Mechanistic bases of metal tolerance: linking phenotype to genotype.

Craig Anderson

2012

PhD Thesis

Cardiff University

School of Biological Sciences

Abstract

Ecotoxicology is currently undergoing a revolution as the result of new technological advances in molecular biology, capable of finely resolving metabolic mechanisms associated with exposure. These high-throughput analyses can detail the evolutionary and ecological implications of exposure in non-model organisms, such as the earthworm, *Lumbricus rubellus*. This terrestrial sentinel has been observed across former mine sites that are highly contaminated with arsenic and have been found to mitigate toxicity at soil concentrations that cause mortality in unadapted populations. This is indicative of the adaptive capacity of natural populations recently exposed to persistent and strong selection pressure. However, mechanisms surrounding adaptation to arsenic in *L. rubellus* have yet to be characterised, and so the effects of exposure are broadly reported with the aim of distinguishing resistance from phenotypic plasticity in natural populations.

Unadapted earthworms were initially used to derive basal phenotypic variation associated with arsenic exposure. Variation in life-history parameters was observed among adult and juvenile *L. rubellus*, establishing relative sensitivity and population-level inferences. A systems biology approach was employed to describe molecular mechanisms associated with arsenic metabolism, encompassing transcriptomic and metabolomic analyses, underpinned by arsenic speciation.

Insight into the genetic bases of arsenic resistance, which enable persistence of *L. rubellus* at highly contaminated sites, was sought. Recombinant inbred lineages derived from adapted populations, were cultivated and their phenotypes relative to arsenic exposure determined. Phylogeographic analyses were used to interrogate genetic variation among populations inhabiting former mine sites as well as proximal control sites. A mitochondrial marker defined cryptic species across the UK, but did not establish soil chemical profiles relative to clade occurrence. RADseq better resolved genetic variation at these sites, determining that soil geochemistry is strongly associated with genetic variation. Furthermore, genomic markers inferred genetic erosion, found to selectively reduce variation at sites relative to a single clade.

Acknowledgements

Primarily, thanks tremendously to my supervisors Dave, Claus and Pete for donating to me generous portions of their time, effort and talent. I wouldn't have gotten here without their guidance and enthusiasm.

Thanks so much to my collaborators for all of their help and infinite patience- Mark Hodson, Ben Elsworth, Jake Bundy, Manuel Liebeke, Karim Gharbi, John Davey and Pablo Orozco ter Wengel.

A massive thanks to Dan Pass and Pier Francesco Sechi, who have been both great friends and colleagues.

Thanks to all the great students at CEH, including Rory O'Connor, Fingers Phil, Leah Cuthbertson, Christine Tansey and especially Fraser Sinclair.

Thanks to the Danish contingent at CEH and Copenhagen- Nils, Hanna, Lisbeth and Laura.

Cheers to Chris Goulebourne for all of the seriously good times and posh meals.

Thanks to the guys from Plymouth- Matt, Alex, Raul. Maybe Lenny.

Big thanks to Harry, Mel, Stu, Yaser, Luis and Ceri for time well spent in Cardiff.

I'm also extremely grateful to all the fantastic support I received at Cardiff, including Steve Turner, Esh, Julian, Mike O'Reily, Joan, Carsten and Mike Bruford.

Thanks to the Oxford skaters for all of the amazing times and trips- Sam, Tom, Hugo Boss, Louie, Doddy, Leddy and JK.

Thanks to Aaron and James for everything.

Cheers to John Morgan, for showing everyone how it's done.

Thanks to my parents and sister for taking an interest, supporting me and putting up with me as I wrote up.

Thanks to my wonderful girlfriend Sarah, though this is probably worse because of her...

Finally, I'd like to thank Tom and Jenny. I absolutely wouldn't be where I am without their friendship, generosity and company. I am forever indebted.

Contents

Chapter 1

Introduction

1.1 General Introduction

Organisms can maintain fitness by responding to environmental variation through phenotypic regulation. The genetic framework that supports the dynamism of this system is the product of multi-generational selection. This is the nature of adaptation, in that it permits survival and reproduction of the individual, influencing variation that is found within populations (Morgan et al. 2007). The adaptive response is therefore a portrait of evolutionary history that declares the range and impact of environmental variation experienced by the collective biological systems of a population (Meyers and Bull 2002). As populations are put under selective pressure, more favourable genetic information that is conducive to survival under contemporary specific environmental scenarios will be retained. If the disturbance is too great, reduced genetic diversity will result due to directional selection, constraining the ability of populations to respond appropriately to future perturbations (van Straalen and Timmermans 2002).

Adaptive mechanisms permissive to the management of environmental variation appear to be dependent upon the complexity of the exposed system (van Straalen and Timmermans 2002) as well as the temporal occurrence of exposure within specific life-histories (Doroszuk et al. 2006). The more powerful and frequently occurring environmental variation is, the more character defining and better integrated the response will tend to be. It is likely that more highly developed taxa tend to exist within an environment prior to reproductive activity for longer, and therefore are more likely to experience environmental fluctuation as a result of individual behaviour or the temporal nature of the disturbance, i.e. diurnal and seasonal climate fluctuations, allowing for a temporary adaptive response.

The ability to maintain a stable phenotype as the result of increased system complexity is dependent upon the molecular mechanisms set in place to maintain and modify such a system, and their regulation relative to the environmental stimulation has a temporal significance. Acute exposure results in variation at the molecular level, whereas prolonged, multi-generational exposure results in genetic pruning and exaggeration that transforms and defines populations (Tully and Ferriere 2008). These different responses reflect the means by which populations meet short and long-term

challenges, and in the case of selection can potentially establish the stability of contemporary populations on a generational basis within the limits of imposed abiotic changes. Our ability to achieve such an insight is dependent largely on the resolution within biological systems gained through technological advances as well as the more recent call to integrate different fields on a massive level through what is now termed "Systems Biology". That is, the pooling of data from across the "-omics" spectrum to demonstrate the adaptive capacity of the system as a whole, focussing upon responses expressed under specific scenarios.

Adaptation to environmental perturbation occurs through wide-ranging, integrated mechanisms that are effective relative to the temporal nature of exposure. Short-term response is elicited under phenotypic plasticity in conjunction with canalization (that is, as a means of maintaining robustness (fitness) under fluctuating environmental parameters) via maternal effects and the response to environmental stimuli, which are immediate but temporary. Semi-permanent epigenetic alterations also occur and allow for "memorization" of reactions to perturbations and can lead to permanent genetic basis of adaptation (Walsh et al. 2010), which predominantly occurs when a long-term, multi generational selective pressure is imposed upon populations. Genetic variation can lead to changes among adaptive mechanisms that ultimately reflect the characteristics of the organism and the journey taken to achieve such phenotypes. Through resolution of the particular mechanisms under which an organism responds to environmental stimuli, one may gain insight into adaptive mechanisms.

1.2 Phenotypic Plasticity

The phenotype is the expression of the genotype in response to environmental and developmental cues. Adaptive "phenotypic plasticity" represents an acclimation in response to environmental changes from a single genotype (Agrawal 2001). It is one of the fundamental mechanisms that endow an organism with the ability to mitigate deleterious effects of heterogeneous or unpredictable environments through alterations in the phenotype that maintains fitness. This is dependent upon the organism's ability to sense variation within the environment and transduction of this signal towards responsive transcriptional machinery (Roelofs et al. 2010). In relation to this, Schlichting and Smith (2002) proposed the hierarchical levels of phenotypic plasticity, specifically highlighting molecular machinery such as the transcriptome and the proteome as tools for coordinating responses to environmental stimuli. Changes in gene and protein expression can lead to physiological or behavioural changes in the organism as a direct response to stress. This includes the possibility to modify

developmental trajectories, and also the ability of an individual organism to change its phenotypic state or activity (e.g. via metabolism) (Fusco et al., 2010).

Sensory perception of environmental stimuli has a key role in eliciting a responsive phenotype, reflecting conditions under which ancestral populations developed. The ability of a diverse array of organisms to react to their environments is well documented (Derby and Sorensen 2008) and a range of stimuli relative to environment, feeding, predation, mating, etc., have been observed to elicit a profound response, which are beginning to be defined in relation to plasticity (Reale et al. 2007; Sih et al. 2004). Behaviour is considered to be a highly plastic phenotype that is likely to allow for an almost reflexive response to environmental stimulation (Briffa et al. 2008), as a result of a direct interface of regulatory mechanisms with specific cues (Hazlett 1995). For example, *Daphnia pulex* respond behaviourally by moving through the water column relative to the position of predation threats (Nesbitt et al. 1996). Environmental stimuli can also provoke variation at the molecular level, which can mitigate perturbation directly or affect mechanisms further up the hierarchy, towards protein formation, physiology and also behaviour given the permanency and strength of stimuli. Such responses may be observed during induction of seasonal phenotypic variation, such as that among avians (Pulido 2007).

Drosophila melanogaster used to assess phenotypic plasticity of the transcriptome was able to provide a preliminary demonstration of gene expression across 4 fitness-related phenotypes following the effects of 20 different environmental variations (Zhou et al. 2012). A primary set of environmentally responsive, genetically variable signatures of plasticity were noted, accounting for 15% of gene expression. The data suggests that a specific proportion of the transcriptome tightly controls discrete phenotypes expressed following environmental disruption. Nussey et al. (2007) suggested that these types of variation represent “labile” phenotypes that can be repetitively expressed, enabling organisms to enrol the necessary biochemical or physiological adjustments throughout life-histories to ensure survival and reproduction (Morgan et al. 2007). This type of phenotypic control is differentiated from management of developmental trajectory as a response to specific environmental stimuli that reflects a distinctive, “non-labile” adaptation that occurs only once during an organism’s life. For example, the nematode, *C. elegans*, alters both larval and reproductive development by entering into the dauer stage, which is a developmentally arrested larval phenotype that has reduced sensitivity to environmental variation as a result of detecting high density populations or shortages in food (Harvey et al. 2008). When nematodes expressing arrested phenotypes encounter more favourable conditions, they can benefit from leaving the

dauer stage and progressing towards their development into an adult. This signifies the role of development upon movement into specific phenotypes (the phenotypic landscape), which is theorised to allow for future movement into others that are more appropriate and plays a role in the development and evolution of adaptive phenotypes (Alberch 1991).

Phenotypic plasticity is able to confer advantages when entering new environments, such as in *Daphnia lumholtzi*, which form a defensive head spine in response to chemical cues from predators (Engel and Tollrian 2009). This adaptation has become the dominant phenotype among invasive populations that occur in North America and is considered to play a key role in the range expansion of this species and its competitive advantage over natives. Such phenotypic plasticity is considered to be favoured over genetic variation for achieving an effective phenotypic response to environmental variation, as niche accessibility is improved (de Jong 2005). This in part is a result of canalisation, which is the stability of phenotypes despite environmental variation. This is reflected in the occurrence of alternative, discrete phenotypes that are limited in occurrence as opposed to wide-ranging, continuous differentiation, reinforced by the work of Zhou et al. (2012). Thus, the same phenotype may occur following a range of different environmental stimuli as organisms can respond to manage perturbation, but maintain a level of stability so as to maintain fitness optima. In this way, canalisation acts to constrain selection on novel phenotypes (Flatt 2005). Phenotypic plasticity therefore permits the occurrence of organisms within a range of environmental conditions, preventing speciation through elimination of stringent population subdivision that may occur (Moczek 2010).This is reliant upon the means by which individuals are able to perceive and respond to environmental variation in a temporary (i.e. non-transgenerational) manner, but is fundamentally based upon the history of a species and its ancestors. There are, however, exceptations as to the reach of phenotypic variation beyond parental lineages.

1.3 Maternal effects

Juveniles have been found to be more sensitive to environmental perturbation in a number of cases (Adam et al. 2010; Qiu and Qian 1998). This is potentially because they are often smaller in size, have fewer biological resources for managing environmental perturbation and are less robust. For example, they may often have thinner epidermises and a greater surface-area-to volume ratios relative to adults, which may be important in the context of dermal exposure and resulting toxicity. In many invertebrates, juveniles develop within an egg, characterised by the maternal

condition under which it was developed, containing substances derived directly from the mother following stimulation by environmental factors (Ho and Burggren 2010). On occasion, juvenile fitness is improved through maternally derived adaptive responses that mange resources relative to deleterious circumstances. In doing so, these act to buffer their impact upon offspring (Burgess and Marshall 2011). These “maternal effects” are exemplified in an investigation by Mitchell and Read (2005), who demonstrated the direct repercussions of maternal exposure to parasites in *Daphnia magna*, whose offspring were found to have reduced susceptibility to bacterial infection.

Subsequent to birth and hatching, parents can also affect offspring indirectly (Marshall and Uller 2007), through allocation of fewer resources in an attempt to maintain personal fitness when stressed. For instance, offspring size (Marshall and Keough 2007) and by proxy, egg size (Bernardo 1996), are key indicators that reflect transgenerational adaptive strategies. These life-stages, along with the number of offspring produced, crucially reflect how energy is invested following disturbance. The interplay between life-history stages can be important drivers of variation in population growth rate. Analysis of such effects using population life-history based models can provide an insight into the mechanisms deployed for managing acute exposure. Commonly, it appears that organisms will forgo temporary reductions in fecundity to maximise fitness throughout the disturbance. For example, exposure of earthworms to c60 nanoparticles elicited demographic variation, including reduced cocoon production, which is attributed to a diversion of resources towards increased maintenance and repair (van der Ploeg et al. 2011). Deviation in resource allocation may also be evident when analysing growth response of juveniles, which may mitigate toxicity using that which would otherwise be intended for biomass production (Kammenga et al., 2003).

1.4 Epigenetics

In a similar, and in some cases overlapping manner to maternal effects, epigenetic variation can be inherited directly from parents. Alternatively epigenetic changes can result as a consequence of within generation exposure to environmental stimuli. In both cases the presence of epigenetic modification to DNA chemistry or morphology can affect gene expression profiles. The persistent nature of many types of epigenetic modification provides affects on the phenotype of offspring that can be semi-permanent in nature. These changes entailing epigenetic “marks” can occur relative to underlying genetics without any actual deviation in genetic sequence. Types of recognised epigenetic marks include molecular modifications to DNA that include cytosine

methylation and histone modification, regarded as key mechanisms often associated with transcriptional variation (Schones and Zhao 2008). A role for various types of RNA has also been observed, whereby transcripts can affect both initiation and maintenance of epigenetic marks (Shirayama et al. 2012; Zhai et al. 2008). However, new epigenetic mechanisms continue to be observed, including hydroxymethylation (Ficz et al. 2011), adenine methylation (Ficz et al. 2011) and regulation of the transcriptome via mRNA methylation (Meyer et al. 2012), reflecting that this field is still in its infancy.

Eukaryotes maintain a regulatory system to ensure that genetic material is highly condensed as chromatin, consisting of tightly wound nucleosomes that incorporate 147 bp of material (Luger et al. 1997). At the core of each unit is an octomer of histone proteins that mediate access of the genetic material by the transcriptional machinery; two H3-H4 dimmers form a centralised histone tetramer, flanked by 2 H2A-H2B dimmers that contact DNA at 14 points (Li et al. 2007; Luger et al. 1997). Such tightly controlled packaging of genetic material means that it is available for transcription when the correct developmental or environmental stimuli call for it. The histone-DNA interface is mediated by chaperones and remodelling complexes, permitting access to transcriptional machinery. RNA polymerase II, along with transcriptional elongation factors temporarily displaces a single H2A-H2B dimer, disentangling the sequence and permitting transcript elongation. As this machinery moves along the sequence, the octamer reforms. When transcription (and therefore RNA polymerase occupation) is high, the entire octamer can be reversibly displaced to facilitate the needs of the cell (Kulaeva et al. 2007). This process allows for the implementation of histone variants such that form less stable dimers and are more resistant to modification, which better facilitate frequent transcription (Schwartz and Ahmad 2005; Workman 2006). Such epigenetic modifications are thought to be exchanged transgenerationally through the germ line (Jablonka and Raz 2009), while there is also evidence of variation in histone modifications reflecting deviation relative to environmental variation in model organisms (Hall et al. 2011; Ho and Burggren 2010; Seong et al. 2011).

DNA methylation is another well studied epigenetic mark, debatably better so than histone modifications as methodological progress has exploited recent technological advances in DNA sequencing. This specific epigenetic mark entails methylation of cyotsines by methyltransferases and is often more stable than histone modification, potentially lasting throughout the lifetime of the organism (Waterland et al. 2010). DNA methylation of gene promoters and/or coding regions is considered to inhibit transcript elongation and prevents aberrant transcriptional initiation in susceptible genes (Mandrioli 2007) that also endows a mechanism for silencing genetic parasites (Regev

et al. 1998), while within vertebrates, it is generally considered that methylation has deviated through limitation to promoters for involvement of gene silencing. However, DNA methylation is considered to be ancient and represents differing mechanisms for specific effects in different groups (Zemach et al. 2010), varying in location and frequency, as well as in maintenance and turnover.

Transgenerational epigenetic inheritance is differentiated from maternal effects through functioning primarily via variation in germ cells, however, distinguishing between the two remains difficult. For example, agouti mice are well acknowledged to demonstrate heritable epigenetic variation via the maternal germ line in respect to fur colour, whereby methylation of a transposable element upstream of the agouti locus results in darkening of the coat, from yellow to black (Morgan et al. 1999). However, evidence shows that maternal effects may not be discounted. This conclusion is derived from embryo transfer and breeding experiments, which reflect that although the intrauterine environment is not responsible for maternal effects, the cytoplasm of oocytes may play a role in phenotypic variation (Morgan et al. 1999). The agouti phenotype is also interesting because of sensitivity to environmental deviation, whereby mice supplemented with a diet rich in methyl groups were less likely to produce yellow pups (Wolff et al. 1998). Environmental stress that may result in randomly formed epigentically differentiated alleles that may undergo selection, potentially leading to rapid local adaptation (Finnegan 2002). Thus, transgenerationally heritable epigenetic variation can provide an indication of environmental variation that imposes selective pressure upon organism and is therefore, justifiably, gaining momentum in environmental fields such as ecotoxicology (Head et al. 2012).

Epigenetic variation remains relatively poorly understood and is expensive to deploy, therefore studies examining environmental effects outside of model organisms are infrequent. Although DNA methylation is recognised now within a wide-range of organisms (Zemach et al., 2010) and studies can benefit from advances in sequencing technology, only a small number of studies have surveyed epigenetic variation at suitable resolution, remaining as yet unfounded in natural populations of metazoans. Walsh et al. (2010) documented methylation in the pea aphid (*Acyrthosiphon pisum*), finding homologues for DNA methyltransferases found in vertebrates. Further, the majority of methylation was observed at exons, with further evidence of methylation among genes responsible for endocrine regulation. Epigenetic variation relative to environmental perturbation is likely to be a fundamental adaptive response recognised among environmental sentinel organisms such as *Daphnia magna* (Vandegehuchte et al. 2009; Vandegehuchte et al. 2010) and among earthworm species (Maldonado

Santoyo et al. 2011), which have seen focussed genome sequencing efforts. The foundations of such research are likely to be relative to relevant environmental variables with a mechanistic role in epigenetic modification, such as arsenic, which is thought to incorporate methyl groups during metabolism that are also used by DNA methyltransferases. Furthermore, epigenetic modifications can lead to genetic variation, such as through demethylation of methylated cytosines, giving uracil that becomes thymidine following repair (Walsh et al. 2010). These mutations signify a more permanent means by which environmental variation can impact populations, potentially eliciting an adaptive response that can transform the ability to maintain fitness under perturbation.

1.5 Resistance

Environmental deviation can lead to selection for phenotypes that offer increased fitness. Although this is true of the other mechanisms for adaptation already mentioned, it differs in that it reflects only a proportion of the population who are able to persevere via traits that mitigate disturbance through genetic differentiation from those unable to. Ultimately, this results in a shift in population structure as demonstrated among a culture of *D. melanogaster* that were maintained in the dark for 57 years (1400 generations)(Izutsu et al. 2012). This “dark fly” population were found to have higher fecundity under the conditions that they had been selected for and were found to vary genetically from the strain from which they were derived. Whole genome sequencing was used to determine a range of candidate genes that were differentiated in the dark fly culture, including one for response to light and another for olfactory reception, reflecting the nature by which this strain is now genetically adapted. Long-term selection experiments such as that by Izutsu et al. (2012) provide a powerful means by which the genome-wide effects of a specific adaptation can be controlled and recognised. Predominantly, these make use of the genetic resources and relatively short generation times of model organisms, with studies recently being reviewed by Burke (2012).

Although a genetic basis for adaptive variation (or “resistance”) can be lost in the same way it is gained, it is ultimately considered to be permanent. Genetic bases for adaptation are particularly important as they can most definitively infer selection events, and provide insight into the ecology of adaptation. The extent of the genetic variation required to address specific selection pressure provides insight into not only the extent of exposure, but also the mechanisms involved and commonality of such adaptations. These may therefore provide insight into the occurrence of fundamental

evolutionary phenomena such as convergent evolution. Of course, this will be based upon the mechanisms indicted through the environmental perturbation.

Evolutionary theory suggests that adaptive traits may arise from either standing genetic variation or from single new mutations and are termed as “hard” and “soft” selective sweeps, respectively (Hermisson and Pennings 2005). Hard sweeps are limited by the rates and consequences of new mutations and have a strong effect on all ancestral neutral variation that is tightly linked to the selected allele through elimination by hitchhiking, i.e. loci linked to selected traits and genome wide effects (Barton 2000). Such an event will mean that adaptations will spread quickly to fixation through natural selection (Pritchard et al. 2010). This is similar to that seen in Andersen et al. (2012), who determined that a series of strong selective sweeps have resulted in low levels of genetic variation in world-wide sampling of *Caenorhabditis elegans*. One sweep, specifically, was considered to have occurred within the last few hundred years and is thought to be a result of human movement. Soft sweeps reflect when a trait is selected for in an allele that is already segregating within the population, but has previously assumed to be neutral or mildly deleterious under normal circumstances. Soft sweeps are limited by the quality and amount of available genetic variation and will therefore take longer to become fixed in populations (Hermisson and Pennings 2005). This is endorsed under a different scenario whereby multiple independent mutations that each confer similar adaptive advantages occur over a single locus occur. Evidence for this has been observed during insecticide resistance in *D. melanogaster*, in which simple and complex resistance evolved quickly and repeatedly within individual populations (Karasov et al. 2010).

