough grassland	SS 83690 84574	283690	184574	150	13.5	8.5	4	19.6	999	148	9.7
Port Talbot Blast Furnace	Industrial polluted	Rough grassland	SS 79001 85463	279001	185463	6.1	13.5	8.5	4	19.6	999	148	9.7
Porton Down ECN	Unpolluted	Improved pasture	SU 19575 37692	419575	137692	102	14.1	6.2	1.4	21.9	749	122	47.6
Roman Gravels Mine	Mining polluted	Rough grassland	SJ 33592 00339	333592	300339	368	14.1	5.6	1.3	21.6	668	126	51.8
Scunthorpe blast furnace	Industrial polluted	Arable	SE 94800 15840	494800	415835	19.7	13.4	5.7	0.9	21.3	613	115	49.8
Scunthorpe Control	Unpolluted	Arable	SE 93156 12000	493156	412000	10.7	13.4	5.7	0.9	21.3	613	115	49.8
Shipham control	Unpolluted	Improved pasture	ST 46312 59409	346312	159409	54	14.6	7	2.6	21.7	899	134	28.9
Shipham mine	Mining polluted	Improved pasture	ST 44799 57273	344799	157273	169	14.6	7	2.6	21.7	899	134	28.9
Snowdown ECN control	Unpolluted	Rough grassland	SH 63674 55116	263674	355116	748	12	5.9	1.8	18.1	2612	199	50.1

Supplementary Table 2. Vegetation and presence (shaded) or absence (unshaded) for earthworm species at the 23 sites used for the collection of morphotype *L. rubellus*.

Site name	Mean sward height	Livestock	Exposure	*Aporrectodea caliginosa*	*Aporrectodea rosea*	*Lumbricus castaneus*	*Dendrobaena rubida*	*Aporrectoda longa*	*Lumbricus terrestris*	*Lumbricus festivus*	*Octolasion cyaneum,*	*Allolobophora chlorotica*	*Octolasion tyrtaeum tyaneum*
Alice Holt ECN Control	-	Deer	shaded	■				■	■		■		
Avonmouth Control	20 cm	None	part-shaded	■				■				■	■
Avonmouth Incinerator	10 cm	None	open	■		■		■					■
Avonmouth Savalco	35 cm	None	open			■			■				
Caradon Mine	5 cm	horses	open										
Carrick Fell Mine	10 cm	sheep	open						■				
Castell Mine	5 cm	None	open			■				■			
Clydach Smelter	5 cm	None	shaded	■	■	■	■				■		
Cwmystwyth control	10 cm	None	open				■						
Cwmystwyth mine	10 cm	sheep	open										
Devon Great Consouls Control	10 cm	None	open	■	■	■	■		■			■	
Devon Great Consouls Mine	10 cm	None	part-shaded				■						
Didcot Power Station	-	None	Shaded	■		■		■	■		■	■	■
Drayton ECN Control	15 cm	None	open	■	■	■		■	■		■		
Dinas Powys	-	None	shaded	■	■	■	■		■			■	
Ecton Mine	-	None	part-shaded						■				
Parys Mountain	5 cm	None	part-shaded	■									
Port Talbot Control	10 cm	None	part-shaded	■	■	■		■	■			■	
Port Talbot Blast Furnace	10 cm	horses	open	■		■		■	■			■	
Porton Down ECN	8 cm	None	part_shaded										
Roman Gravels Mine	3 cm	horses	open										
Scunthorpe blast furnace	5 cm	None	open	■	■	■		■	■			■	
Scunthorpe Control	6 cm	None	part-shaded	■		■		■	■			■	
Shipham control	10 cm	None	open	■					■		■	■	
Shipham mine	10 cm	cattle	part-shaded			■			■				
Snowdown ECN control	15 cm	sheep	open	■		■	■						

Supplementary Table 3. Arithmetic mean of measured soil chemical properties for pH, loss on ignition and concentrations of a suite of trace lements based on analysis of three samples collected from sites the 23 sites used for the collection of morphotype *L. rubellus*.