The nature of the adaptation required to deal with perturbations will ultimately have the greatest influence on the rate at which populations can achieve physiological optimization, the frequency of individuals able to do so and the inherent population dynamics. It is also important to acknowledge that life-history parameters of a population provides the context specific scope for evolutionary capacity, i.e. fixation of adaptive traits will favour species with short generation times and high reproductive capacities. The role of the adaptation within the system and its stability (considered important for normal development and physiological function) and complexity (i.e. pathways involved as opposed to change in binding of transcribed proteins) can thus inform you of how well conserved the adaptive event is. Pritchard et al. (2010) describe another model that takes into account the effect of adaptive events at the level of the genome. Polygenic adaptation is put forward as an alternative to hard sweeps (though not mutually exclusive) that allows for rapid adaptation but is derived from standing

genetic variation at many loci simultaneously. The authors exemplify human stature as a well-studied polygenic trait, with around 50 candidate loci affecting human height in Europeans. If such a trait comes under selection, it is likely that allele frequency shifts at those loci will occur and although such adaptation will rise in frequency quickly, it will generally not come to fixation in the population as a result of the combinational minor roles played by each. This presses a need to establish the transcriptional variation seen in adapted individuals, as this may affect the means by which genetic networks function (i.e. via epistasis (Gao et al. 2010)).

Strongly deleterious environmental variation can instigate a reduction in not only population size through mortality or reduced reproductive capacity, but also in the genetic variation of populations (Hoffmann and Willi 2008). This loss in contributing members is termed as a population bottleneck and is eloquently described and supported by Ramakrishnan et al. (2005), as to "increase the rate of random genetic drift (Wright 1931) and increase inbreeding (e.g. Saccheri et al. 1999) leading to the loss of genetic variation (e.g. as documented in Hoelzel et al. 1993; Taylor et al. 1994) which may lead to...reduced ability to respond to environmental change (Lacy 1997; Reed et al. 2002). Bottlenecks are also important given their potential role in speciation (Mayr 1963; Slatkin 1996) and founding events (Barton and Charlesworth 1984)".
In addition to the influence of bottlenecks, populations also undergo depressions in genetic variation as a result of mutation rate, directional selection and migration rates as suggested by van Straalen and Timmermans (2002). These authors critically assessed attempts made to explore the occurrence of this "genetic erosion" among natural populations exposed to stressors derived from anthropogenic sources. Genetic variation, as suggested previously, facilitates adaptation to environmental perturbation and therefore reduces the likelihood that members of a population that has undergone a strong selective event will be able to adapt to future events via genetic erosion. Localised selection events may in fact bring infrequent genetic variation to the forefront within adapted populations and if interaction with other populations is maintained, novel environments could actually be a source of variation (Shirley and Sibly 1999), which may be acknowledged through recognition of compatibility and the frequency that closely related populations or species are able to mix (Dasmahapatra et al. 2012).

In summation, there are a range of adaptive mechanisms that permit an organism to persist through environmental perturbation, which is strongly reliant upon the extent of the selective pressure imposed. Adaptation ranges from tolerance at a level of phenotypic plasticity, to intergenerational molecular variation that confers a medium-term memory to benefit development via maternal effects and epigenetic variations,

through to stringent selection whereby genetically differentiated individuals that have higher fitness shape populations. This not only acknowledges some of the major signatures and effects of adaptation, but more roundly demonstrates why adaptation, especially that which is associated with anthropogenic sources of environmental variation, is important to acknowledge: Recognising adaptation to sources of environmental variation informs us of how we indirectly interact with natural populations, how intensely we affect their natural history and the potential life-history and evolutionary outcomes as a result of this interaction. As such the mechanism that govern species adaptation of environmental stress are an important topic for scientific investigation in support of understanding current and future impacts of environmental perturbation upon natural communities.

1.6 Investigating adaptation in natural populations

One of the most appropriate areas for investigating recent adaptation occurs relative to environmental pollutants. In particular, metal contamination is one of the best means of recognising adaptation as it reflects a permanent (non-degraditive) and often strong source of selective pressure that can be imposed upon natural populations over numerous generations. A handful of articles have deterministically demonstrated a genetic basis for resistance in field sampled populations, with arguably the most well documented case of resistance recorded in populations of the springtail, *Orchesella cincta*. It was found that individuals originating from contaminated sites and cultured in uncontaminated laboratory conditions had a higher, multigenerational cadmium binding and excretion efficiency than naïve populations (Posthuma et al. 1993). This was the result of variation seen in a metallothionein promoter, which was responsible for increased transcription observed among adapted springtails (Janssens et al. 2007). Other studies that have shown multigenerational resistance to metal exposure have done so regarding the effects of cadmium on the isopod *Porcellio scaber* (Donker et al. 1993), copper on the marine invertebrate *Styela plicata* (Galletly et al. 2007) and cadmium and nickel on the benthic oligochaete *Limnodrilus hoffmeisteri* (Klerks and Levinton 1989). Additionally, it has been observed that gene duplication for metal chaperones has occurred throughout populations of *Drosophila melanogaster* (Maroni et al. 1987), while Shirley and Sibly (1999) demonstrated that cadmium resistance in the species was dependant on a single gene that conferred pleiotropic effects on life-histories. Finally, Langdon et al. (2009) have noted multigenerational tolerance to arsenic in a population of the earthworm, *Lumbricus rubellus*. The range of taxa in which multigenerational resistance has been observed, the environments that they occur within (terrestrial, freshwater, marine) and range of contaminants whose effects

are mitigated through genetic variation suggests that selection as a response to exposure to pollution may be a relatively common phenomenon in natural populations.

The adaptive capacity of invertebrates suggests that they are good models for investigating resistance to environmental variation. Primarily, many species belonging to different taxa are frequently ubiquitous widely distributed. These organisms are often relatively small and thus take little effort and expense to sample and maintain in the laboratory. Invertebrates are also likely to have short generation times, which are beneficial for breeding analyses that are subsequently important for determining the genetic bases behind adaptive traits. There are also no complex issues demanding ethical evaluation in experimentation as in vertebrates (and cephalopods). *C. elegans* and *D. melanogaster* have proven to be extremely useful for evolutionary studies because of their rapid turnover and extremely well-resolved biology (Izutsu et al. 2012; Lopes et al. 2008; McGrath et al. 2011; Shirley and Sibly 1999). The comprehensive level of analysis undertaken using the two species has meant that they are termed "model organisms"; however, they have faced neglect relative to environmental sciences (Stürzenbaum et al. 2009). In spite of this, the two may be considered unsuitable for determining genetic variation relative to metal exposure. Recently, it was observed that low levels of global differentiation occur across *C. elegans* populations, which demonstrate no differentiation at local and continental scales (Andersen et al. 2012; Dolgin et al. 2008). This is thought to be the result of anthropogenic activity and is similar to global differentiation as seen in *D. melanogaster*, which has been recently dispersed by human action from its origins in west Africa (Blaxter 2011; Haddrill et al. 2005). Comparative methodology not only means that it is easier to observe conservation of specific biological mechanisms in response to exposure, but also ensures that novel findings are more broadly applicable. What's more, van Straalen and Feder (2012) suggest that the two invertebrate models have specialised genomes and are less representative of bilaterlia (two-sided symmetrical animals), citing annelids (ringed worms- earthworms, ragworms and leeches) as a more appropriate representative. Thus, a more appropriate environmental sentinel is required. Of acknowledged organisms that have so far been determined to have especially important ecological role, with evidence for adaptation to environmental perturbation and can facilitate a degree of contemporary analyses, the earthworm stands out as an excellent candidate.

Earthworms are largely responsible for the movement and degradation of organic matter in the terrestrial environment, while their activity enables air and water to more effectively permeate soil (Peijnenburg and Vijver 2009). Their role in determining the

organic content and structure of the soils in which they inhabit has led to them being identified as ecosystem engineers (Lavelle et al. 1997). Earthworms are also considered to be environmental sentinels, so called as to represent the communities in which they are part of during biological monitoring through a level of environmental sensitivity that is general of that for other organisms. Thus earthworms are used in standardised tests for comparable evaluation of contaminants and the effects that they may have upon mortality and life-history end points (OECD 1984; OECD 2004). This, in part, is due to the fact that earthworms adhere to the criteria of experimental suitability, as above. In addition to this, earthworms are long lived, while their movement and subsequent colonisation of new environments is limited as a result of their capacity for migration. This means that long-term exposure of individuals at polluted sites can be anticipated. The results of such studies are used to determine policies regarding environmental contamination and further evoke their potential use as a candidate for exploiting contemporary analytical efforts that allow for insights into adaptation in natural populations and the ecological implications of such phenotypic and evolutionary change.

Of the earthworm species found in the UK, *L. rubellus* is debatably most ubiquitously distributed and appropriate environmental sentinel, occurring throughout polluted and uncontaminated sites (Stürzenbaum et al. 2009), while being comparatively sensitive relative to other species (Langdon et al. 2005a; Spurgeon et al. 2000). Importantly, adaptation to metals has already been studied in *L. rubellus* to the extent that evidence of clear difference between populations originating from contaminated land and naive earthworms, with no previous history of exposure, have been demonstrated (Langdon et al. 2001), with more recent work suggesting a genetic basis for such adaptation (Langdon et al. 2009). To date research focussing upon the bases of metal adaptation in earthworms has yet to take advantage of molecular endpoints that can elaborate upon the mechanisms involved in As toxicity and population response to long-term exposure. Although a number of studies have focussed upon a variety of endpoints relating to As toxicity, the mechanisms remain poorly resolved in *L. rubellus*. This is in fact shared among other taxa, which is likely due to the bivalent nature of inorganic arsenicals found in organisms and their disruption of fundamental cellular operations, with different species acting as a phosphate analogue, denaturing proteins and generating free radicals (Norton et al. 2008). Insights into the metabolism of As in naive populations will not only set a precedence for comparison of *L. rubellus* with other organisms, but will also set a basis for interpreting genetic variation between adapted and non-adapted populations. This effectively allows for discrimination between

temporary acclimation and permanent resistance to offer an insight into key evolutionary and ecological phenomena.

Many ecotoxicological studies begin with the acquisition of life-history data to represent the effects of toxicity at the populations and individual levels of biological organisation. The endpoints interrogated represent mechanisms important for the normal function of fundamental biological mechanisms responsible for population dynamics. Mortality is an easily observable and well defined parameter already used to evaluate the response the respsone of detailed to occur following exposure of earthworms to As exposure (Langdon et al. 2001). Effects upon growth and reproduction following exposure also represents the management of resources for mitigating imposed stressors or the role of toxicants in inhibition of such processes and are useful for examining the effects of exposure to contaminants (Scott-Fordsmand et al. 1998). Juvenile exposure and determination of the effects upon growth rates can further establish degrees of sensitivity within populations and relates to the potential for continuation of populations under such circumstances (Spurgeon et al. 2004b). Collectively, these endpoints have previously been used to derive population-level effects of exposure for projections regarding the long-term effects of exposure at specific concentrations via demographically based (e.g. matrix) models. These models have the potential to improve understanding of environmental stress imposed upon natural populations (Klok 2007; van der Ploeg et al. 2011).

Beyond assessment of acute and chronic demographic trait effects, molecular endpoints are essential for defining the mechanistic variation. In earthworms, measurement of internal As chemistry via speciation has proven popular. Arsenic speciation interrogates the manner by which As is metabolised. The approach relies on the use of analytical methods such as high profile liquid chromatography inductively coupled plasma mass spectrometry (HPLC-ICP-MS) to identify, as well as quantify, the different As species that occur within exposed earthworms. A range of investigations have focussed upon organic arsenicals in an attempt to associate potential mechanisms for detoxification with novel forms that include arsenosugars (Geiszinger et al. 2002) and arsenobetaine (Langdon et al. 2002) as well as methylated arsenicals, widely thought to relate to metabolism (Challenger 1945). The role of these species in the metabolism of As by earthworms is less clear, given that they may actually be a derivative of microbial action (Button et al. 2012; Ritchie et al. 2004). More recent work has established that inorganic arsenicals are the dominant forms found in earthworms (Button et al. 2009; Button et al. 2012). Clearly, As speciation is an important aspect of determining the interaction of arsenicals with biological mechanisms following

exposure. However, such analyses cannot single-handedly provide the resolution required to fully characterise the metabolic pathways responsible for metabolism of and potentially resistance to As in *L. rubellus*.

Earthworms have previously seen the deployment of molecular techniques to better understand metal toxicity; however, current genetic resources remain limited, while the level of genetic diversity among populations (both in the UK and outside) has not been adequately defined. Fortunately, new techniques allow for -omics analyses, encapsulating all of the variation observable within respective endpoints. Genomic, transcriptomic and metabolomic methodologies are becoming more widely available for holistic output of molecular endpoints. These were recently endorsed by van Straalen and Feder (2012) for considering the ecological and evolutionary implications of environmental risk assessment. Collectively, these can provide a means by which mechanisms for As metabolism can be determined, which will permit insight as to the means by which adaptation is derived, as mentioned previously. This is in keeping with what is now “systems biology”, whereby multiple endpoints are used to better resolve biological mechanisms (Williams et al. 2011). Bundy et al. (2008) described a systems biology approach to describing Cu toxicity in *L. rubellus*, using ^{1}H -NMR and cDNA microarrays to examine metabolomic and transcriptomic endpoints, respectively. Recent advances mean that more powerful methods for transcriptomic analysis permit *de novo* sequencing of an entire transcriptome, yet to be characterised in *L. rubellus*.

The basis of such methodological progress is shared with genomic endpoints, as both may take advantage of contemporary high throughput sequencing capabilities. New methods for genetic analysis can also allow for differentiation of genotypes to posit variation relative to adaptation (Davey et al. 2011). Previous methods that provide a generalised overview of markers across the genome implemented restriction enzymes. Combining these approaches with new sequencing methods now permits a far higher density of markers for analysis (Davey et al. 2011). This is exemplified by the outputs relating to genetic diversity in *C. elegans* populations. Thus, Barriere and Felix (2005) were able to detect the low levels of genetic diversity that occurs between sites, while Andersen et al. (2012) were able to determine generalised views of variation across chromosomes and are able to suggest the means by which low diversity occurs globally, using Restriction Associated DNA sequencing (RADseq). Following digestion by a restriction enzyme, adaptors are ligated onto size selected fragments, which are then selectively amplified (Baird et al. 2008) prior to sequencing. What’s more, the range of SNPs derived from the thousands of markers in RADseq can be used to produce concatenated sequences for phylogeographic analysis for better definition

regarding the ecology of metal adaptation. The use of this approach to investigate genetic variation in natural populations has been illustrated in a study conducted by Emerson et al. (2010). These authors were able to use RADseq to elucidate the phylogeography of the pitcher plant mosquito (*Wyeomyia smithii*), by initially establishing mitochondrial genotypes before using RADseq to better resolve adaptive radiation. The results highlighted the value of phylogeographic analyses for defining genetic variation relative to environmental variation. To date, phylogenetic analysis in *L. rubellus*, relative to toxicity, have been implemented only a study by Andre et al. (2010a). The authors described variation across an abandoned lead and zinc mine. The study indicated that populations were potentially under selection and that the extent of genetic differentiation could be related to metal concentrations, leading to genetic erosion and association of specific haplotypes with increasing soil concentrations. This work was based upon mitochondrial work which has already been used to determine that at least two clades occur in the UK, thus supporting the occurrence of cryptic speciation among earthworms previously described using a mitochondrial marker (King et al. 2008). Further, it is unknown as to the adaptive potential of each of the two relative to specific metals, or even variation in life-history parameters. Any analyses made of natural populations will be required to acknowledge this basal level of variation. As mentioned previously, it will be important to demonstrate adaptive phenotypes, which can be difficult to determine if the trait encompasses complex genetic variation. When crossed with molecular methods in natural populations, this should not only give insight as to the mechanisms responsible, but should provide insight to the ecology mechanisms behind adaptation.

1.7 Research Aims

As outlined above, new advances in contemporary sequencing technologies now enable for better identification of genetic variation occurring relative to the adaptive mechanisms that can be used to address response to environmental stress, such as metal contamination. More fully resolving the metabolic mechanisms relating to adaptive traits can provide insight into specific genetically based traits that enable continuation in heavily polluted sites, important for recognising the long-term effects upon communities. Molecular endpoints will benefit from anchoring by life-history parameters, which connect toxicity and metabolism with long-term, population level effects.

Recent evidence in the environmental sentinel, *L. rubellus* has suggested a genetic basis for As resistance. Although the genetic resources and related ecological findings

for this species are at present limited, holistic -omics endpoints will allow for a swift influx of large-scale data sets for addressing this. The summation of these points will provide an insight into the ecology of *L. rubellus* relative to anthropogenic sources of pollution, while establishing the long term effects of exposure upon populations relative to the temporal nature of adaptation. The following investigations will therefore be used to achieve the goal of deriving phenotypes relative to the temporal nature of exposure to As among populations of *L. rubellus*:

To describe the effects of As exposure upon *L. rubellus* relative to life history parameters using a matrix model, while investigating the effect of exposure upon haplotypes at recognised mitochondrial levels.

To implement -omics endpoints to interrogate As metabolism in unadapted populations, based upon life-history. Metabolomic and transcriptomic analyses will be used to more fully resolve phenotypic plasticity relating to As speciation in *L. rubellus.*

To observe the effects of As exposure among inbred lineages that links phenotypic variation relating to resistance in natural populations.

To demonstrate genetic variation relative to metal contamination in natural populations of *L*. *rubellus* across the UK using a recognised mitochondrial marker as well as those derived from RADseq, for high throughput analysis of genomic markers.

Thus, through integrating a range of analyses across multiple levels of biological organisation, the earthworm *L. rubellus* will be used to distinguish adaptive mechanisms incorporated for management of As toxicity.

Chapter 2

Life-History Effects of Arsenic Toxicity in Clades of the Earthworm, *Lumbricus rubellus*.

2.1 Introduction

The long-lasting, deleterious effects of metal contamination have been the subject of many ecotoxicological analyses (Walker et al. 2006). Arsenic (As) is one of several trace elements that has contributed towards aquatic and terrestrial contamination. Generally, soil surface concentrations of As are relatively low (<11 mg/kg As; Emmett et al., 2010), however, processes such as mineral extraction can lead to localisations that bare enormous concentrations, ranging from between 2 and 3 orders of magnitude above background levels (Klinck et al. 2005). These soils are capable of enforcing a significant toxic burden upon natural populations. Consequently, As has been the subject of many environmental investigations and has more recently been recognised as a contemporary toxicant of concern for humans such as in Asia (Lu and Zhang 2005; Stroud et al. 2011).

Earthworms are sentinels for terrestrial systems on account of their definitive ecological roles. Consequently they have been widely used in ecotoxicological studies attempting to resolve the deleterious effects associated with metal exposure (Lowe and Butt 2007). Previous investigations involving the species *Lumbricus rubellus*, have attempted to resolve the effects of As on specific life-history parameters (Langdon et al. 2003b). Basic toxicological data has established that adult earthworms lose condition when exposed to 2,000 mg/kg As (Langdon et al. 2009; Langdon et al. 1999), contributing to significant levels of mortality in laboratory exposures. Exposure to As-contaminated field soils has also been shown to cause weight loss in exposed earthworms (Shin et al. 2007). Such single endpoint analyses are vital for determining the short term ecological implications of exposure; however, identifying multiple traits affected by perturbation can provide additional insight into population level impacts.

Integration of toxic effects on multiple traits has been proposed as a more appropriate means of extrapolating endpoints to better describe population responses to exposure (Forbes et al. 2008). Population growth rate, in particular, has been found to provide good insight into long-term population level susceptibility (Klok et al. 2006; Kooijman 2000). For earthworms, Spurgeon et al. (2003a) developed an approach to effect assessment that allows for the comparison of toxic effects between different life-

stages. Applying the stratified testing approach revealed that juveniles are particularly sensitive to cadmium and copper toxicity (Spurgeon et al. 2004b). Furthermore, reduced cocoon production as a result of exposure has also been linked to delayed maturation (Spurgeon and Hopkin 1996). These examples highlight the value of integrating sublethal endpoints within a demographic context. Here, the approach of Spurgeon et al. (2003a) is utilised to quantify the toxic effects of As for a range of relevant life-cycle traits in *L. rubellus*.

Although demographic studies integrate effects more comprehensively than studies concentrating on a single trait, they do not explicitly acknowledge the role of genetic variation (Calow 1996). Variation in sensitivity between individuals for a given endpoint may be indicative of genetically distinct cohorts within populations that respond differently to the selective pressure of the toxicant (Morgan et al. 2007). This is particularly relevant in *L. rubellus*, as recent evidence shows that populations inhabiting the extensively surveyed former As and copper mine site at Devon Great Consols (DGC), UK, are able to maintain populations in soils containing in excess of 8 000 mg/kg As (Button et al. 2009; Langdon et al. 2009; Langdon et al. 1999). Further, evidence has suggested a high degree of localised genetic differentiation occurs in populations across metal contaminated sites (Andre et al. 2010a). This is suggestive of genotype specificity in the ability to tolerate exposure and promotes the inclusion of genotypes to identify any possible sources of variation within ecotoxicological data (Calow 1996; Morgan et al. 2007). In this study, the influence of genotype (as defined by sequencing of an established mitochondrial marker) upon *L. rubellus* life-history parameters relative to As sensitivity is examined.

2.2 Materials and methods

Established protocols were used to assess As effects on *L. rubellus* adult and juvenile parameters and total tissue As, as well as phylogenetic analysis for juvenile worms. The exposure protocols were based on Spurgeon et al. (2003a). Earthworms used throughout were field-collected *L. rubellus* purchased from Lasebo (Nijkerkerveen, Holland) and maintained in an uncontaminated culture of artificial soil consisting of a 1:1:1 mix of loam soil: peat: composted bark for two months.The detailed procedures used are summarised below.

2.2.1 Test Medium

Soils used for both adult and juvenile tests consisted of commercially available clay loam soil (Broughton Loam, Kettering, UK), that was sieved (2 mm) and then mixed

with 3% composted bark (LBS Horticultural, Colne, UK) to give a suitable test medium (pH 7.1, organic matter approx. 5%) as previously used and fully characterised by Hooper et al. (2011) and Spurgeon et al. (2004b). For the dry medium, 1400 g and 350 g were used for the adult and juvenile exposures, respectively. Identical As exposure concentrations were used in both the adult and juvenile tests, permitting integration of trait parameters. Distilled water or a solution of sodium arsenate ($Na_2HAsO_4 \cdot 7H_2O$) (Sigma Aldrich, Dorset, UK) was added to provide a moisture content of 60% of the established water holding capacity and soil concentrations of 0, 3, 12, 36, and 125 mg/kg arsenic as reflecting results within the available literature (Fischer and Koszorus 1992; Langdon et al. 2002; Meharg et al. 1998; Yeates et al. 1994). Soils were then left for 7 days to allow initial speciation equilibrium of As (Spurgeon et al. 2003a).

2.2.2 Adult Arsenic Exposure

Prior to exposure, fully clitellate *L. rubellus* were collected and acclimated under test conditions for 48 hours. In the adult test worms were incubated as a group within 1400 g dry weight of test soils held within a polypropylene container of dimensions 180 mm x 180 mm x 93 mm). For each of the five concentrations, 5 replicates were used. For each replicate, 5 adult worms were washed, blotted dry and collectively weighed before addition to the test soil. 8 g (dry wt) of dried horse manure was contaminated with corresponding As concentrations and rewetted to 80% moisture content before addition across the soil surface in each container. Containers were then covered to limit water loss and kept at 14 ^{o}C in constant light for 28 days. After 14 days, any remaining manure on the soil surface was removed and replaced by a further 8 g of fresh manure as food. At this time, worms were also retrieved from the soil, weighed, and visually inspected for phenotypic characteristics before being returning to test conditions. After 28 days, excess manure was removed and the worms were again weighed and inspected before, then immediately frozen and homogenised under liquid nitrogen. These samples were then stored at -80 ^{o}C and used to determine metal/metalloid tissue concentrations. Cocoon production rates (cocoons/ worm/ day) were determined by sieving the soil following exposure, and comparing the number of cocoons collected with survival data to calculate cocoon production rates. Worms were not depurated to maintain the integrity of samples for later metabolomic and gene expression studies (Chapter 3).

2.2.3 Juvenile Arsenic Exposure

In this test, each concentration consisted of 25 replicates, with each replicate containing an individual worm. To initiate the test, 350 g dry weight of soil was contaminated with As as in the adult exposure, before an individual hatchling was

weighed and added. Initially, 1 g (dry wt) of dried horse manure was contaminated with corresponding chemical concentrations and rewetted to 80% moisture content before addition across the soil surface. Containers were then covered to limit water loss and kept at 14 ^{o}C in constant light. Survival, growth and development of worms were recorded over 280 days, every 28 days. At each time interval, remaining manure was removed from the soil surface. The worm in each container was then retrieved and weighed prior to return into the same soil and addition of 1 g dry weight of fresh, spiked food. When individuals had developed a clitellum, they were considered to be adult. From this time food was increased to1.5 g dry weight to cater for higher metabolic requirements. Following completion of the experiment, ~10 mg of tissue was removed from the posterior end of each individual using a scalpel. Tissue samples were frozen in liquid nitrogen and stored at -80 ^{o}C for genotyping using a mitochondrial marker.

2.2.4 Juvenile Genotyping

DNA was then extracted using the QIAamp DNA mini kit (Qiagen, Paisley, UK) prior to polymerase chain reaction (PCR) amplification according to Andre et al. (2010a) using cytochrome oxidase gene (COII) forward (TAGCTCACTTAGATGCCA) and reverse (GTATGCGGATTTCTAATTGT). PCR products were checked by gel electrophoresis and purified for sequencing using 0.25 U each of Exonuclease I and Shrimp Alkaline Phosphatase (NEB, Hitchin, UK) incubated at 37 ^{o}C for 45 minutes and 80 ^{o}C for 15 minutes to inactivate the reagents. Purified PCR products were then sequenced as in Andre et al. (2010), using ABI PRISM® BigDye v3.1 Terminator Sequencing technology (Applied Biosystems, USA).

Sequences were aligned and trimmed for tree construction using the Maximum Likelihood method and General Time Reversible model with a gamma distribution Mega (v5.01). Tree topology was supported by bootstrap analyses over 1000 iterations and the use of outlier sequences from the other lumbricid species. Sequences for *L. rubellus* associated with specific mitochondrial clades already documented in the UK were incorporated as anchor sequences.

2.2.5 Soil Arsenic Analysis

Approximately 5 g of soil were removed from 5 replicates of each concentration from both the adult and juvenile toxicity tests following termination of the exposure. Soils were dried for 48 hours at 80 ^{o}C before being passed through a 2 mm sieve. Soils were analysed for total As concentration as described in Spurgeon et al. (2011). Briefly, following aqua regia digestion using a microwave digestion, samples were quantified using a Perkin Elmer Elan DRC II inductively coupled plasma mass spectrometer (ICP-

MS) (Perkin Elmer 4300DV, Cambridge, UK). Quality control within the analyses was conducted using the standard reference material ISE 192 (International Soil Exchange, The Netherlands). Certified values for reference materials corroborated well with data in this analysis, being within 10% of mean certified concentration for As concentrations.

2.2.6 Tissue Metal Analysis

Metal tissue analysis was conducted for worms exposed in the adult test using homogenised tissue derived from all surviving worms within each replicate. Worms were not depurated following exposure**.** These samples were analysed for total As concentration using 0.5 g of dried sample digested with 10 ml of 70% HNO3 (Ultrapure) at 200 ^{o}C for 15 minutes within a microwave digestion vessel. Samples were run on the Perkin Elmer DRCII ICP-MS. Each run included the certified reference materials TORT-2 and DOLT-4 (National Research Council, Canada) for quality control. Certified values for reference materials corroborated well with data in this analysis, which were within 1% of certified values for As.

2.2.7 Statistics

Effects on mortality, weight change, cocoon production and juvenile maturation were checked for normality using the Kolmogorov–Smirnov statistic and then further analysed using analysis of variance (ANOVA) in Minitab (v15) ($\alpha = 0.05$). Cocoon production was corrected for mortality events, thereby allowing cocoon production rate (cocoons/worm/week) estimation. Tukey's test was used for the post-hoc pairwise comparison of cocoon production rates between the different treatments. An LC50 for juvenile *L. rubellus* exposed to As was calculated the survival data at 280 days with the trimmed Spearman-Karber method (Hamilton et al. 1977). The EC50 for effects on cocoon production was calculated by fitting cocoon production rate data for each replicate using a logistic model in Sigmaplot (v10.0).

The patterns of juvenile growth relative to exposure concentration were fitted using a Gompertz model in SigmaPlot 12.0 using mean mass values for the sampling intervals throughout the exposure. This model is summarised as below.

$$W^t = A + Ce^{-e^{-B(t-M)}}$$

Where W^t is the weight at time t; A the lower asymptotic weight (g), C the upper asymptotic weight (g), M the age at the inflection point (d), and B is the Gompertz growth constant (days^{-1}). From these parameters, maximum growth (at the point of inflection) was calculated using the expression,

$BC^{-e^{-1}}$.

The Gompertz model has been shown to provide a good description of the patterns of earthworm growth following trace metal exposure (Spurgeon et al. 2004b). The influence of haplotype variation on both the length of the juvenile period and size at maturation under different As treatments was assessed using a General Linear Model (GLM), with sequenced haplotype and As concentration as the fixed factors including the combined factor also within the analysis. Individuals that did not mature were not included in this analysis.

2.2.8 Population Model

To integrate the toxic effects of As on multiple life-cycle traits, life-history data were used to predict the intrinsic rate of increase (λ) using separately parameterised models for each test concentration. Stage structured matrix population models were used as they are mathematically manageable and carry enough biological information for risk assessment (Caswell 2001; Forbes et al. 2011; Neubert and Caswell 2000). A three stage matrix population model (Figure 2.1) was used. This incorporates i) a cocoon stage (during which early development of the young worms occurs) in which worms either stay in cocoons, hatch to become juveniles, or die; ii) a juvenile stage during which worms may stay juvenile, mature to adults, or die; and iii) an adult stage in which individuals either remain as adults producing cocoon at a given rate, or are removed from the population. A value for λ can then be calculated as the dominant eigenvalue of this projection matrix.

To parameterise the models for each of the nominal concentrations used in the toxicity test (0, 3, 12, 36, 125 mg/kg As), separate 3 x 3 matrices were populated with values describing these stage specific values and transitions based on prior knowledge and experimental data. For the cocoon stage, experimental data were not available and so previous cocoon hatching information was used to provide default values to be used for all test concentrations. Cocoons were assumed to hatch after 56 days, with 80% producing a single hatchling per cocoon (Kammenga et al. 2003). This simplification may result in underestimation of population growth rates if As exposure has an effect on the developing juvenile.

For the remaining values, the matrices were parameterised using experimental data for each life-stage specific for each nominal soil As concentration in the adult and juvenile tests. The duration of adult exposure corresponded to the projection interval of the model (28 days), therefore adult survival and cocoon production rates were used

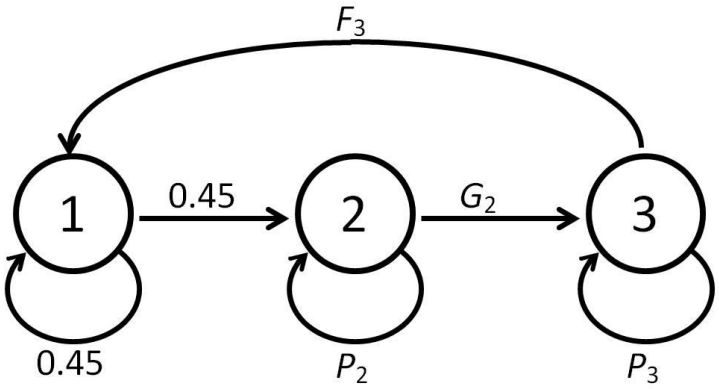


Figure 2.1 A stage-structured life-cycle model for *L. rubellus*, where nodes 1, 2 and 3 represent cocoon, juvenile and adult life-stages, respectively. P_i is the likelihood of remaining as the current stage, G_i is the probability of progression to the next stage and F_3 is cocoon production/worm/week, where $_i$ is life-stage. The probability of cocoon survival and progression is fixed from literature data (see text).

directly as parameters. Juvenile parameter values were derived from the measurement of survival to adulthood and maturation time relative to exposure. Error measurements for λ were estimated via parameter estimates and associated standard deviation data over 1000 iterations.

2.3 Results

2.3.1 Soil Physico-Chemistry

Background As concentrations in uncontaminated soil were 16.5 ± 0.6 mg/kg. Negating background concentrations by subtracting from totals indicated that measured values were within 15% of nominal concentrations. For simplicity, concentrations are referred to according to nominal concentrations throughout this text, although measured concentrations were used for the LC50 and cocoon production EC50 calculations.

2.3.2 Adult and Juvenile Mortality

There was good survival of worms in the Adult toxicity test, with no significant reductions found relative to exposure, as determined by ANOVA ($P > 0.05$). In controls 88% of worms survived the experiment, falling to only 80% among worms at the highest concentration used (125 mg/kg). In the juvenile test, survival rates over the experiment were much lower in comparison to those in the adult test. ANOVA indicated a significant effect of As concentration on juvenile survival to 280 days ($P < 0.001$). This reflects a stark contrast, whereby all worms in control conditions remained alive throughout the experiment, but survival was reduced to 12% at 125 mg/kg As. Using

survival data of juveniles at 280 days, an LC50 of 67 mg/kg (95% confidence interval= 57.44-78.35) could be estimated for this life-stage.

2.3.3 Cocoon Production

Cocoon production rates in control soils were in accordance with the results of previous work. For example, Klok et al. (2007) found an overall mean of 0.51+/- 0.07 cocoons/ earthworm/ day when analysing the effect of density upon reproduction in unpolluted soils. This compares to the rates of 0.7 +/- 0.3 seen in the control population in this study. Reduced cocoon production was seen in adult worms exposed to the higher soil As concentrations (Figure 2.2). ANOVA confirmed that cocoon production was significantly affected by As exposure ($P < 0.001$), with post-hoc assessment using Tukey's test indicating that cocoon production was significantly reduced at 125 mg/kg, while all remaining treatments were not significantly different from the control. The strong decline in rates of cocoon production observed resulted in a logistic model that described a steep decline in reproduction between the two highest treatment concentrations. On the basis of this fit ($R^2 = 0.57$) an EC50 of 118 mg/kg As (95% confidence interval not available) could be estimated.

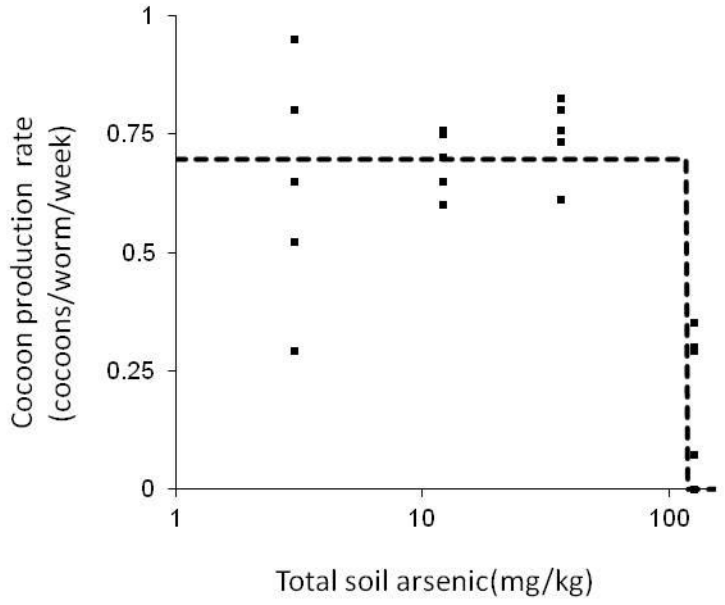


Figure 2.2 Cocoon production rate of *L. rubellus* exposed to a range of concentrations of As for 28 days. Points show cocoon production rate for individual replicate containers (cocoons/worm/week), dashed line shows fitted logistic model.

2.3.4 Adult Earthworm Tissue Arsenic Concentrations

Following log-transformation to ensure normality of the data, earthworm tissue As concentrations significantly increased relative to soil concentrations (ANOVA $P < 0.001$) (Table 2.1). Post-hoc comparison indicated that worms exposed to soil As concentrations of 12, 36, and 125 mg/kg had significantly higher tissue As concentrations (2.0 ± 0.5, 5.9 ± 0.7, and 26.1 ± 4.06 mg/kg, respectively) compared to worms exposed to the control soil (0.8 ± 0.2 mg/kg As).

2.3.5 Growth and Maturation

Juvenile *L. rubellus* mass was measured at 28 day intervals over a 280 day exposure. Figure 2.3 demonstrates that higher soil As concentrations resulted in reduced mean growth rate and final body size. Worms exposed to the higher concentrations were significantly affected (ANOVA $P < 0.001$), reaching only 0.084 g by the end of the exposure, as compared to 1.8 g in control worms. Post-hoc analyses further demonstrated that worms exposed to 36, as well as 125 mg/kg, were significantly smaller after 280 days than worms in the control and lower concentration treatments.

The temporal patterns of juvenile weight change were fitted using Gompertz curves parameterised using mean mass values taken from each sampling interval throughout the exposure (Figure 2.3). Fitted models effectively describe *L. rubellus* growth, as reflected by the high R^2 values (> 0.99) for all concentrations except 125 mg/kg (R^2 = 0.8422). Model fits indicated that growth rate decreases as the concentration of As increases (Table 2.1). The upper asymptotic weights confirmed the trends for reduced final mass at higher soil As concentrations. Between growth curves, estimated values for the point of inflection range from 194 – 90 days, with the lowest value seen at the higher soil As concentrations, potentially due to the limited increase in body mass occurring at the end of the experiment when body tissue As concentrations can be expected to be highest. At the sampling interval closest to the point of inflection of control worms (196 days), worms at 125 mg/kg As were significantly smaller than all other treatments (ANOVA, $P < 0.001$), while those at 36 mg/kg were smaller than worms exposed to 3 mg/kg and control worms. This reflects the smaller maximum growth rates estimated for these treatments.

Maturation status was used to establish the length of the juvenile period for each individual, which is considered to be significantly dependent upon As concentration (ANOVA $P < 0.001$). Most notably, earthworms exposed to 125 mg/kg of As failed to

Table 1. Coefficients of determination (R^2), Gompertz parameters and the maximum growth rates for the growth curves for juvenile *L. rubellus* exposed to arsenic concentrations over 280 days (d). Gompertz parameters include; A the lower asymptotic weight (g), C the upper asymptotic weight (g), M the age at the inflection point (d), and B is the Gompertz growth constant (days^{-1}). Mean masses at sample intervals closest to the point of inflection of controls (196 d) are shown, as well as in the middle (140 d) and end (280 d) of the exposure regime. Modelled population growth rate (λ) over 1000 iterations and the mean tissue As concentration for adults exposed to As for 28 d are also shown. In columns, values with a common superscript are not significantly different ($P < 0.05$).

Exposure concentration mg/kg As	R^2	C (± SE)	B (± SE)	M (± SE)	A (± SE)	Mean mass at time interval (g) 140 d (± SE)	196 d (± SE)	280 d (± SE)	Maximum Growth rate (g/ d)	Lambda (λ) (± SE)	Arsenic tissue concentration mg/kg As (± SE)
0	0.998	2.87 ± 0.331	0.009 ± 0.001	194 ± 12.7	0.0067 ± 0.0314	0.57 ± 0.11 ^{a}	1.08 ± 0.22 ^{a}	1.77 ± 0.35 ^{a}	0.00910	1.153 ± 0.005	0.84 ± 0.09
3	0.999	3.289 ± 0.396	0.007 ± 0.001	223 ± 15.8	-0.013 ± 0.025	0.52 ± 0.11 ^{a}	0.94 ± 0.20 ^{a}	1.68 ± 0.36 ^{a}	0.00860	1.119 ± 0.003	1.13 ± 0.06
12	0.997	2.562 ± 0.411	0.008 ± 20.2	207 ± 19.1	0.0125 ± 0.0318	0.46 ± 0.10 ^{a}	0.87 ± 0.18 ^{ab}	1.48 ± 0.32 ^{a}	0.00744	1.138 ± 0.002	1.95 ± 0.21
36	0.993	1.324 ± 0.160	0.0105 ± 16.5	150 ± 10.2	0.0017 ± 0.034	0.45 ± 0.10 ^{a}	0.66 ± 0.14 ^{b}	1.02 ± 0.22 ^{b}	0.00511	1.069 ± 0.003	5.87 ± 0.32
125	0.842	-0.0763 ± 0.0130	-0.0874 ± 0.182	90.2 ± 14.0	0.095 ± 0.006	0.11 ± 0.04 ^{b}	0.08 ± 0.05 ^{c}	0.09 ± 0.04 ^{c}	0.00245	0.98 ± 0.005	26.14 ± 1.81

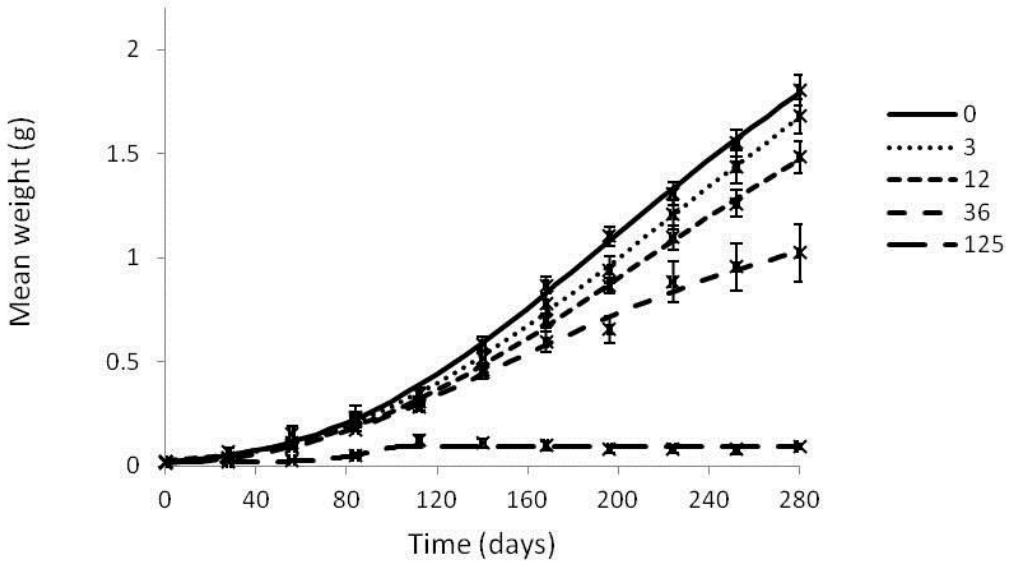


Figure 2.3 Mean weight (+/- SE) for *L. rubellus* exposed to a range of As concentrations in soil, over 280 day at each 28-day time interval, with points appearing proximal to relative growth curves. Fitted curve correspond to best fit Gompertz growth curves estimated based on mean time interval weights.