Site name	Soil pH	% Soil loss on ignition	Al (mg/kg)	As (mg/kg)	Ba (mg/kg)	Cd (mg/kg)	Co (mg/kg)	Cr (mg/kg)	Cu (mg/kg)	Fe (mg/kg)	Hg (mg/kg)	Mn (mg/kg)	Mo (mg/kg)	Ni (mg/kg)	Pb (mg/kg)	Sb (mg/kg)	Se (mg/kg)	Ti (mg/kg)	V (mg/kg)	Zn (mg/kg)	Ca (mg/kg)	P (mg/kg)	S (mg/kg)
Alice Holt ECN Control	5.16	11.15	8737	13.2	19.3	0.1	6	12.6	12.4	16500	0.17	78	0.3	9.5	27.1	0.4	0.5	7	24.6	43	1640	424	202
Avonmouth Control	5.86	14.23	11130	20.9	356	2.3	10	46.3	75.1	26133	0.76	539	1.2	25.1	207	1.3	1.7	32.9	26.9	697	68633	524	1029
Avonmouth Incinerator	6.5	7.94	11000	36.9	439	50.6	12.3	23.6	204.3	26133	1.07	1010	2	25.7	1943	14.9	5.6	34.4	31.8	4640	39267	1600	3190
Avonmouth Savalco	6.3	24.51	14733	7.8	75.1	2.6	8.1	25.1	29.8	19233	0.25	663	0.6	18.4	99.6	1.2	0.6	95.5	29.6	299	4777	1233	734
Caradon Mine	4.64	16	5147	407	33.7	1.17	2.03	3.9	609	20633	0.33	221	2.57	4.27	69.1	3.17	1.49	27.4	13.8	43.5	342	776	884
Carrick Fell Mine	4.85	10.8	16300	737	33.5	3.07	8.18	16.4	59	33300	2.06	837	18.10	11.4	173	7.01	3.85	165	92.6	282	1607	961	1457
Castell Mine	5.03	8.15	14333	14.7	15.7	9.03	12.5	16.4	60	46267	0.5	844	1.16	16.5	210	2.28	1.35	8.1	26.8	1792	321	506	312
Clydach Smelter	5.3	10.2	6780	44.1	142	1.61	90.1	32.3	465	25067	0.34	1369	2.68	1799	278	3.57	14.43	70.9	47.2	370	39277	945	1230
Cwmystwyth control	5.23	26.89	16167	19.8	14.6	0.1	6.7	18.7	19.7	38033	0.12	467	0.9	14.3	626	1.6	1.2	6.6	24.6	116	209	688	709
Cwmystwyth mine	5.46	21.13	20033	49	29.8	0.2	41.2	24.1	28.8	50433	112	2597	1.5	23.3	657	2.5	1.4	22	33.3	127	551	689	534
Devon Great Consols Control	5.58	22.79	21533	310	45.5	0.4	14.8	31.7	107.3	45800	0.05	585	1	27.5	68	1.4	2.3	41.8	38.7	140	2213	1177	919
Devon Great Consols Mine	5.19	38.9	17300	6270	45.9	0.2	25.7	17.8	2647	79600	0.64	630	1.3	22.3	225	13.9	2.6	38.4	38.5	277	3340	286	1503
Didcot Power Station	7.31	4.17	10570	14.5	139	0.53	8.89	14.7	41	22900	0.34	309	0.93	24.1	102	2.42	1.40	15.1	26.9	181.00	37600	901	1250
Drayton ECN Control	5.79	21.65	12833	25.1	48.4	0.4	4.9	15.8	46	28533	0.1	805	2.3	10.7	30.2	0.6	0.5	34.3	32	176	35000	2000	1183
Dinas Powys	6.9	8	17166	23	85.4	1.2	16.1	39	42.6	24933	0.1	1120	1.7	45.1	109	*	1.2	41.5	45.7	770	12766	2070	*
Ecton Mine	5.82	10.4	1403	136	537	61.4	33.2	5.4	5787	12267	0.35	965	101	72.4	1553	92.1	5.42	20.1	20.2	6047	116200	203	9280
Parys Mountain	4.44	4.95	681	1480	96	8.77	7.78	5	3673	175667	3.34	35.7	37.5	3.17	29033	210	42	54.4	19	2333	1749	26.7	22467
Port Talbot Control	5.66	16.99	15300	14.5	127	0.6	11.3	37.7	33.8	31467	0.27	1287	0.9	17.4	39.1	0.8	1.2	15.7	62.9	480	13903	1427	1083
Port Talbot Blast Furnace	5.98	12.3	10323	14.5	287	0.8	4.7	129	35.5	34967	0.33	3550	1.7	16.8	117	2.2	1.1	230.7	140	341	89133	1227	1680
Porton Down ECN	6.75	12.93	6153	21	88.3	0.7	8	20	39.2	13567	0.37	611	0.9	17.3	109	1.4	0.5	46.3	17.2	184	148333	3100	1423
Roman Gravels Mine	6.4	8.48	10633	13.7	123	14.2	14.3	13.3	99	30133	0.33	582	1.17	24.8	1125	2.48	1.33	15.7	18.4	1788	6367	455	878
Scunthorpe blast furnace	6.43	9.14	10803	40.6	90.3	0.3	13.1	42	25.4	59900	0.33	1227	1.5	29.5	124	1.8	1	68.3	128.7	183	13500	1497	453
Scunthorpe Control	6.79	6.04	7197	24.6	47.4	0.4	4.9	15.4	20.2	21767	0.12	535	2.3	11.2	30.0	0.6	0.5	33.6	30.7	69.1	130667	1253	1167
Shipham control	6.07	16.1	12967	37.1	302	2.1	7.5	23.1	19.1	26500	0.03	465	0.7	16.4	163	0.9	0.4	53.9	34.8	328	3520	972	584
Shipham mine	6.03	33.7	9907	867	1526	404	12.5	21	85.1	85100	12.3	2277	5.1	31.6	7260	48.8	3.8	51.4	28.5	31833	16117	2370	3547
Snowdown ECN control	5.52	22.07	38600	17.5	10.1	0.3	30.6	78.2	19.5	70733	0.19	1350	0.5	30.4	37	0.5	1.4	1530	239	114	1111	524	658