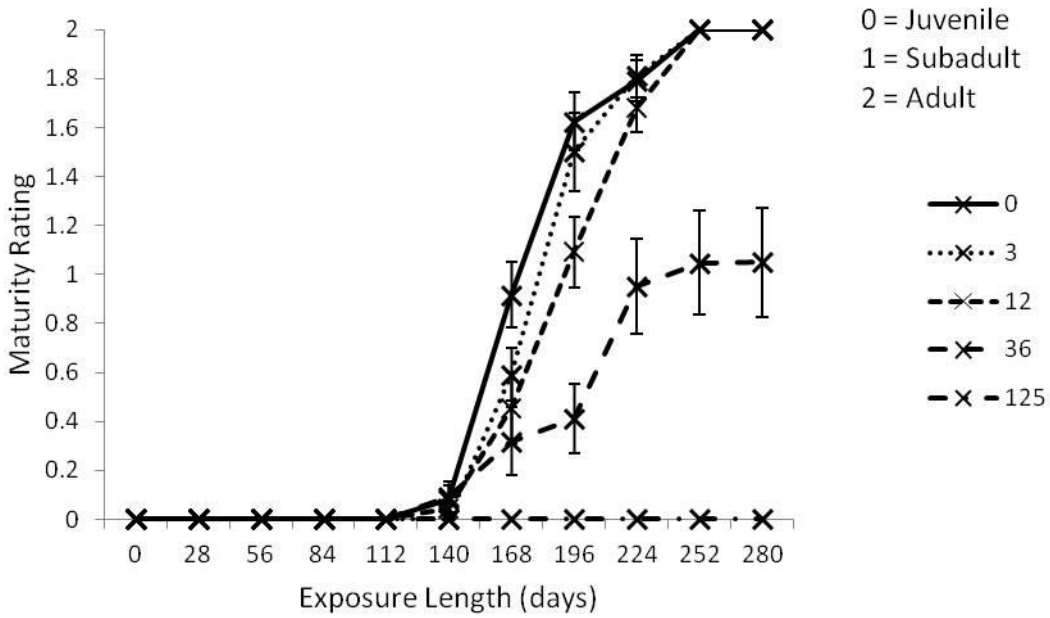


Figure 2.4 Mean maturity rating (± SE) for *L. rubellus* exposed to a range of As concentrations in soil, recorded at sample intervals of 28 days, over 280 days. Individuals that developed a clitellum were determined to be adults (score of 2) those with only tubercle pubertatis as subadults (score of 1) and those with neither as juveniles (score of 0).

mature over the course of exposure (Figure 2.4). Juveniles at 36 mg/kg As also had a significantly lower proportion of mature adults relative to lower exposure concentrations, with only 50% of surviving individuals developing a clitellum after 280 days. Earthworms exposed to 0, 3 and 12 mg/kg As were all mature by the end of exposure, with average maturation times of 207, 211, and 223 days, respectively.

2.3.6 Effects of Arsenic on Intrinsic Rate of Population Increase

Values for adult survival and reproduction were taken directly from the results obtained for each tested concentration in the adult exposure. Values for juvenile survival and maturation were calculated from the juvenile test. Earthworms that failed to mature over the duration of the exposure were assumed to have matured by the putative next sample interval (set at 308 days). This is almost certainly an unrealistic assumption for worms at 125 mg/kg, therefore estimation of λ at this concentration is likely to be lower than the given value. Based upon fitted models, control populations are expected to increase with a λ of 1.15 (see Table 2.1 for all values). Values at 3 and 12 mg/kg As are similar to that for the controls (1.12 and 1.14 respectively). Populations at 36 mg/kg As had a lower population growth rate (1.07) than controls, though it remained above 1, demonstrating reduced stability. At 125 mg/kg As, model estimates suggest that the population will eventually go extinct, with λ at 0.98. Given the severe effects on maturation and the assumption of maturation at 308 days, it may be expected that extinction may in fact occur at rates greater than estimated by the model.

2.3.7 Genetic Variation, Juvenile Development and Arsenic Sensitivity

Earthworm genotypes were determined by sequencing of the COII mitochondrial gene. A Maximum Likelihood model constructed from the obtained sequences resulted in identical clade construction which identified 3 major clades demonstrated further by the posterior probabilities and bootstrap values (Figure 2.5). Some discrepancies were seen within clade 2, although the other two clades were well supported (bootstrap values over 70%), with greater than 7.5% genetic differentiation occurring between identified clades (Hillis and Bull 1993). Outlying sequences from *Lumbricus terrestris* and *Lumbricus castaneus* COII genes were used to define potential misidentification of early stage juveniles, with an average sequence divergence of 20.2% and 19.6% respectively. Comparison of genetic variation between and within clades, as well as the frequency of clades in exposure regimes are shown in Table 2.2. Anchor sequences of 2 previously recognised British lineages (using AFLP and COII sequences) were placed among 2 distinct clades, though not among the most numerous clade found in this study (Andre et al. 2010a). Branches with Bootstrap values below 70% were collapsed.

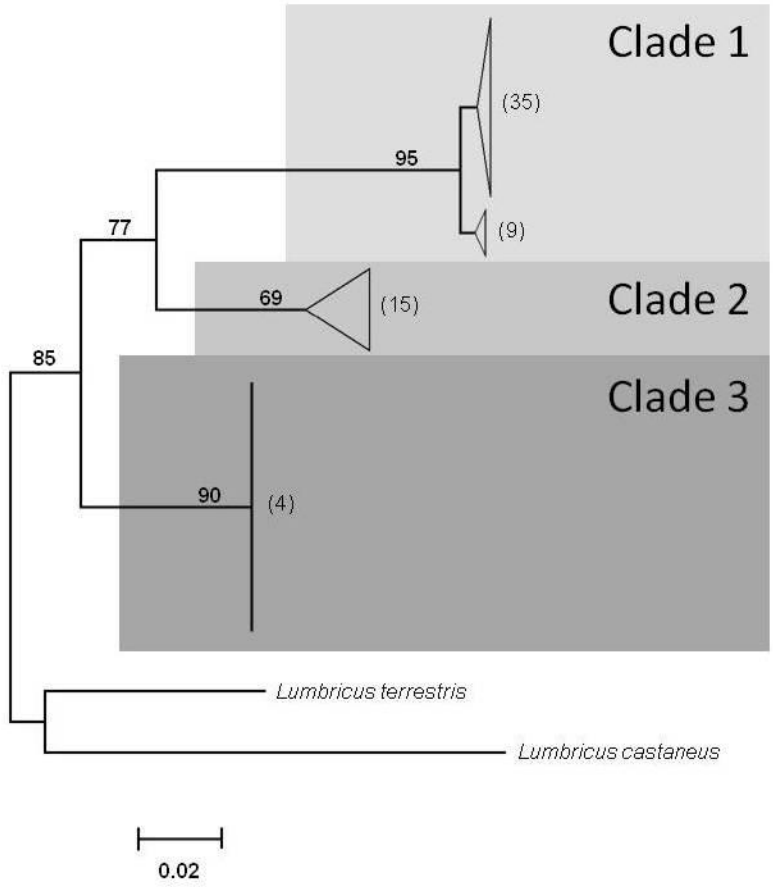


Figure 2.5 Maximum Likelihood phylogenetic tree of the COII mitochondrial gene of 63 *L. rubellus* juveniles exposed to arsenic. Three clades with greater than 7.5% sequence variation are identifiable and the sequences belonging to 2 previously identified genotypes, A and B, have been included as anchors and lie within clades 3 and 2, respectively. Sequences from *L. terrestris* and *L. castaneus* were used as outliers. The number of individuals pertaining to the same haplotype is identifiable next to highlighted branches, with bootstrap support indicated above each node. Values lower than 70% were condensed.

Considering clade effects, GLM showed that the date at which individuals matured was significantly affected by the concentration of As ($F=3.99$, $P<0.05$). Further, haplotype was also found to have a significant effect on maturation time ($F = 10.06$, $P < 0.01$). There was, however, no interaction between the two variables. Post-hoc testing indicated that worms in clade 2 matured at an earlier date relative to worms from clade 1, which represent the majority of worms in this experiment. Individuals belonging to clade 3 were not included in the analysis as only 4 individuals, occurring at a single exposure concentration were present among the sampled worms. For size at maturation, there was no significant effect of either genotype or As exposure

concentration (F=3.2 *P*>0.05) on mass at maturation. The low frequency of individuals from clade 2 (Table 2.2), however, indicate that further analysis is required to more fully document genotype specific variation among life-history parameters.

2.4 Discussion

Our analyses demonstrated both acute and sub-lethal toxicity effects of As. Life-history stages appear to be differentially sensitive, with effects on traits seen at lower concentrations in tests with juvenile worms than those occurring within the adult test. This effect could, however, also relate to the shorter exposure duration used in the adult toxicity test. Integration of affected parameters within a simple demographic model indicated negative effects of As on intrinsic rate of population increase, which will ultimately result in population extinction when at high concentrations. Significant effects particularly seen in juvenile survival and maturation, as well as reduced cocoon production, are mainly responsible for this population level effect of As exposure.

In responding to the presence of elevated As concentrations in soil, the iteroparous nature of *L. rubellus* (Sims and Gerard 1985) is likely to allow for optimisation for maximum reproductive output to exploit contemporary circumstances. Lika and Kooijman (2003) stated that indeterminate growers (who keep growing after maturation), such as the earthworm, were likely to reduce allocation of resources away from reproduction, towards maintenance in cases where individual are exposed to sub-optimal conditions. Widdows et al. (1988) pioneered analyses to provide a quantitative assessment of the energy status of animals relative to growth rate, or rather “scope for growth”. Their seminal study demonstrated reduced growth in the mussel, *Mytilus edulis*, under stress imposed by copper and hydrocarbons, specifically demonstrating this diversion of resources. Dynamic energy budget models have been used previously to explore population level effects of metal exposure in *L. rubellus*, using similar life-history parameters. Disturbance caused through exposure to c60 nanoparticles elicited similar demographic variation as seen during this investigation, and included reduced cocoon production. These effects are attributed to a diversion of resources towards increase maintenance and repair (van der Ploeg et al. 2011).

The multifaceted effects of As exposure on resource allocation are evident in the growth response of the juvenile earthworms, who may instigate diversion of resources for mitigation of toxicity that would otherwise be intended for biomass production (Kammenga et al. 2003). Increased sensitivity to toxicity relative to size can be the result of the proportional surface area available to exposure and also potentially to feeding limitations based upon physiological characteristics such as mouth size

Table 2. The number of individuals across exposure regimes are shown relative to those included in phylogenetic analysis of *L. rubellus* exposed to As via inference from the mitochondrial COII gene. Maximum pairwise distances were calculated to show variation between and within established clades.

Phylogenetic association	Number of individuals at each concentration (mg As/kg)					Representation among sequenced (%)	Total representation (%)	N	Maximum variation in clade (%)	Maximum genetic variation against clade 1	Maximum genetic variation against clade 2
	0	3	12	36	125						
Clade 1	13	8	5	15	3	69.8	35.2	44	1.5	N/A	17.2
Clade 2	4	3	4	3	1	22.2	12.0	14	5.3	17.2	N/A
Clade 3	0	0	4	0	0	8.0	3.2	4	0	15.7	10.8
Unsequenced	8	14	12	7	21	N/A	28.0	35			
Mortality	0	4	3	5	22	N/A	21.6	27			

(Jager et al. 2006; Kooijman 2000). The consequences of such increased exposure and/or sensitivity is clearly demonstrated here both by the increased mortality seen among juveniles at a concentration that only induces sublethal toxicity in adults and also by the impacts on juvenile growth and maturation at a concentration where adults are otherwise unaffected (e.g. 36 mg/kg As). In addition to these effects on resource allocation, it is also likely that As toxicity may directly inhibit growth, maturation and cocoon production. In reality, however, it is not feasible to distinguish resource allocation and toxic effects at the levels of biological organisation under investigation here.

Juvenile maturation time has been shown to have a critical effect upon the intrinsic rate of population increase in earthworms (Kammenga et al. 2003). Reduced values for λ were observed at 36 mg/kg As and higher- an effect underpinned by reduced juvenile growth and maturation relative to lower concentrations. At these higher concentrations, assumptions of maturation for juveniles that failed to develop after 280 days may mean that real population growth rates at this concentration may actually be lower than estimated. The high levels of mortality among juveniles exposed to 125 mg/kg As evidently has profound negative implications for population growth rate. Gompertz fits and maturation data in this treatment provide evidence of very limited growth, whereby juveniles would not be expected to mature and population extinction is assured. When juvenile growth in populations exposed to 125 mg/kg As was hypothetically replaced with that of control populations, λ increased to 1.02. This demonstrates that reduced cocoon production still has implications for population growth and so effect on this parameter contributes to effects on populations exposed at intermediate concentrations. Overall though the importance of growth and development to effects on λ corroborate classical evolutionary theory that delayed reproduction is considered to be the most influential demographic trait affecting fitness (Cole 1954).

In this investigation, a significant increase in As tissue concentration among exposed adults relative to soil concentrations was observed. A similar increase in tissue As was also seen by Button et al. (2009b), who found a linear association of As tissue concentrations with a measure for bioaccessibility, which also increased relative to soil As concentration in *L. rubellus* collected from the Devon Great Consols (DGC) mine site. This comparability suggests similar potential toxicity in these field site worms. Earthworms at DGC have previously been found to maintain condition when exposed to contaminated soils (8 000 mg/kg As) that prove lethal to naive populations (Langdon et al. 2009). Since the data presented here approximate for the relative effects of As upon unadapted, natural population, the selection pressure imposed and lower

sensitivity of the DGC mine site worms suggests that the greater exposure and toxicity may have resulted in selection for tolerance in the field population.

Previous work by Andre et al. (2010a) defined localised genetic differentiation in populations of *L. rubellus* occurring over a former Lead and Zinc mine site in the UK, identifying two morphologically identical clades found to be genetically distinct by approximately 13%. These results identified a need for phylogenetic analysis of As exposed worms during this investigation to distinguish clade specific responses to toxicity. Genotype specific differences in life-history variation during the juvenile exposure were seen, whereby individuals belonging to clade 2 matured earlier than those of clade 1, though there was no significant variation in mass at maturation between clades. There is no evidence for variation in maturation time or size of genotyped individuals relative to exposure and therefore the effects of As upon population growth rate for the two clades are likely to remain consistent.

It is feasible that the earlier maturation observed in clade 2 may confer a competitive advantage. Without cocoon production data or establishing the length of clade-specific lifetime, the fitness consequences of this trait difference are not fully elaborated. It can be assumed that this early maturation could result in an increase in intrinsic rate of increase that would likely see that clade dominate if not associated with trade-offs. However, the low frequency of single clade inhabitance at sites in the UK, as found in work by Andre et al. (2010) may suggest otherwise. As a result of genotyping individuals in the toxicity test, one may be able to better represent the scenario relevant to natural populations, and in doing so distinguish differences relevant to multiple haplotypes of a species complex that varies greatly, genetically. Additionally, this provides information that is pre-emptive of emerging demographic variation seen in the field. Future analyses would benefit from establishing clade specific experiments or through acknowledging genetic variation, as previously suggested by Calow (1996), though establishing and maintaining clade specific cultures remains a challenge. Toxicological analyses, such as the OECD tests incorporating *Eisenia fetida* and *Eisenia andrei*, may not incorporate monophyletic species and therefore warrant further investigation to address the role of genotype on phenotype to provide greater understanding of the basis of sensitivity of this species complex and validity of view this complex as a single species in the derivation of environmental quality guidelines (PerezLosada et al. 2005).

2.5 Conclusion

Soil As concentrations of 125 mg/kg As are estimated to evoke population extinction in the earthworm, *L. rubellus* through affects on juvenile and adult life-history parameters. Significant mortality was observed among juveniles, although not among adults, which produced fewer cocoons in response to exposure. Juvenile growth and maturation was significantly affected at the highest As concentrations tested, notably at 36 mg/kg, suggesting increased sensitivity of juvenile compared to adults, which were not adversely affected at this concentration for measured traits. Deviation of resources to manage toxicity may be responsible for the sublethal perturbation seen in this investigation and corroborates with previous studies. Phylogenetic analyses of juveniles established 3 genetically distinct clades, 2 of which were found to vary in time to maturation, but did not show phenotypic variation relative to As exposure. This investigation further resolves genetic complexity of *L. rubellus* and provides phenotypic anchors for molecular endpoints. The clades present within *L. rubellus* also indicate the presence of substantial genetic variation that may be the basis of adaptive traits.

Chapter 3

Arsenic Toxicity and Metabolism in the Earthworm, *Lumbricus rubellus*.

3.1 Introduction

Long-term mining contamination has previously been found to adversely affect terrestrial soil invertebrate communities that are responsible for ecosystem services (Park et al. 2011; Spurgeon and Hopkin 1995). Earthworms form a large portion of the invertebrate community at many such sites and have a key role in aeration of the soil and decomposition of organic matter. Their important ecological role and ubiquitous distribution means that earthworms can be used as an important environmental sentinel (Smith et al. 2005). Specifically, *Lumbricus rubellus* is considered to be an informative species for ecotoxicology, garnering wide-ranging endpoints from many levels of biological organisation; ranging from ecological, physiological, epigenetic and genetic parameters. In particular, the species has been used to determine the effects of a metal pollution in natural populations (Langdon et al. 2003b; Spurgeon et al. 2003b). One of the most important traits recognised in this area is the adaptive potential seen among the *L. rubellus* species complex. Populations inhabiting former As mine sites have been characterised in respect of the metabolic and adaptive mechanisms responsible for resistance (Langdon et al. 2009; Langdon et al. 2001). Although As toxicity is well understood, the mechanisms for metabolism and therefore adaptation remain relatively poorly resolved (Hughes 2002; Miller et al. 2002).

Arsenic toxicity encompasses a number of mechanisms imposed via flexible chemical characteristics, whereby speciation plays a significant role in determining the nature and extent of disruption imposed upon biological systems (Petrick et al. 2000). The pentavalent form of As (arsenate, AsV) is toxic on account of its ability to act as a phosphate analogue, whereas the trivalent form (arsenite, AsIII) is able to react with sulphydryl groups of enzymes and proteins, thus preventing function (Norton et al. 2008). The relevance of speciation relative to toxicity is further exemplified via the occurrence of high concentrations of arsenobetaine (AB) within marine biota, which is considered to be relatively non-toxic (Francesconi 2010) and is an example of the way organisms are able to mitigate toxicity for the protection of biological processes. Furthermore, there is evidence that the various As species are capable of instigating more generalised cell stress, through production of free radicals that affect a variety of

endpoints, including cellular integrity, metabolic activity and cell proliferation (Miller et al. 2002). Previous work (Chapter 2) has shown that exposure to high concentrations of As can impact on the growth and maturation of *L. rubellus*, suggesting that As instigates a diversion of resources to help maintain cellular function and manage toxicity.

As metabolism has previously been considered to occur via the Challenger pathway (Challenger 1945), whereby a series of reduction and methylation steps were proposed to manage and potentially detoxify As. Investigations have posited that complexation may form the basis of alternative mechanisms (Francesconi and Edmonds 1998), and the inability of plants to methylate inorganic arsenic (Lomax et al. 2012), considered to be key features of the Challenger pathway, demonstrate that metabolic mechanisms may be diverse across a variety of taxa. Previous analyses of *L. rubellus* found at sites heavily contaminated with As have mostly focussed upon As speciation to provide insight into metabolic mechanisms. This work has provided initial evidence for complexation via AB and arsenosugars (Button et al. 2009; Langdon et al. 2003a). However, the majority of accumulated tissue As remains as inorganic AsV or AsIII, while there is more recent evidence that establishes the key role of glutathione (GSH) in the role of managing the imposed toxic burden (Hayakawa et al. 2005). Although such studies provide a unique perspective into potential biological activity in response to exposure and resulting As chemistry, the mechanistic understanding of the underlying biology of complex traits requires a systems biology analysis to resolve metabolic mechanisms and therefore candidates for adaptation (Garcia-Reyero and Perkins 2011).

Previous research has already benefitted from marrying multiple -omics endpoints to provide mechanistic evidence of copper toxicity in *L. rubellus*. Bundy et al. (2008) combined cDNA microarrays and nuclear magnetic resonance (NMR) based metabolic profiling to determine the effects at a range of concentrations, that were underpinned by classic life-history endpoints. Their analyses were able to resolve toxicity at the molecular level, highlighting interference of copper upon energy metabolism among exposed organisms. Subsequent studies have also been able to provide an unprecedented level of understanding of toxicity. For example, Williams et al. (2009) were able to provide markers that highlight disrupted metabolic pathways in the three-spined stickleback (*Gasterosteus aculeatus*) using transcriptomic and metabolomic endpoints, following exposure to environmentally relevant concentrations of the polycyclic aromatic hydrocarbon, dibenzanthracene. Similar analyses implemented by

Zhang et al. (2012) were able to demonstrate the deleterious effects of drinking water contaminated with a range of industrial effluents in mice.

So far, transcriptomic analyses using microarrays have been used to explore inorganic As toxicity among metazoans, although to date these studies have yet to be conducted in tandem with other -omics endpoints. Lam et al. (2006) focussed upon differential expression in the liver of zebrafish (*Danio rerio*) exposed to AsV and described variation among genes that are classically implicated with As metabolism, as well as iron (Fe) homeostasis, ubiquitin-dependent protein degradation and antioxidant activity. Other investigations have used As exposures in human and mouse models to explore As toxicity and also the use of As as a therapeutic agent. Within these studies, a range of candidate genes have demonstrated differential expression in processes including lipid metabolism, apoptosis and immune response (Andrew et al. 2007; Andrew et al. 2008; Andrew et al. 2003; Argos et al. 2006; Kozul et al. 2009; Nelson et al. 2009; States et al. 2012; Wu et al. 2009; Zhang et al. 2012).

Recent progress in transcriptomic analyses allows for construction of a transcriptome using RNAseq, which can represent the majority of genes in the genome, even when a reference genome is unavailable. RNAseq is a relatively novel technique whereby total mRNA can be sequenced, before data is shuttled into a range of bioinformatic pipelines for construction of *de novo* transcriptomes and relative transcript abundance. As such, RNAseq can provide a unique and comprehensive insight into gene expression (Wang et al. 2009). However, recent studies have determined that 30M short-length reads are necessary to provide adequate depth of coverage to fully resolve both frequent and rare transcripts, meaning that cutting-edge sequencing technologies remain expensive (Wang et al. 2011).

To provide insight into the biological effects of As in earthworms as a means of recognising adaptive responses, the effects of As upon *L. rubellus* were investigated by employing a range of –omic endpoints. Tissue derived previously from adult *L. rubellus* in which the sublethal effects of As exposure upon life-history parameters (Chapter 2), were used to define the effects of As exposure at molecular levels. Arsenic speciation by HPLC-ICP-MS was implemented to better associate findings with As metabolism, forming a basis for comparison against previous investigations. RNAseq and ^{1}H -NMR was then used to derive transcriptomic and metabolomics insights towards the biological reaction to exposure and a systems biology analysis. Following recent investigations, it was hypothesised that the role of inorganic As will feature predominantly and act towards demonstrating the role of GSH in metabolism and

detoxification (Nemeti et al. 2012). With respect to this, it is expected that metabolomic and transcriptomic endpoints will signify the importance of the methionine cycle, which has particular relevance to GSH production, which is an established mechanism for As metabolism.

3.2 Methods

The earthworms used for the analysis of As speciation and associate gene expression and metabolite change were obtained from the study of the toxic effects of As on adult earthworm survival and reproduction as described in Chapter 2. The procedure used for this exposure is summarised below. For the analysis of earthworm tissue As speciation and concentration, metabolite and gene expression levels, all surviving worms in each test replicate were homogenised to produce a single earthworm tissue sample representative of each container. These samples were analysed and can be considered as true biological replicates.

3.2.1 Tissue and Soil metal concentrations

In addition to determination of As soil concentrations by methods in 2.2.5, Cu, Zn, Cd, W, Pb, P, Ca and Fe were quantified as described in Spurgeon et al. (2011). As in As tissue analysis described in 2.2.6, concentrations of Al, Mn, Fe, Co, Ni, Cu, Zn, Mo, Cd, Sb, Pb, Hg, and Cr were also determined among homogenised tissue derived from adult *L. rubellus*.

3.2.2 Arsenic Speciation

Prior to extraction of earthworm samples, the stability of As species (AsIII, AsV, mono-methylarsonate (MMA), dimethylarsinate (DMA) and AB) was established under the extraction conditions employed by separately spiking earthworm powder material with each of the As species. Recoveries of spiked arsenic species were on average 103.3%, with no evidence of interconversion between species (particularly between AsIII and AsV). Metal speciation was measured in earthworms exposed to 0, 36 and 125 mg/kg As, while methodologically based upon the analyses performed by Button et al., (2009). Simply, samples were digested using water and methanol prior to analysis by HPLC-ICP-MS, which is capable of quantifying a range of species associated with biotransformation of As (Watts et al. 2008a).

Homogenised earthworm tissue (0.97 g average dry weight) was weighed directly into 30 ml round-bottom Nalgene centrifuge tubes (Thermo Scientific, Langenselbold, Germany). 10 ml methanol : water (1 : 1 v/v) was then added and the tubes shaken on

an orbital shaker at 175 rpm for 4 hours at 20 °C. The extracts were centrifuged at 3000 rpm for 10 minutes (JA-10 rotor, Beckman Coulter, High Wycombe, UK) and the supernatant transferred to a 500 ml acid washed glass evaporation vessel before the methanol was evaporated off using a Syncore Analyst (Buchi, Zurich, Switzerland), samples were then reconstituted to 10 ml using deionised water and analysed immediately. Arsenic speciation analysis employed a Perkin Elmer series 200 HPLC system (autosampler, quaternary pump and temperature controlled column oven) that was coupled to a Perkin Elmer Elan DRC II ICPMS by connecting the analytical column to the ICP nebuliser with polyetheretherketone tubing and low dead volume connectors (all from Perkin Elmer, Cambridge, UK). The HPLC-ICP-MS couple used an automatic switching valve for easy equilibration of the HPLC column without overloading the ICP-MS system with HPLC eluent. The two instruments were coupled in such a way that the injection of each sample solution via the autosampler and subsequent separation and measurement was synchronised automatically using the ICP-MS software (Chromera, Perkin Elmer, Cambridge, UK), enabling reproducible sample injections. The HPLC system used was a Hamilton PRP-X100 (10 µm, 4.1 x 250 mm) (Sigma-Aldrich, Dorset, UK) anion exchange column with a reverse phase guard column (RP-1, Phenomenex, Sigma-Aldrich, Dorset, UK). Arsenic species in the earthworm extracts were separated using a gradient elution with ammonium nitrate buffer adjusted to pH 8.7 with ammonia (i.e. flow rate; 1ml/ minute, 10 minute gradient between 4 mM and 60mM NH_4NO_3). This method achieved good separation for AB, AsIII, AsV, MMA, DMA and the elution profile for speciation is shown in Figure 3.1. The Perkin Elmer Elan DRC II ICPMS was used for all As measurements using standard operating conditions (i.e. monitored ^{75}As, with external calibration for each As species in the range 0-250 µg/L). The ICP-MS calibration standards for AsIII and AsV were prepared using aqueous 1000 mg/l stock solutions (Spex Certiprep, Loughborough, UK) and for AB, MMA and DMA from the corresponding salt (Sigma-Aldrich, Dorset, UK). The derived data could not be directly compared to tissue concentrations due to the lack of certified reference materials (CRMs) in the analysis, meaning that the validity of absolute quantification could not be assured. Therefore, the results for each of the measured As species were provided as proportions of the total signal. This was calculated by dividing values for each replicate by the sum of mean values for each concentration. There were also no other observable peaks representing other As species suggesting that the main As endogenous metabolites were included within the analysis. The data on the proportions of each As species across the treatments was analysed using analysis of variance (ANOVA) based upon the total signal. For those metabolites where significant differences were found, Tukey's test was used to resolve significant variation between treatments.

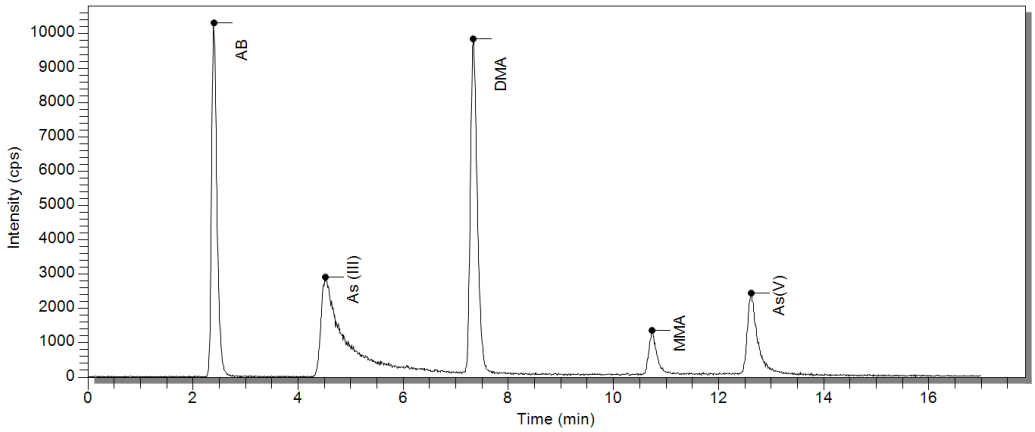


Figure 3.1 HPLC-ICP-MS anion gradient elution profile for a mixture of arsenic species observed in adult *L. rubellus* exposed to sodium arsenate. This profile is derived from a mixture of standards in solution and peaks are annotated: Arsenobetaine (AB), arsenite (As(III)), dimethylarsinate (DMA), mono-methylarsonate (MMA), arsenate (As(V)).

3.2.3 Metabolomic Analysis

Metabolomic analysis was conducted upon all replicates of adult *L. rubellus* exposed to As in Chapter 2. For the extraction of the samples from each biological replicate for metabolite analysis, 0.5 ml of 1 mm zirconia silica beads (Stratech, Suffolk, UK) and 1.5 ml of 70% of HPLC grade acetonitrile was added to approximately 200 mg homogenised tissue in a 2 ml eppendorf tube. The mixture was shaken using a bead beater (MiniBeadBeater 1 Stratech, Suffolk, UK) for 90 s and then centrifuged at 14,000 rpm for 10 minutes. The supernatant was transferred to a new tube and kept on dry ice. Samples were further processed for ^{1}H -NMR analysis as described in detail by Bundy et al., (2008). Briefly, dried extracts were resuspended in NMR buffer and analysed at 300 K on a 14.1 T DRX600 Advanced NMR spectrometer (Bruker BioSpin, Rheinstetten, Germany). Further details of the standardised analysis can be found in Beckonert et al. (2010).

The derived spectra were processed and analysed with Chenomx NMR Suite 4.6 (Chenomx, Edmonton, AB, Canada) as described in detailed by (Bundy et al. 2008), which fits idealised spectra based upon standards and estimates compound concentration. Each dataset derived from the Chenomx spectoral fits were normalized to the total sum of the metabolite concentrations to partially compensate for the differences in relative concentrations that occur between the high and low abundance metabolites prior to use of the data for further statistical analysis. Initially, univariate

analyses were used to determine the difference and significance of exposed concentrations against control replicates via fold-change and t-test ($P < 0.05$), respectively. For initial visualisation of the data, results were used to generate volcano plots for each of the test concentrations to demonstrate the relationship between fold change and significance. Subsequently, the correlation of candidate metabolites with nominal concentrations was employed to select for those that significantly varied relative to exposure using the Spearman's correlation method. To control for false discovery rate among metabolites found to significantly correlate with As exposure, the Benjamin-Hochberg procedure was implemented using a false discovery rate of 0.2 (Benjamini and Hochberg 1995) For increased stringency, the Bonferroni correction was implemented with a family error rate of 0.05 (Simes 1986). With respect to the strength of the Bonferroni correction in false discovery rate, observed outputs will only be considered as an indication of the most prolific variation associated with As exposure. The relative concentrations of metabolites were calculated relative to the average concentration of the most common metabolite. The overall relationship between candidate metabolites was described using principle component analysis (PCA), using XLSTAT (v.2012), as well as a range of associated univariate methods.

3.2.4 Transcriptomic Analyses

Transcriptomic analysis was performed upon adult *L. rubellus* that had been exposed to 0, 36 and 125 mg/kg As due to financial constraints. Total RNA was purified from each replicate sample which comprised a homogenised tissue sample from all surviving individuals within a test replicate. RNA extracts were made using a Qiagen RNeasy kit (Crawley, UK) as according to the manufacturer's protocol. Initially, approximately 10 mg of tissue was homogenised by repeated extrusion using a sterile needle and syringe. An on-column DNA digest conducted using the Qiagen RNase-free DNase set was implemented to ensure that there was no contaminating DNA, according to the manufacturer's instructions. RNA was then eluted from the column in 60 µl RNase free water. Samples were then quantified and assessed for quality using a Bioanalyser 2100 (Agilent Technologies, Stockport, UK). Although an RNA integrity number could not be determined as only a single major ribosomal RNA peak is definable (Winnebeck et al. 2010), RNA was not observably degraded, which is normally determined through an increase small fragments. Replicates from each concentration were pooled equally and kept at -20 °C before being sent for sequencing at the Genepool sequencing facility, UK. An Illumina truseq library (Illumina, Essex, UK) with an insert size of 150 bp was constructed for pooled samples from 0, 36 and 125 mg/kg As before sequencing on contemporary Illumina platforms.

Eighty million 100 bp paired-end reads from the 3 concentrations were processed prior to assembly of a *de novo* transcriptome using the reads trimmed relative to quality using the programme *sickle*, incorporating default settings (Joshi 2011). Reads were prepared for Velvet (1.1.04) using the programme shuffleSequences, before use of VelvetOptimiser to find the optimal parameter settings for Velvet deriving a k-mer length of 23 (Zerbino and Birney 2008). Trinity was also used to assemble raw reads using default parameters (Grabherr et al. 2011). Both assemblies selected for a minimum contig length of 200 bp. CAP3 was used (with default parameters) to produce a single, non-redundant assembly derived from both Oases and Trinity assemblies, while also working to reduce complexity having used pooled samples from whole tissue RNA extraction (Huang and Madan 1999).

Using the alignment programme Bowtie, 10M reads from each concentration were aligned to a database created using the programme Bowtie_build, providing an insight to the quality of each of the assemblies (Trapnell et al. 2012). The script, Contig_stats.pl (Sujai Kumar, the University of Edinburgh), was also implemented to derive a series of assembly statistics, including the number of assembled contigs, N50 length (the smallest contig size representing half of all contigs), number of contigs represented by N50, the mean length of contigs in the N50 and GC content.

Parts of the “Tuxedo” pipeline commonly used for transcriptomic analysis were implemented to signify candidate genes and pathways, with default parameters implemented throughout (Trapnell et al. 2012). Tophat was initially used to align raw reads to the consensus transcriptome before Cufflinks was used to assemble transcripts. Cuffmerge was used to merge together Cufflinks assemblies in preparation for Cuffdiff, which determined differentially expressed transcripts. Splice variants were not determined as a result of producing consensus sequences using CAP3, which robustly combines similar contigs such as those through determination of splice sites. Normalisation was addressed through calculating fragments per kilobase of transcript per million fragments mapped (FPKM), which accounts for variation in gene length and abundance. BLASTX was used to annotate the consensus transcriptome, incorporating matches with an e-value threshold of 10^{-5} from the *Caenorhabditis elegans* (WS220.66), *Drosophila melanogaster* (BDGP5.66), *Mus musculus* (NCBIM37.66) and *Homo sapiens* (GRCh37.66) peptide databases downloaded from NCBI (ftp://ftp.ensembl.org/pub/release-66/fasta/) (Altschul et al. 1997). Gene objects were also annotated using the NR database (ftp://ftp.ncbi.nih.gov/blast/db/) and annotated draft *L. rubellus* lineage B genome (unpublished). Anything over 2-fold expression was manually corrected with analogous protein accessions from the other databases for

recognition by Ingenuity Pathway Analysis (IPA), while human annotations were preferred. Concatenation of annotations across the databases was performed in Microsoft Access (2007). Highly differentiated candidate pathways that are influenced by As exposure were determined by IPA. As replicates were pooled, fold-change was used as a proxy for differential expression, with ± 2 fold change indicative of differentiation for inclusion in networks.

3.3 Results

3.3.1 Tissue Metal Concentrations

The total As concentrations were found to vary significantly relative to exposure ($F = 168.54$, $P < 0.001$) and are explicitly discussed in Chapter 2. Analysis of the concentrations of trace metals in the tissues of the exposed worms indicated no significant deviations relative to As exposure. For Fe, however, the measured concentration in tissues were found to significantly increase in earthworms exposed to 125 mg/kg As ($F = 5.60$, $P = 0.003$) (Figure 3.2). This indicates a modification of earthworm Fe handling as a consequence for high level As exposure.

3.3.2 Arsenic speciation

Arsenic species were individually transformed to realise proportions relative to different exposure concentrations, in light of being unable to correct for extraction efficiency using a CRM. Arsenic species in *L. rubellus* exposed to sodium arsenate demonstrated an increase in inorganic species relative to exposure concentrations. Generally, inorganic arsenicals account for over 92, 96 and 100% of As found in earthworms exposed to 0, 36 and 125 mg/kg, respectively. The As species detected in the control samples included both MMA and AsV, of which only the latter occurred in great quantities The presence of these two species represents the putative typical composition of As in earthworms that reside in soils containing background concentrations. Following As exposure, analysis of measured species confirmed that the proportion of the total measured As as AsV and AsIII differed significantly in relation to As exposure (ANOVA, $F = 9.00$, $P = 0.004$ for AsV and $F = 280.49$, $P < 0.001$ for AsIII, respectively). Post-hoc analysis using Tukey's test indicated that earthworms exposed to 125 mg/kg As had higher proportions of both of these metabolites than earthworms exposed to lower concentrations. Amongst methylated arsenicals, the proportion of As present as MMA was found to differ significantly, whereby higher relative levels were found among controls than in exposed samples ($F = 3.99$, $P = 0.047$). The cumulative percentage of

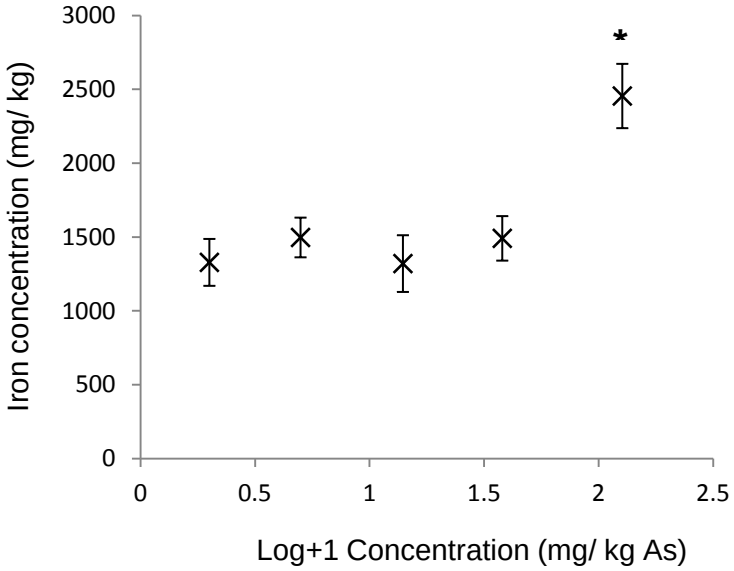


Figure 3.2 Mean Fe concentrations (+/- SE) in homogenised tissue of adult *L. rubellus* exposed to a range of As concentrations for 28 days. ***** = significant, where $P < 0.05$.

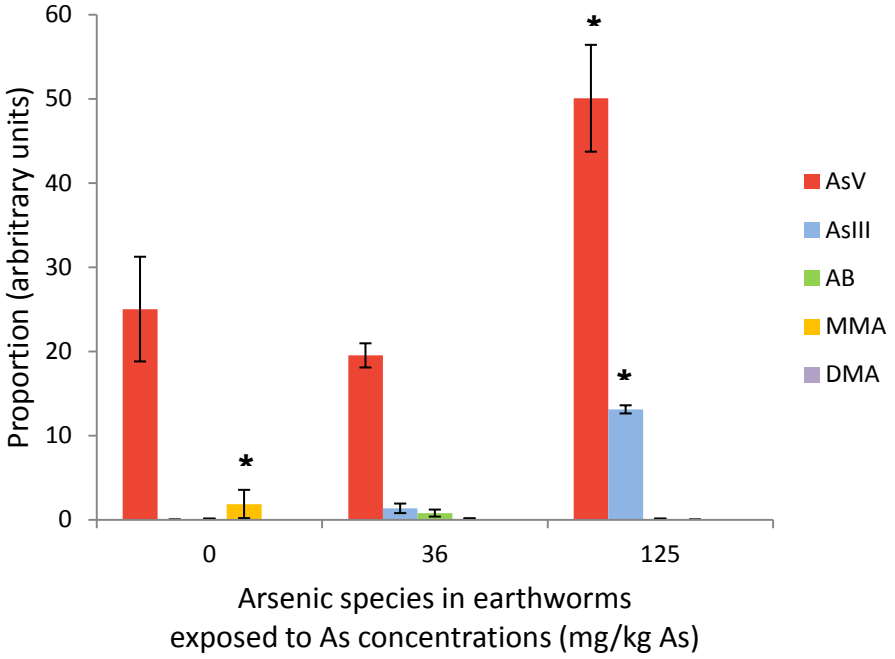


Figure 3.3 As species determined in adult *L. rubellus* by HPLC-ICP-MS following exposure to 0, 36 and 125 mg/kg As in soil for 28 days. Quantities of each species of each replicate were calculated as a proportion of the total signal (see 3.2.2)(+/- SE). ***** = species specific significance, where $P < 0.05$.

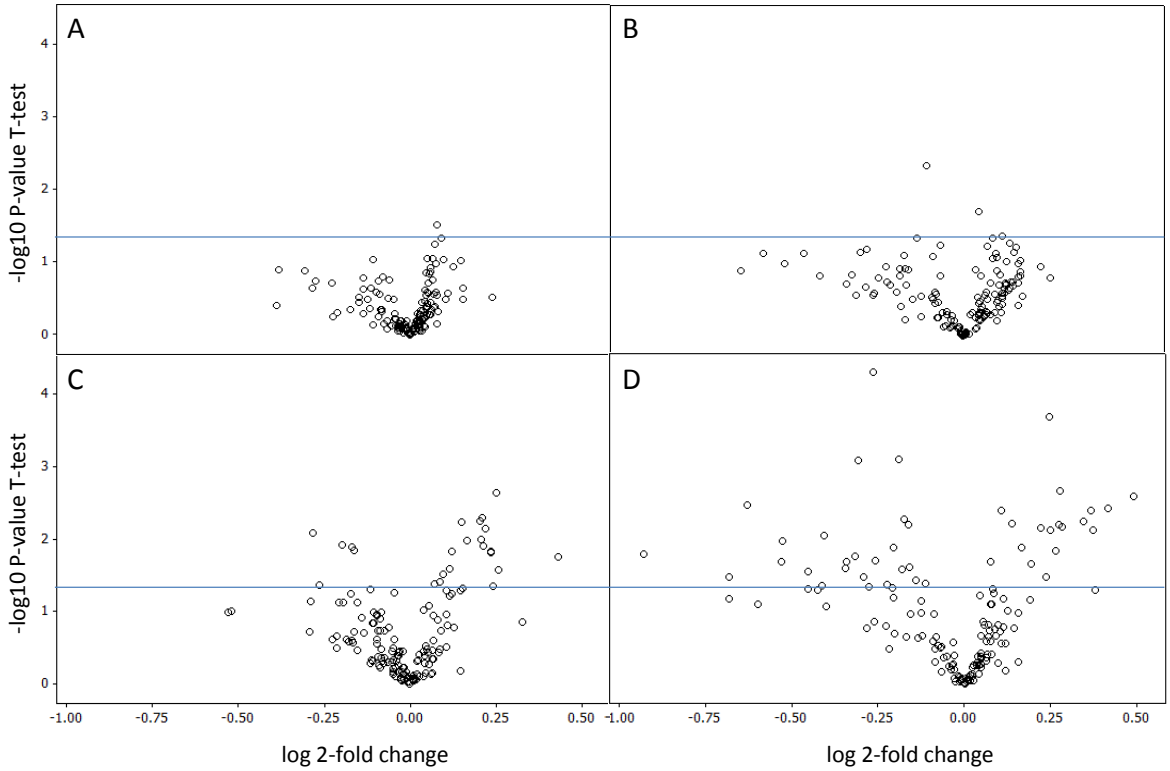


Figure 3.4. Volcano plots displaying fold change in the quantity of metabolites derived by ^{1}H -NMR against significance, comparing homogenised tissue from adult *L. rubellus* exposed to **A**) 3, **B**) 12, **C**) 36 and **D**) 125 mg/kg As against controls. α = 0.05, represented by the blue marker.

AsV and AsIII were found to correlate with increasing soil As exposure (r = 0.66, P = 0.007), while that of AB, MMA and DMA were found to be negatively correlated with As exposure; trends generally observable in Figure 3.3. As AsIII is produced through direct reduction of AsV, correlation with total tissue As is indicative that both tissue As and speciation data adequately represent As in earthworms even though they were not depurated.

3.3.3 Metabolomics

Increasing variation in the concentrations and number of metabolites, as well as increasing significance of these changes was associated with elevated levels of As exposure, as observable in (Figure 3.4). In total, 37 metabolites were recognised using Chenomx software, representing amino acids, sugars, nucleotides and metabolic intermediates among others. Relative abundance of metabolites was based upon the

concentration of aminoisobutyric acid (AIB), with only lombricine and 2-hexyl-5-ethylfuran-3-sulfonic acid (HEFS) at concentrations occurring above 10% of that of AIB.

The scores plot derived from PCA analysis (Figure 3.5) was used to demonstrate the relationship between replicates at each of the exposure concentrations. The first 2 principle components of the PCA explained 38.3% of the variation present within the metabolite data-set. The analysis was able to show that earthworms exposed to 36 and 125 mg/kg As could be clearly differentiated from those exposed to lower concentrations. Thus the samples that were exposed to the two highest concentrations of As are towards the bottom right of the scores plot shown in Figure 3.5. It is also clear that earthworms exposed to 3 and 12 mg/kg As cluster together with controls, reflecting trends seen in tissue As (Table 2.1). Analysis of the loadings plot indicates the contribution of metabolites recognised in Chenomx to the components (Figure 3.6). A number of metabolites are shown to be located in the same region of the plot that is associated with the highest exposure of samples observable within the scores plot. In particular a group of 5 metabolites are found in the corresponding bottom right region of the plot. These are glutamate, glycine, methionine, pyruvate and threonine,

Correlations of the relative levels of individual metabolites were assessed against the log concentration of As exposure to further establish the relationships between changes in As exposure (based on soil concentrations) and metabolite concentrations. Among the metabolites found to be affected relative to soil As concentration, amino acids methionine and glycine were both found to be most strongly correlated with As exposure. These two metabolites remained strongly correlated Bonferroni's correction which is widely recognised as the most stringent of available methods for multiple sample correction. Other metabolites found to be significantly correlated with As exposure concentrations are documented in Table 3.1 and include glutamate, leucine and phenylalanine, as well as histidine, scyllo-inositol and succinate, the latter three of which were each negatively correlated with As exposure. In all cases these three metabolites were significant following a Benjamini-Hochberg procedure, suggesting the validities of the relationship to As exposure observed in the data-set. For each of the metabolites, separate analysis by ANOVA indicated that only two that were found to correlate significantly with As exposure were found to differ significantly ($P < 0.05$). Earthworms exposed to 125 mg/kg As had significantly higher concentrations of glycine ($F = 8.66$, $P < 0.001$) and methionine ($F = 7.61$, $P = 0.001$) than those exposed to 0, 3 and12 mg/kg As. Variation in the concentration of these metabolites is observable in Figure 3.7.

Metabolite	Abbreviation [A]	Relative Concentration (%) [B]	Correlation (P)	P-value
ADP	ADP	0.67	-0.105	0.616
AMP	AMP	0.16	-0.288	0.167
Acetate	Ac	2.14	-0.130	0.535
Alanine	Ala	6.11	0.348	0.095
Aminoisobutyric acid	AIB	100	-0.017	0.936
Asparagine	Asn	1.87	0.173	0.407
Aspartate	Asp	2.17	0.175	0.401
Betaine	Bet	1.58	0.061	0.771
Choline	Cho	0.51	0.205	0.325
Fumarate	Fum	0.30	0.004	0.986
Glucose	Gluc	3.47	-0.163	0.435
Glutamate	Glu	3.97	0.497	0.017 ***
Glutamine	Gln	3.07	-0.002	0.993
Glycine	Gly	2.13	0.693	0.001 ***
HEFS	HEFS	10.79	0.332	0.112
Histidine	His	1.79	-0.412	0.048 ***
Isoleucine	Ile	2.11	0.177	0.396
Lactate	Lac	3.67	-0.346	0.098
Leucine	Leu	7.59	0.411	0.049 ***
Lombricine	Lom	93.12	-0.098	0.638
Lysine	Lys	3.33	0.237	0.255
Malate	Mal	8.37	-0.247	0.236
Maltose	Mlt	3.32	-0.263	0.207
Scyllo-Inositol	s-Ins	1.60	-0.579	0.006 ***
Methionine	Met	0.60	0.762	0.000 ***
N,N-Dimethylhistine	DMH	0.73	-0.369	0.077
Nicotinate	Nic	0.33	0.028	0.893
O-Phosphocholine	o-Pco	1.15	0.015	0.942
Phenylalanine	Phe	1.95	0.470	0.024 ***
Pyruvate	Pyr	0.19	0.347	0.096
Succinate	Succ	1.82	-0.526	0.012 ***
Threonine	Thr	3.85	0.184	0.379
Tryptophan	Trp	0.19	0.091	0.664
Tyrosine	Tyr	1.19	0.269	0.197
UDP-GlcNAc	UDP	0.36	-0.024	0.909
Valine	Val	3.29	0.240	0.251
myo-Inositol	m-Ins	1.21	-0.259	0.214

Table 3.1 List of metabolites recognised via ^{1}H -NMR spectra from homogenised tissue of adult *L. rubellus* following exposure to As. visible metabolites assigned in earthworm extracts. Pearson's correlation for metabolite concentrations against nominal As soil concentrations are shown, as is significance ($P < 0.05$) including after multiple testing correction, denoted by: * = ANOVA, * = Benjamini-Hochberg procedure, * = Bonferroni correction. [A] Corresponds to the labels in Figure 3.6. [B] Mean value, expressed as a percentage of the most concentrated metabolite (aminoisobutyric acid).

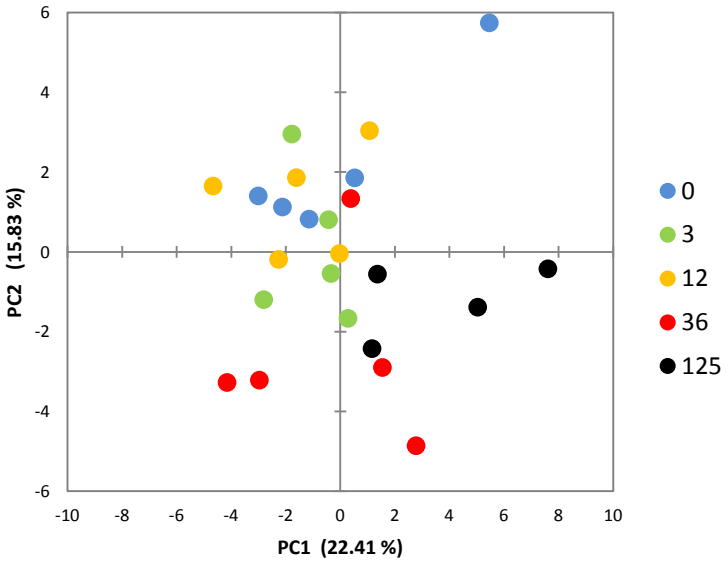


Figure 3.5 PCA scores plot of NMR spectral data showing relationship between metabolite profiles in adult *L. rubellus* following exposure to As for 28 days. Points represent replicates from a range of exposure concentrations.

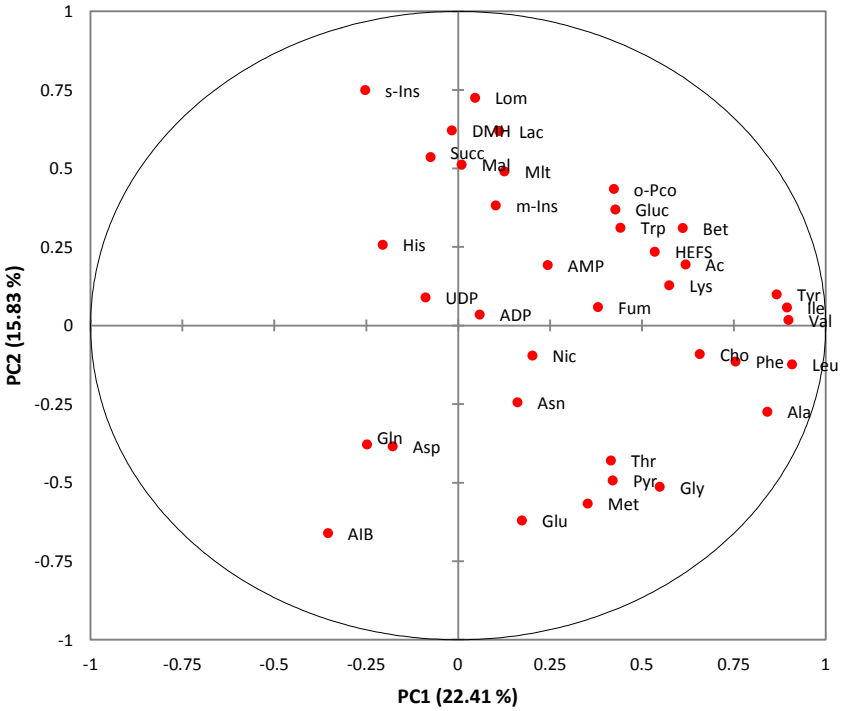


Figure 3.6 PCA loadings plot for individual metabolites that define profiles relating to exposure concentration of As in adult *L rubellus*. Metabolites are identifiable via abbreviations in Table 3.1.

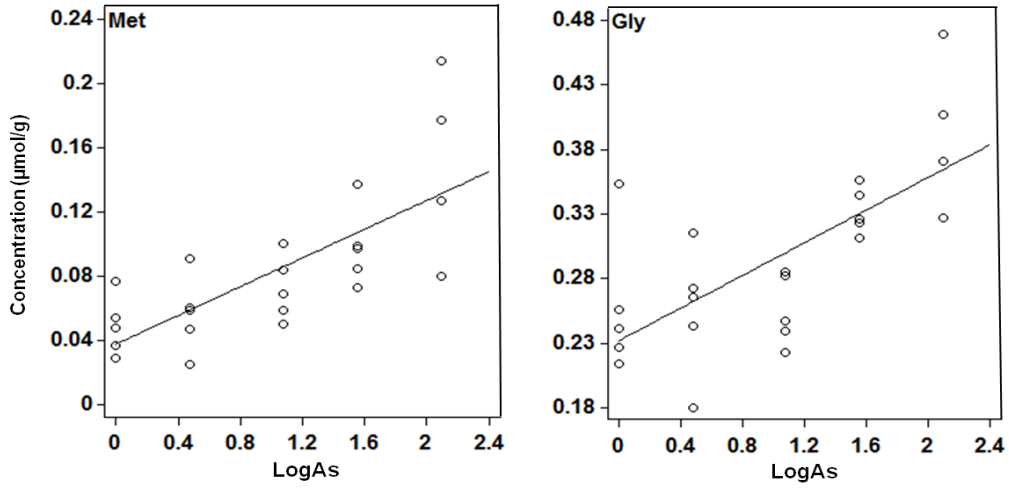


Figure 3.7 Effects of As exposure upon concentrations of methionine (**Met**) and glycine (**Gly**) (µmol/g) in adult *L. rubellus* exposed to As for 28 days, observed using Pearson's correlation. Points represent replicates from a range of exposure concentration

3.3.4 Transcriptomics

Total RNA from homogenised tissue representing replicates exposed to 0, 36 and 125 mg/kg As were sequenced to derive a *de novo* transcriptome for preliminary gene expression analysis. In a review by Martin and Wang (2011), the authors recognise that there are currently no recognised standards for assessing the quality of transcriptomic assemblies, although the authors do make a suggestion of several metrics, such as the number of correctly assembled bases, their completeness and level of contiguity, which all require well-established reference transcripts that vary in length. Although these metrics cannot all be adhered to in this instance, a series of more basic measures could be implemented for agreement with the set derived by Kumar and Blaxter (2010) who compared assembly methodology and included construction of a consensus transcriptome using CAP3, as in this investigation. These basic metrics are used to assess assemblies, while an analysis for completeness where the number of raw reads accounted for within the transcriptome is also used (Table 3.2).

Within their analysis, Kumar and Blaxter (2010) considered there to be ~30, 000 different transcripts with a mean length of 1.2 kb in the *C. elegans* transcriptome. These values represent a primitive metazoan which has seen an array of genome and transcriptome sequencing efforts, and therefore is a suitable proxy for the quality of the assemblies and the efficacy as representatives of the *L. rubellus* transcriptome. A large number of contigs greater than 200 bp in length, > 150,000, were produced through assembly using both the Oases and Trinity assembly tools. It can be hypothesised that

Table 3.2 Transcriptome assembly metrics for Oases, Trinity and Consensus *L. rubellus* transcriptomes, as well as EST library (Owen et al., 2008) and the *L. rubellus* draft genome. Included in metrics are a measure of alignment using lots of 10 M raw reads (100 bp in length), which contributed to transcriptome assembly as well as a measurement using the N50, which is the smallest contig size below which half of the assembly is represented.

	Num Contigs	**N50 for contigs > 200 bp**	**Contigs >200 bp in N50**	**Mean length for contigs > 200 bp**	**GC content**	**Mean of 10M reads aligning (n=3, ± SD)**
Oases	152828	1770	25841	1023.6	43.85	34.75 ± 1.21
Trinity	163282	1257	24957	735.6	43.6	62.67 ± 2.27
Consensus	34820	2524	7274	1689.4	43.9	70.69 ± 2.82
EST	8129	672	2985	611	43.6	15.58 ± 0.72
Genome	58254	597	11810	441.6	52.5	7.15 ± 0.53

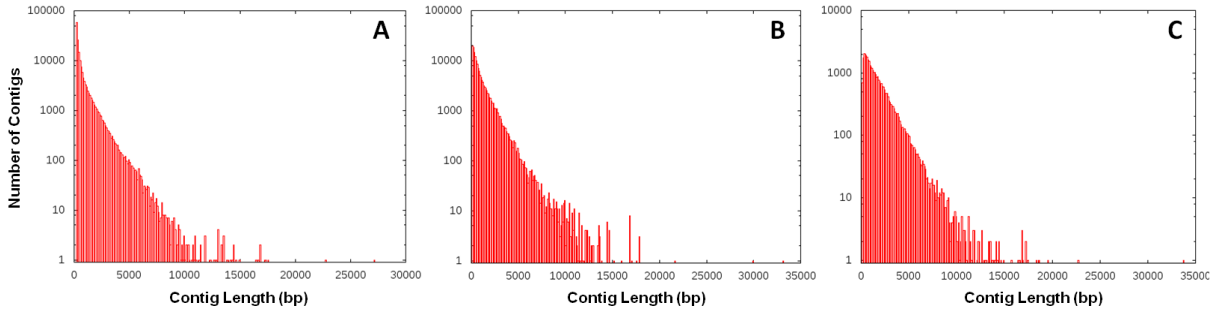


Figure 3.8 Contig length of gene bodies derived from **A**) Oases, **B**) Trinity and **C**) consensus *L. rubellus* transcriptome assemblies.

the identification of such a relatively large number of potential transcripts may be a result of the degree of genetic variation present within samples. This is likely a result of analysis of pooled samples constructed from the total RNA of homogenised tissue from 59 individuals represented within biological replicates (owing to mortality seen during the exposure).

Although the number of contigs is relatively similar between assemblies, Trinity produced many smaller contigs than Oases, as shown by the spread of contig sizes and frequencies, observable in Figure 3.8. The frequency of large chimeric sequences were relatively low in each of the assemblies, though Trinity assemblies remained smaller. The consensus transcriptome that was composed of the outputs derived from both Oases and Trinity assemblies and combined using CAP3 allowed for a stronger refinement of the transcriptome, removing the high levels of diversity and therefore providing a smaller number of contigs (~34 000), somewhat similar to the number observed in *C. elegans* (Kumar and Blaxter 2010). A similar proportion of reads were represented by the N50, with a greater mean contig length occurring (1.7 kb). Previously constructed *L. rubellus* expressed sequence tag (EST) and draft genome resources were compared to demonstrate their suitability for analysis and represent the extra variation within a dedicated transcriptome. Though it is clear that the EST library is not a complete reference, the suitability of the assembled transcriptome for transcriptomic analyses is less clear. As a measure of suitability of the *de novo* transcriptome for RNAseq analysis and comparison between biological treatments, raw sequence reads were aligned to each of the assemblies and additional genetic resources available for *L. rubellus*. This demonstrated that there is a relatively low capacity of the Oases assembly to integrate as much of the variation within observed by the Trinity. When combined using CAP3, over 70% of raw reads aligned to the assembly, relative to 16% in the EST library and only 7% in the genome. This is

representative of the power of combining assemblies, the non-coding sequences derived from total RNA, variation in splice sites and potentially of the large amounts of variation seen between the species that compose the *L. rubellus* complex in which the genome and transcriptome differ. Of the 34 223 gene objects that were ultimately derived from the outputs of the combined assembly, a total of 41% were annotated using BLASTX. This extent of BLAST homology is synonymous among other *de novo* transcriptomes (Barreto et al. 2011; Karatolos et al. 2011) and is very similar to the 43% of *L. rubellus* ESTs annotated in a study by Owen et al. (2008). This indicates that the transcriptome resource provides a reliable basis for the alignment and quantification of gene expression in the biological samples investigated.

3.3.5 Differential expression and pathway analysis

Differential expression analysis in this instance provides insight into gene regulation following As exposure. Of the gene objects in the consensus transcriptome, a total of 473 and 650 were found to deviate ± 2 fold relative to comparisons of controls with those exposed to 36 and 125 mg/kg As, respectively. Double the number of gene objects up-regulated were found to be down-regulated in both worms exposed to 36 and 125 mg/kg As. Of the differentially expressed transcripts, an average of 40% were BLAST annotated of which 49% were recognised within IPA. All annotated genes that were found to be differentially expressed, either positively or negatively, by two fold and the comparison of relative expression in earthworms exposed to 36 and 125 mg/kg As are detailed in Table 3.3. Annotated gene objects of interest that were not included in pathway analysis are included.

The analysis of the relationships of gene expression change and the link to effects in major biological pathways was undertaken using the IPA software system. This software can be used to establish networks of interaction, composed from affected biological pathways and processes as determined from compiled data derived from literature of key model organisms. Given the expected limitations association with both the extent of BLAST homology of earthworm transcripts as well as the representation of know homologous transcripts within the pathways included in the IPA, this analysis is a partial representation of the systemic nature of gene expression in *L. rubellus* following As exposure. Nonetheless, a number of relevant changes in key biological pathways were found to be associated with As exposure. The most prolific pathways among networks in earthworms exposed to 36 mg/kg As included those relating to cell development, growth and signalling (Figure 3.9 A). Transcriptional variation relative to controls was further eluded to as an affected molecular function, while genes responsible for managing oxidative stress and Fe metabolism are also up-regulated. A

A

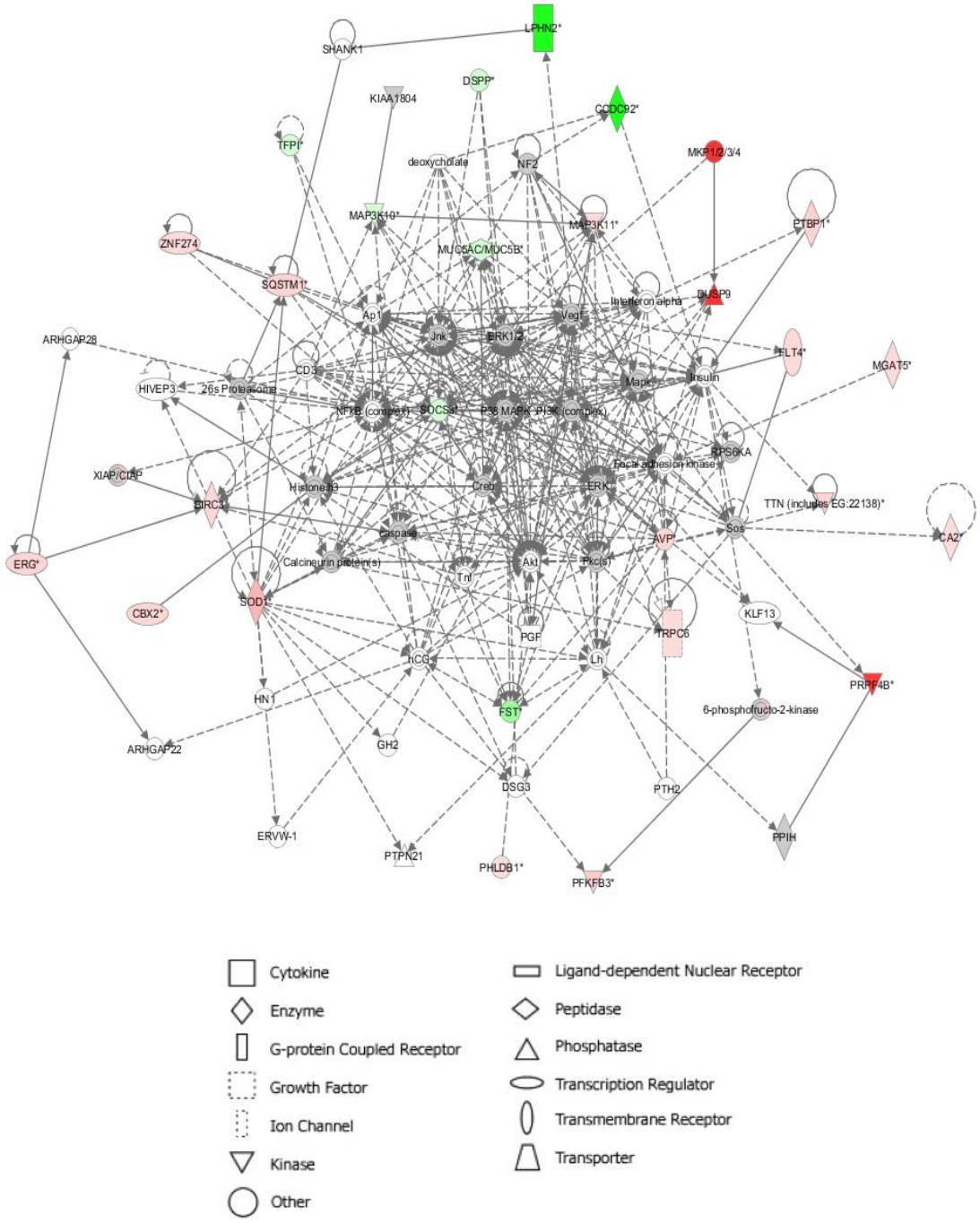


Figure 3.9 A comparison of differentially expressed genes (DEGs) occurring between adult *L. rubellus* exposed to 0 and 36 mg/kg As, reflecting pathways involved in cellular development, growth, proliferation and signalling. Elements shown in red are up-regulated and in green are down-regulated, while grey demonstrates that transcription is unchanged between treatments. White nodes reflect that a transcript closely involved in processes described by the network has not been found by IPA in the dataset. Shading is relative to the global extent of DEGs, however all DEGs are either above or below the threshold of 2 or -2, respectively. Symbols of DEGs are listed in Table 3.3. Solid and dashed edges define direct and indirect relationships between nodes, respectively.

B

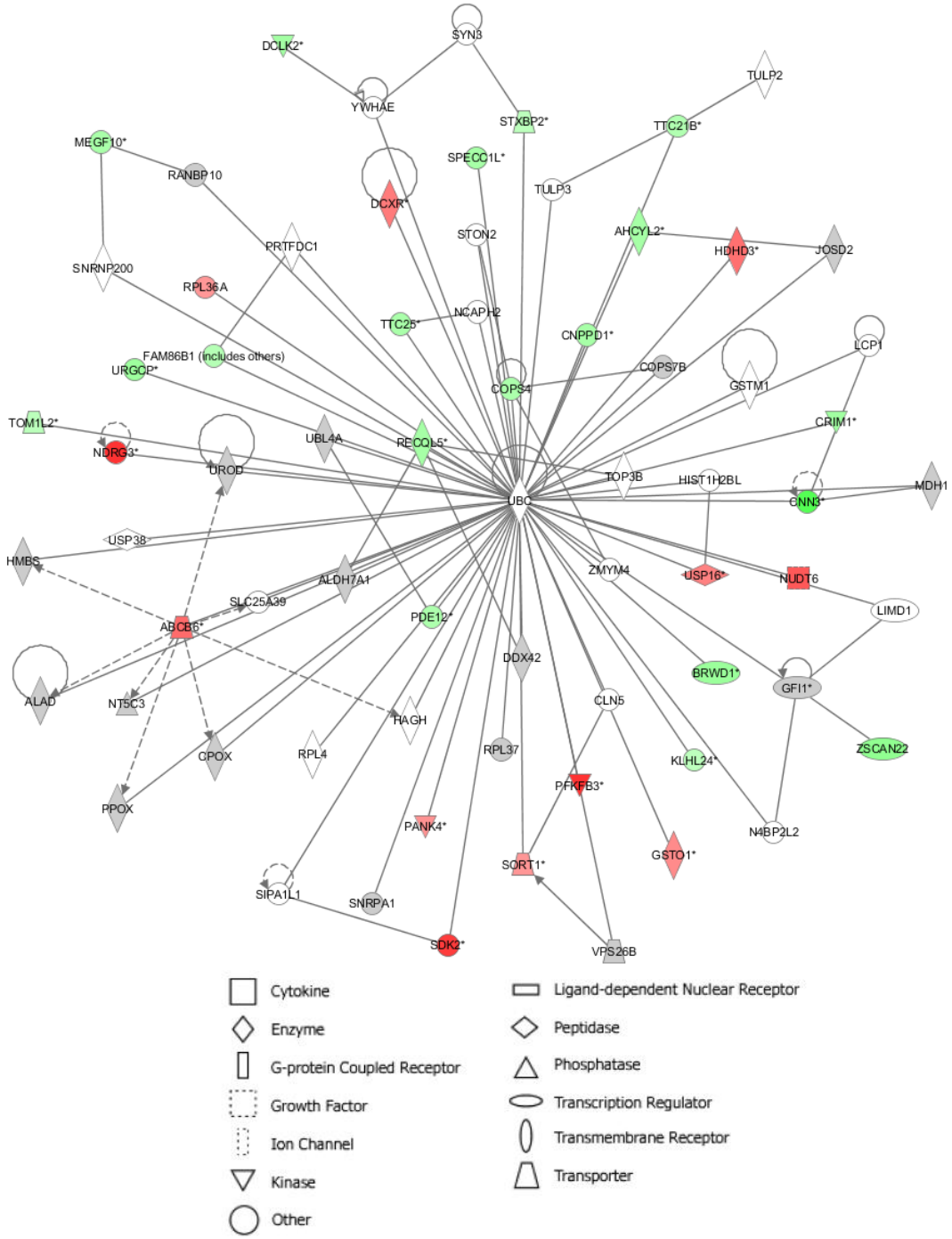


Figure 3.9 A comparison of differentially expressed genes (DEGs) occurring between adult *L. rubellus* exposed to 0 and 125 mg/kg As, reflecting pathways involved in hematological disease, metabolic disease and small molecule biochemistry. Elements shown in red are up-regulated and in green are down-regulated, while grey demonstrates that transcription is unchanged between treatments. White nodes reflect that a transcript closely involved in processes described by the network has not been found by IPA in the dataset. Shading is relative to the global extent of DEGs, however all DEGs are either above or below the threshold of 2 or -2, respectively. Symbols of DEGs are listed in Table 3.3. Solid and dashed edges define direct and indirect relationships between nodes, respectively.

C

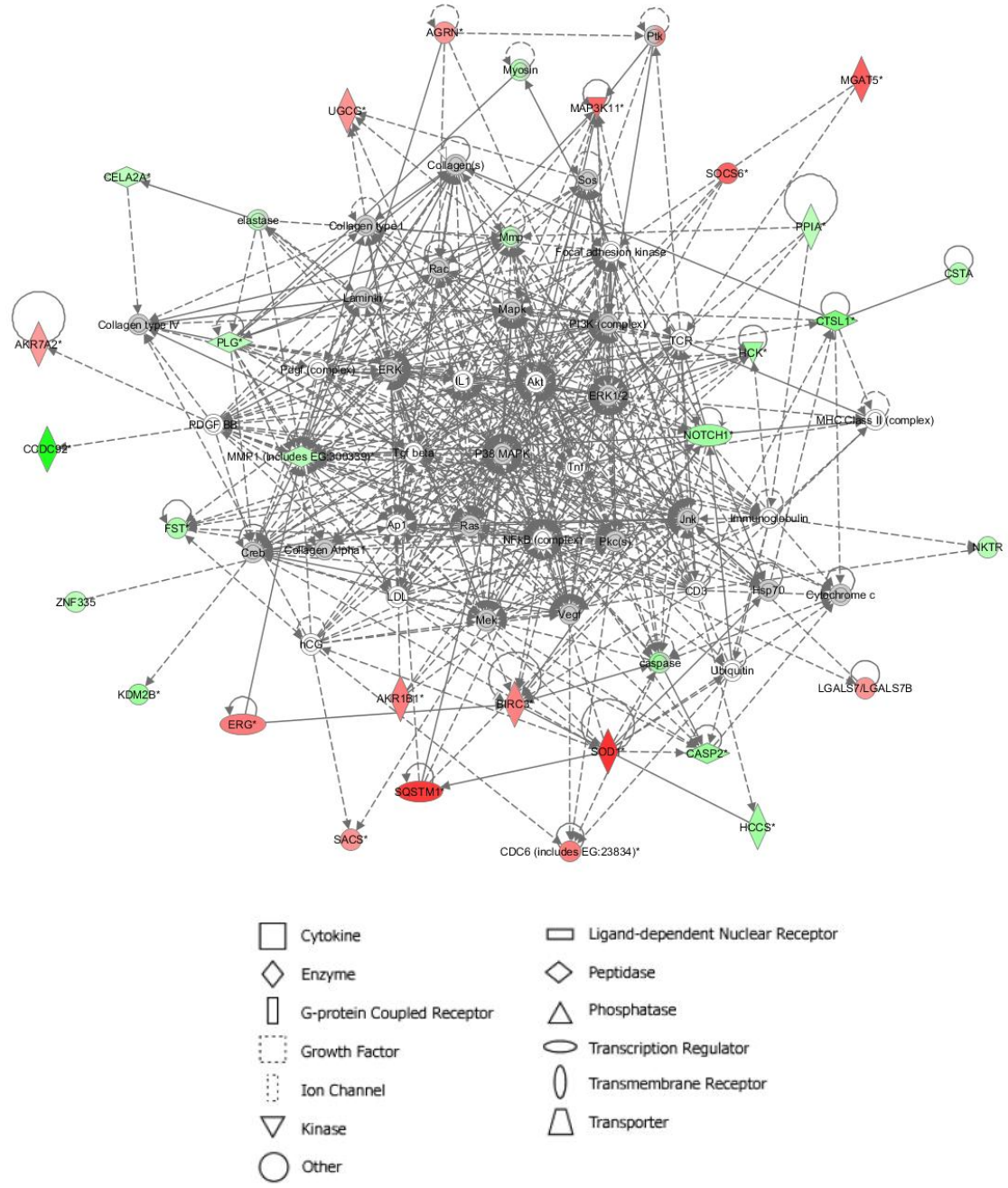


Figure 3.9 A comparison of differentially expressed genes (DEGs) occurring between adult *L. rubellus* exposed to 0 and 125 mg/kg As, reflecting pathways involved in cell death, tissue morphology and cancer. Elements shown in red are up-regulated and in green are down-regulated, while grey demonstrates that transcription is unchanged between treatments. White nodes reflect that a transcript closely involved in processes described by the network has not been found by IPA in the dataset. Shading is relative to the global extent of DEGs, however all DEGs are either above or below the threshold of 2 or -2, respectively. Symbols of DEGs are listed in Table 3.3. Solid and dashed edges define direct and indirect relationships between nodes, respectively.

Table 3.3 Differential gene expression values for adult *L. rubellus* exposed to As for 28 days, comparing those at 36 (**A**) and 125 (**B**) mg/kg As with controls. Fold-change beyond a threshold of above 2 or below -2 are shown and is alphabetized based upon the gene symbol referred to by IPA. Ensemble Gene ID, the location of activity within cells and gene product are also listed. Additional genes that are referred to specifically in the text but are not recognized by IPA, were manually located using BLASTX annotation and are highlighted in yellow at the end of the list.

A

Fold Change	ID	Symbol	Entrez Gene Name	Location	Type(s)
2.211	ENSG00000130377	ACSBG2	acyl-CoA synthetase bubblegum family member 2	Cytoplasm	enzyme
-2.014	ENSP00000423551	ADAMTS6	ADAM metallopeptidase with thrombospondin type 1 motif, 6	Extracellular Space	peptidase
2.534	ENSP00000369647	AVP	arginine vasopressin	Extracellular Space	other
2.605	ENSP00000436855	BIRC3	baculoviral IAP repeat containing 3	Cytoplasm	enzyme
-2.469	ENSP00000389014	C5orf42	chromosome 5 open reading frame 42	unknown	other
2.213	ENSG00000104267	CA2	carbonic anhydrase II	Cytoplasm	enzyme
2.569	ENSG00000173894	CBX2	chromobox homolog 2	Nucleus	transcription regulator
-13.35	ENSG00000119242	CCDC92	coiled-coil domain containing 92	Cytoplasm	enzyme
-2.11	ENSG00000133063	CHIT1	chitinase 1 (chitotriosidase)	Extracellular Space	enzyme
-2.91	ENSG00000060718	COL11A1	collagen, type XI, alpha 1	Extracellular Space	other
2.015	ENSP00000355718	DLL1	delta-like 1 (Drosophila)	Plasma Membrane	enzyme
-2.532	ENSP00000382213	DSPP	dentin sialophosphoprotein	Extracellular Space	other
14.173	ENSG00000130829	DUSP9	dual specificity phosphatase 9	Nucleus	phosphatase

2.699	ENSG00000157554	ERG	v-ets erythroblastosis virus E26 oncogene homolog (avian)	Nucleus	transcription regulator
2.382	ENSG00000037280	FLT4	fms-related tyrosine kinase 4	Plasma Membrane	transmembrane receptor
2.94	ENSP00000393884	FRRS1	ferric-chelate reductase 1	Plasma Membrane	transmembrane receptor
-5.981	ENSP00000426315	FST	follistatin	Extracellular Space	other
2.253	ENSG00000168301	KCTD6	potassium channel tetramerisation domain containing 6	unknown	ion channel
-15.864	ENSG00000117114	LPHN2	latrophilin 2	Plasma Membrane	G-protein coupled receptor
-2.272	ENSP00000253055	MAP3K10	mitogen-activated protein kinase kinase kinase 10	Cytoplasm	kinase
2.347	ENSP00000309597	MAP3K11	mitogen-activated protein kinase kinase kinase 11	Cytoplasm	kinase
2.41	ENSG00000152127	MGAT5	mannosyl (alpha-1,6-)-glycoprotein beta-1,6-N-acetyl-glucosaminyltransferase	Cytoplasm	enzyme
2.048	ENSG00000086504	MRLP28	mitochondrial ribosomal protein L28	Cytoplasm	other
-2.561	ENSP00000384815	MUC5AC/MUC5B	mucin 5AC oligomeric mucus/gel-forming	Extracellular Space	peptidase
6.534	ENSG00000101079	NDRG3	NDRG family member 3	Cytoplasm	other
2.537	ENSG00000149761	NUDT22	nudix (nucleoside diphosphate linked moiety X)-type motif 22	unknown	other
2.195	ENSG00000157881	PANK4	pantothenate kinase 4	Cytoplasm	kinase
3.148	ENSG00000170525	PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	Cytoplasm	kinase
2.184	ENSG00000019144	PHLDB1	pleckstrin homology-like domain, family 8, member 1	unknown	other
-3.354	ENSG00000141956	PRDM15	PR domain containing 15	Nucleus	other
13.053	ENSP00000337194	PRPF4B	PRP4 pre-mRNA processing factor 4 homolog B (yeast)	Nucleus	kinase
2.84	ENSG00000011304	PTBP1	polypyrimidine tract binding protein 1	Nucleus	enzyme
2.378	ENSG00000103769	RAB11A	RAB11A, member RAS oncogene family	Cytoplasm	enzyme
13.604	ENSP00000374410	RHBDF2	rhomboid S homolog 2 (Drosophila)	Cytoplasm	other
2.111	ENSP00000338218	RNF213	ring finger protein 213	Plasma Membrane	enzyme

2.103	ENSP00000404375	RPL36A	ribosomal protein L36a	Cytoplasm	other
-3.28	ENSG00000138326	RPS24	ribosomal protein S24	Cytoplasm	other
2.065	ENSP00000316909	SLC17A8	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 8	Plasma Membrane	transporter
2.493	ENSG00000160785	SLC25A44	solute carrier family 25, member 44	Cytoplasm	transporter
2.346	ENSP00000379326	SLC5A12	solute carrier family 5 (sodium/glucose cotransporter), member 12	unknown	transporter
-2.32	ENSP00000339834	SNRNP48	small nuclear ribonucleoprotein 48kDa (U11/U12)	Nucleus	other
-2.39	ENSP00000330341	SOCS3	suppressor of cytokine signaling 3	Cytoplasm	other
4.805	ENSP00000270142	SOD1	superoxide dismutase 1, soluble	Cytoplasm	enzyme
2.085	ENSP00000251809	SPAG1	sperm associated antigen 1	Cytoplasm	other
2.731	ENSG00000161011	SQSTM1	sequestosome 1	Cytoplasm	transcription regulator
-2.007	ENSP00000318233	STXBP2	syntaxin binding protein 2	Plasma Membrane	transporter
-2.462	ENSP00000233156	TFP1	tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibito	Extracellular Space	other
11.18	ENSP00000260129	TGS1	trimethylguanosine synthase 1	Nucleus	enzyme
-2.016	ENSP00000350630	TLL2	tolloid-like 2	Extracellular Space	peptidase
-3.37	ENSP00000261180	TRHDE	thyrotropin-releasing hormone degrading enzyme	Plasma Membrane	peptidase
2.208	ENSG00000137672	TRPC6	transient receptor potential cation channel, subfamily C, member 6	Plasma Membrane	ion channel
2.817	ENSP00000343764	TTN	titin	Cytoplasm	kinase
-3.245	ENSG00000119801	YPEL5	yippee-like 5 (Drosophila)	unknown	other
3.297	ENSG00000066827	ZFAT	zinc finger and AT hook domain containing	Nucleus	other
-3.246	ENSG00000131115	ZNF227	zinc finger protein 227	Nucleus	transcription regulator
2.336	ENSP00000409872	ZNF274	zinc finger protein 234	Nucleus	transcription regulator
-2.118	ENSP00000388311	ZNF845	zinc finger protein 845	unknown	other

B

Fold Change	ID	Symbol	Entrez Gene Name	Location	Type(s)
2.975	ENSP00000265316	ABCB6	ATP-binding cassette, sub-family B (MDR/TAP), member 6	Cytoplasm	Transporter
2.130	ENSP00000368678	AGRN	Agrin	Plasma Membrane	Other
-2.699	ENSG00000158467	AHCYL2	adenosylhomocysteinase-like 2	unknown	Enzyme
2.616	ENSP00000285930	AKR1B1	aldo-keto reductase family 1, member B1 (aldose reductase)	Cytoplasm	Enzyme
2.021	ENST00000235835	AKR7A2	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase)	Cytoplasm	Enzyme
-3.176	ENSG00000101280	ANGPT4	angiopoietin 4	Extracellular Space	growth factor
-2.510	ENSG00000029534	ANK1	ankyrin 1, erythrocytic	Plasma Membrane	Other
-2.326	ENSP00000405899	AQP12A/AQP12B	aquaporin 12B	Cytoplasm	Transporter
2.694	ENSP00000398375	ARFIP2	ADP-ribosylation factor interacting protein 2	Cytoplasm	Other
2.213	ENSP00000387099	ATHL1	ATH1, acid trehalase-like 1 (yeast)	unknown	Other
2.561	ENSP00000369647	AVP	arginine vasopressin	Extracellular Space	Other
2.613	ENSG00000172318	B3GALT1	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 1	Cytoplasm	Enzyme
-2.122	ENSG00000108641	B9D1	B9 protein domain 1	Extracellular Space	Other
2.502	ENSP00000436855	BIRC3	baculoviral IAP repeat containing 3	Cytoplasm	Enzyme
-2.621	ENSP00000350557	BRD3	bromodomain containing 3	Nucleus	Kinase
-3.048	ENSG00000185658	BRWD1	bromodomain and WD repeat domain containing 1	Nucleus	transcription regulator
-2.594	ENSP00000371488	C1QTNF3	C1q and tumor necrosis factor related protein 3	Extracellular Space	Other
-2.177	ENSP00000446225	C6orf62	chromosome 6 open reading frame 62	unknown	Other
2.728	ENSG00000104267	CA2	carbonic anhydrase II	Cytoplasm	Enzyme

-3.073	ENSG00000106144	CASP2	caspase 2, apoptosis-related cysteine peptidase	Cytoplasm	Peptidase
-2.214	ENSP00000309052	CATSPER1	cation channel, sperm associated 1	Plasma Membrane	ion channel
2.218	ENSG00000173894	CBX2	chromobox homolog 2	Nucleus	transcription regulator
-13.736	ENSG00000119242	CCDC92	coiled-coil domain containing 92	Cytoplasm	Enzyme
2.065	ENSP00000386378	CD207	CD207 molecule, langerin	Plasma Membrane	Other
2.607	ENSG00000094804	CDC6	cell division cycle 6 homolog (S. cerevisiae)	Nucleus	Other
-2.003	ENSP00000352639	CELA2A	chymotrypsin-like elastase family, member 2A	Extracellular Space	Peptidase
-4.417	ENSG00000133063	CHIT1	chitinase 1 (chitotriosidase)	Extracellular Space	Enzyme
-2.252	ENSP00000393719	CLEC17A	C-type lectin domain family 17, member A	Plasma Membrane	Other
-5.232	ENSP00000437665	CNN3	calponin 3, acidic	Cytoplasm	Other
-2.686	ENSG00000115649	CNPPD1	cyclin Pas1/PHO80 domain containing 1	unknown	Other
-3.330	ENSG00000060718	COL11A1	collagen, type XI, alpha 1	Extracellular Space	Other
-2.538	ENSG00000138663	COPS4	COP9 constitutive photomorphogenic homolog subunit 4 (Arabidopsis)	Cytoplasm	Other
-3.400	ENSG00000109472	CPE	carboxypeptidase E	Plasma Membrane	Peptidase
-2.858	ENSP00000280527	CRIM1	cysteine rich transmembrane BMP regulator 1 (chordin-like)	Extracellular Space	Kinase
-2.257	ENSP00000264474	CSTA	cystatin A (stefin A)	Cytoplasm	Other
-3.933	ENSP00000365061	CTSL1	cathepsin L1	Cytoplasm	Peptidase
-2.817	ENSG00000170390	DCLK2	doublecortin-like kinase 2	Cytoplasm	Kinase
2.647	ENSP00000303356	DCXR	dicarbonyl/L-xylulose reductase	Cytoplasm	Enzyme
-2.403	ENSP00000355718	DLL1	delta-like 1 (Drosophila)	Plasma Membrane	Enzyme
2.297	ENSG00000123992	DNPEP	aspartyl aminopeptidase	Cytoplasm	Peptidase
-3.143	ENSP00000381247	EFCAB9	EF-hand calcium binding domain 9	unknown	Other
-2.129	ENSP00000265162	ENPEP	glutamyl aminopeptidase (aminopeptidase A)	Plasma Membrane	Peptidase
2.632	ENSG00000157554	ERG	v-ets erythroblastosis virus E26 oncogene homolog (avian)	Nucleus	transcription regulator

-2.245	ENSG00000118894	FAM86B1	family with sequence similarity 86, member B1	unknown	Other
-3.192	ENSP00000371262	FGL1	fibrinogen-like 1	Extracellular Space	Other
-2.915	ENSP00000391438	FLOT1	flotillin 1	Plasma Membrane	Other
3.065	ENSP00000393884	FRRS1	ferric-chelate reductase 1	Plasma Membrane	transmembrane receptor
-2.456	ENSP00000426315	FST	Follistatin	Extracellular Space	Other
-2.934	ENSG00000114480	GBE1	glucan (1,4-alpha-), branching enzyme 1	Cytoplasm	Enzyme
-2.456	ENSG00000089154	GCN1L1	GCN1 general control of amino-acid synthesis 1-like 1 (yeast)	Cytoplasm	translation regulator
2.105	ENSG00000107249	GLIS3	GLIS family zinc finger 3	Nucleus	transcription regulator
2.006	ENSG00000023572	GLRX2	glutaredoxin 2	Cytoplasm	Enzyme
2.099	ENSG00000177885	GRB2	growth factor receptor-bound protein 2	Cytoplasm	Other
2.328	ENSP00000358727	GSTO1	glutathione S-transferase omega 1	Cytoplasm	Enzyme
-2.844	ENSG00000004961	HCCS	holocytochrome c synthase	Cytoplasm	Enzyme
-2.919	ENSP00000441169	HCK	hemopoietic cell kinase	Cytoplasm	Kinase
2.872	ENSG00000119431	HDHD3	haloacid dehalogenase-like hydrolase domain containing 3	Cytoplasm	Enzyme
2.315	ENSP00000356630	HECA	headcase homolog (Drosophila)	unknown	Other
2.331	ENSP00000295256	HPGDS	hematopoietic prostaglandin D synthase	Cytoplasm	Enzyme
2.223	ENSP00000409332	ISCA1	iron-sulfur cluster assembly 1 homolog (S. cerevisiae)	Cytoplasm	Other
-3.684	ENSG00000120457	KCNJ5	potassium inwardly-rectifying channel, subfamily J, member 5	Plasma Membrane	ion channel
-2.916	ENSP00000261824	KDM2B	lysine (K)-specific demethylase 2B	Nucleus	Other
-2.010	ENSP00000395012	KLHL24	kelch-like 24 (Drosophila)	unknown	Other
2.097	ENSP00000313571	LGALS7/LGALS7B	lectin, galactoside-binding, soluble, 7	Extracellular Space	Other
-2.182	ENSP00000409813	LRP2	low density lipoprotein receptor-related protein 2	Plasma Membrane	Transporter
2.113	ENSG00000104774	MAN2B1	mannosidase, alpha, class 2B, member 1	Cytoplasm	Enzyme
3.034	ENSP00000309597	MAP3K11	mitogen-activated protein kinase kinase kinase 11	Cytoplasm	Kinase

14.206	ENSG00000175471	MCTP1	multiple C2 domains, transmembrane 1	unknown	Other
-6.835	ENSG00000065833	ME1	malic enzyme 1, NADP(+)-dependent, cytosolic	Cytoplasm	Enzyme
-2.633	ENSP00000274473	MEGF10	multiple EGF-like-domains 10	Plasma Membrane	Other
3.210	ENSG00000152127	MGAT5	mannosyl (alpha-1,6-)-glycoprotein beta-1,6-N-acetyl-glucosaminyltransferase	Cytoplasm	Enzyme
-3.502	ENSG00000011143	MKS1	Meckel syndrome, type 1	Cytoplasm	Other
-2.180	ENSP00000322788	MMP1	matrix metallopeptidase 1 (interstitial collagenase)	Extracellular Space	Peptidase
-2.142	ENSG00000086504	MRPL28	mitochondrial ribosomal protein L28	Cytoplasm	Other
-2.268	ENSP00000446086	MYH7	myosin, heavy chain 7, cardiac muscle, beta	Cytoplasm	Enzyme
5.358	ENSG00000101079	NDRG3	NDRG family member 3	Cytoplasm	Other
-2.060	ENSP00000373227	NEK4	NIMA (never in mitosis gene a)-related kinase 4	Nucleus	Kinase
-2.183	ENSP00000232978	NKTR	natural killer-tumor recognition sequence	Plasma Membrane	Other
-2.911	ENSP00000277541	NOTCH1	notch 1	Plasma Membrane	transcription regulator
3.346	ENSP00000306070	NUDT6	nudix (nucleoside diphosphate linked moiety X)-type motif 6	Extracellular Space	growth factor
2.183	ENSG00000157881	PANK4	pantothenate kinase 4	Cytoplasm	Kinase
-2.495	ENSP00000309142	PDE12	phosphodiesterase 12	Cytoplasm	Other
4.109	ENSG00000170525	PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	Cytoplasm	Kinase
-2.264	ENSG00000165443	PHYHIPL	phytanoyl-CoA 2-hydroxylase interacting protein-like	Cytoplasm	Other
-2.005	ENSP00000308938	PLG	Plasminogen	Extracellular Space	Peptidase
-2.083	ENSP00000419425	PPIA	peptidylprolyl isomerase A (cyclophilin A)	Cytoplasm	Enzyme
-2.482	ENSG00000162997	PRORSD1P	prolyl-tRNA synthetase associated domain containing 1, pseudogene	unknown	Other
2.067	ENSG00000117360	PRPF3	PRP3 pre-mRNA processing factor 3 homolog (S. cerevisiae)	Nucleus	Other
-2.649	ENSG00000011304	PTBP1	polypyrimidine tract binding protein 1	Nucleus	Enzyme

-2.703	ENSG00000108469	RECQL5	RecQ protein-like 5	Nucleus	Enzyme
-2.011	ENSP00000379507	RORB	RAR-related orphan receptor B	Nucleus	ligand-dependent nuclear receptor
2.086	ENSP00000404375	RPL36A	ribosomal protein L36a	Cytoplasm	Other
-2.502	ENSP00000265814	RUNX1T1	runt-related transcription factor 1; translocated to, 1 (cyclin D-related)	Nucleus	transcription regulator
2.144	ENSP00000371735	SACS	spastic ataxia of Charlevoix-Saguenay (sacsin)	Nucleus	Other
2.211	ENSP00000262878	SAMHD1	SAM domain and HD domain 1	Nucleus	Enzyme
-13.888	ENSG00000248496	SCAND3	SCAN domain containing 3	unknown	Other
3.928	ENSP00000373378	SDK2	sidekick cell adhesion molecule 2	unknown	Other
-2.836	ENSG00000138674	SEC31A	SEC31 homolog A (S. cerevisiae)	Cytoplasm	Other
-4.035	ENSP00000392027	SF3A3	splicing factor 3a, subunit 3, 60kDa	Nucleus	Other
-2.142	ENSP00000404243	SLC12A8	solute carrier family 12 (potassium/chloride transporters), member 8	unknown	Transporter
3.096	ENSG00000160785	SLC25A44	solute carrier family 25, member 44	Cytoplasm	Transporter
-3.308	ENSP00000445340	SLC5A8	solute carrier family 5 (iodide transporter), member 8	Plasma Membrane	Transporter
3.116	ENSG00000170677	SOCS6	suppressor of cytokine signaling 6	Cytoplasm	Other
5.689	ENSP00000270142	SOD1	superoxide dismutase 1, soluble	Cytoplasm	Enzyme
-12.641	ENSP00000360292	SORBS1	sorbin and SH3 domain containing 1	Plasma Membrane	Other
2.103	ENSG00000134243	SORT1	sortilin 1	Cytoplasm	Transporter
-2.681	ENSG00000100014	SPECC1L	sperm antigen with calponin homology and coiled-coil domains 1-like	unknown	Other
3.956	ENSG00000161011	SQSTM1	sequestosome 1	Cytoplasm	transcription regulator
-2.027	ENSP00000318233	STXBP2	syntaxin binding protein 2	Plasma Membrane	Transporter
2.149	ENSP00000346440	TCF4	transcription factor 4	Nucleus	transcription regulator
-2.113	ENSP00000401917	TEPP	testis, prostate and placenta expressed	unknown	Other
2.104	ENSP00000335094	TMEM17	transmembrane protein 17	Extracellular Space	Other

-2.054	ENSP00000438621	TOM1L2	target of myb1-like 2 (chicken)	unknown	Transporter
-4.524	ENSP00000261180	TRHDE	thyrotropin-releasing hormone degrading enzyme	Plasma Membrane	Peptidase
-2.378	ENSG00000123607	TTC21B	tetratricopeptide repeat domain 21B	unknown	Other
-2.598	ENSG00000260703	TTC25	tetratricopeptide repeat domain 25	Cytoplasm	Other
2.443	ENSP00000343764	TTN	Titin	Cytoplasm	Kinase
-2.578	ENSP00000341289	TUBB4B	tubulin, beta 4B class IVb	Cytoplasm	Other
2.140	ENSG00000148154	UGCG	UDP-glucose ceramide glucosyltransferase	Cytoplasm	Enzyme
-2.948	ENSP00000402803	URGCP	upregulator of cell proliferation	Cytoplasm	Other
2.527	ENSG00000156256	USP16	ubiquitin specific peptidase 16	Cytoplasm	Peptidase
-2.100	ENSP00000397098	ZNF335	zinc finger protein 335	Nucleus	Other
2.867	ENSP00000441855	ZSCAN2	zinc finger and SCAN domain containing 2	Nucleus	transcription regulator
-3.370	ENSG00000182318	ZSCAN22	zinc finger and SCAN domain containing 22	Nucleus	transcription regulator
-3.056	-	-	glutamate receptor 1	-	-
-4.119	-	-	glutamate decarboxylase 1	-	-
2.035	-	TaPCS1	glutathione gamma-glutamylcysteinyltransferase	-	-
-2.122	-	-	locus notch homolog protein 3	-	-
-2.179	-	sno	strawberry notch	-	-

greater number of gene objects were differentially regulated in earthworms that had been exposed to 125 mg/kg As, providing a more comprehensive data-set for IPA analysis (Figure 3.9 B and C). Among the top networks composed in IPA, pathways synonymous with diseases and disorders feature, with signatures of drug and lipid metabolism, as well as small molecule biochemistry and transport occurring. Pathways involved in oxidative stress and Fe metabolism were again implicated with As exposure, while others associated with As metabolism and xenobiotic metabolism were also found to differentiate relative to controls. Variation seen in gene expression between earthworms exposed to 36 and 125 mg/kg As was not the focus of this preliminary work, however analysis of differentially expression is listed in Appendix A.

3.4 Discussion

The aim of this investigation was to better resolve As metabolism and toxicity in the earthworm, *L. rubellus*, with the aim of providing insight into As toxicity and also a basis for understanding potential mechanisms of resistance that may be relevant to natural populations. To achieve this, a range of -omic analyses have been used, focussing upon the effects of As exposure on gene expression and metabolism in the earthworm *L. rubellus*. Specifically, results were anchored using HPLC-ICP-MS to determine variation in As species, before ^{1}H-NMR based metabolic profiling and RNAseq were implemented to more fully derive functional biological deviation at transcriptomic and metabolomic levels. It was expected that As metabolism in *L. rubellus* is analogous to findings determined in other metazoans systems previously discussed.

Importantly, these analyses are underpinned by previously derived phenotypic data. The life-history effects of As on adult earthworms used to derive these data were previously used to derive the long-term effects of As exposure upon populations, outlined in Chapter 2. Briefly, it was found that the top concentration of 125 mg/kg As induced increased mortality and also strong detrimental effects on cocoon production, which when taken together would be anticipated to have impact on the long-term stability of earthworm populations exposed to this exposure concentrations. Although no significant levels of variation were seen in adults exposed to 36 mg/kg As, growth and maturation was significantly affected in juveniles. This objectified both earthworms at 36 and 125 mg/kg As as concentrations of interest for molecular analyses, representing levels at which the effects of exposure may be observable. Total tissue As concentrations were significantly elevated in earthworms exposed to both 36 and 125 mg/kg As.

3.4.1 Arsenic Speciation

As speciation was used to assess the changes in major As metabolite pools across each of the exposure concentrations, reflecting the shifting concentrations of varieties to provide insight into the probable mechanisms for metabolism while allowing for a demonstration of relevance to other studies.

Previous work by Button et al. (2009), focussed upon As speciation among earthworms collected from sites located on the former As mine at Devon Great Consols (DGC), where populations of *L. rubellus* have been inferred to have a genetically based adaptation to As (Langdon et al. 2009). It was found that inorganic As was dominant and increases relative to soil As concentrations, while the levels of organic species remained consistently low. Within this investigation, it was confirm that following exposure to AsV, inorganic As species make up the majority of As found in naive earthworms and increased with soil exposure. AsV is found in control worms reflecting the presence of background levels of As in the unspiked soils. This is confirmed among other investigations looking at compartmentalisation within *L. rubellus*, finding that background As is assimilated and stored in inaccessible cellular regions, such as in cell membranes (Chapter 4). There is slightly less AsV in earthworms exposed to 36 mg/kg As, however, there is more AsIII. This may represent the mobilisation of detoxification strategies. Further evidence for this is obvious at the top concentration, where AsV at higher concentrations, while AsIII is also much higher than in worms exposed to lower concentrations. Both are significantly correlated with total tissue As. The negative correlation of other species enforces association with organic ligands such as GSH and is likely to not be metabolised via the Challenger pathway.

Previous suggestions that AB is a mechanism for detoxification (Langdon et al. 2003a) have since been reconsidered (Button et al. 2012), with evidence in marine biota demonstrating that accumulation is relative to metabolism by microorganisms (Ritchie et al. 2004). AB was observed to occur at very low levels among earthworms exposed to 36 mg/kg As, but not among other concentrations. This low level occurrence could be related to bacterial activity in soils. Unpublished data by Pass et al. has demonstrated a relationship between the community structure of the gut microbiota in natural populations of *L. rubellus* relative to soil As concentrations. The use of sterilised loam as the major constituent in test soils and oven dried manure is suggestive of reduced microbial activity among exposures and this may in part explain the relatively low level of AB that were detected within exposure earthworms. Low levels of methylated As were also observed, with this metabolite only forming a significant component of the total range of As species among controls. The low occurrence of soil

As and absence of a relationship of methylated arsenicals with exposure concentrations suggests that these species are not significantly associated with an active metabolic mechanism in earthworms. This suggests that our molecular analyses are relevant to other investigations using As speciation, whereby similar mechanisms of metabolism are likely to occur.

3.4.2 Metabolomics

Metabolomic analysis provided a direct measure of changes in the concentrations of quantified small metabolites and so can derive a more insightful view of how the organisms mitigate As toxicity. As a result, a variety of compounds that reflect mechanisms under regulation in response to As could be detected. The shared role of many metabolites in several pathways means that it is not always possible to directly link change in an individual metabolite unequivocally to the induction or suppression of individual metabolic pathways. However, the metabolome can represent a discrete temporal view of biological activity through reduced dimensionality to understand the relationships between biological samples and the major metabolites that are associated with variation.

No previous work has focussed the scope of metabolomics upon As metabolism and toxicity, though previous work has used ^{1}H-NMR in tandem with transcriptomic analyses to understand the effects of Cu toxicity in *L. rubellus* (Bundy et al. 2008). Within this investigation, it was possible to indentify a number of metabolites that were both positively and negatively correlated with As exposure. Of these, only glycine and methionine were strongly associated with exposure using both correlation and analysis of variance analysis. These two metabolites thus appear to have key, dose-dependent roles in the cellular management of As. The two are already synonymous with mechanisms of involved in As metabolism and are known to play pivotal roles in the folate and methylation cycles, as well as the closely affiliated pathway for the production of GSH shown in Figure 3.10 (Hayakawa et al. 2005).

It is known that methionine is derived from homocysteine and N5-methyl-tetrahydrofolate via methionine synthase, synonymously producing tetrahydrofolate. Tetrahydrofolate, in turn, is converted back to N5-methyl-tetrahydrofolate by means of cystathionine B-synthase and serine, which is converted to glycine. Donation of methyl-groups is considered to occur via S-adenosyl methionine, derived from methionine and is also associated with DNA methylation. Therefore, it is potentially possible that variation in observed methionine concentrations may be consistent with large scale changes in DNA methylation profiles such suggested in both toxicological and

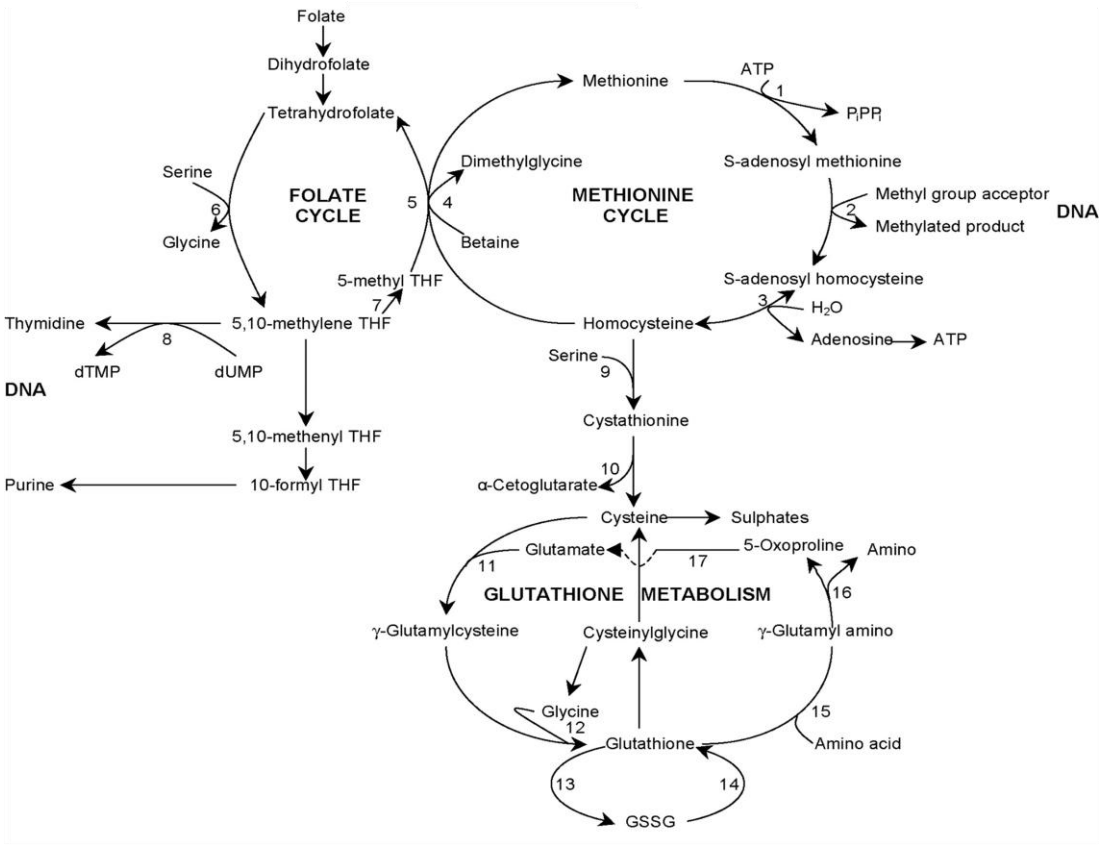


Figure 3.10 The folate and methionine cycles, including glutathione metabolism, adapted from (Ebisch et al. 2007). 1. Methionine-adenosyltransferase; 2. Methyltransferase; 3. S-adenosylhomocysteine hydrolase; 4. Betaine-homocysteine-methyltransferase (BHMT) (zinc dependent); 5. Methionine-synthase (vitamin B12 and zinc dependent); 6. Serineoxidase; 7. Methylenetetrahydrofolate reductase (MTHFR; vitamin B2 dependent); 8. Thymidylate-synthase; 9. Cystathionine-β-synthase (vitamin B6 dependent); 10. Gamma (γ)-Cystathionase (vitamin B12 dependent); 11. γ-Glutamyl-cysteine-synthase; 12. Glutathione-synthase; 13. Glutathione peroxidase; 14. glutathione disulphide (GSSG) reductase; 15. γ-Glutamyl transpeptidase; 16. γ-Glutamyl cyclotransferase; 17. 5-Oxoprolinase. dTMP, deoxythymidine monophosphate; dUMP, deoxyuridine monophosphate; GSSG, glutathione disulphide; Pi, orthophosphate; PPi, pyrophosphate; THF, tetrahydrofolate.

biological responses to As exposure (Reichard and Puga 2010). Homocysteine is also directly associated with the production of GSH, which directly associates with glycine by means of glutathione synthase, though this is thought to be shed during arsenolysis of GSH (Nemeti et al. 2012). The significant increase in both methionine and glycine among earthworms at the top exposure give good cause to implicate regulation by classically associated pathways following sublethal exposure to As (Figure 3.7).

Of the other metabolites significantly affected by As exposure, the changes seen in endogenous glutamate concentrations is consistent with the roles of methionine and glycine as a derivative of GSH catabolism and composite of the precursor to GSH, gamma glutamylcysteine. Both leucine and phenylalanine are also linked to these metabolites. Specifically, leucine is a derived from the conversion of L-glutamate to 2-oxoglutarate, while L-methionine can degrade to L-phenylalanine, normally produced via the addition of keto-phenylpyruvate to L-glutamate. Phenylalanine and glutamate are also closely associated with alpha-ketogluterate, a precursor of succinyl-CoA in the Krebs cycle. Condensation of glycine and succinyl-CoA by ALA synthase, involved in heme synthesis, directly precedes succinate production and could suggest why it is negatively correlated with As exposure. ADP-arsenate is also synthesised in the presence of succinate and provides further evidence for lowering concentrations relative to As exposure (Gresser 1981). It may also be worth noting that phenylalanine is the direct precursor to phenylethylamine, which is a stimulant associated with neuromodulation and transmission, while glycine and glutamate are important co-agonists for N-methyl D-aspartate (NMDA) receptors. This may further implicate that As may affect cell signalling and energy, already associated by other investigations (Kumagai and Sumi 2007; Luo et al. 2012; Partridge et al. 2007; Szinicz and Forth 1988) and could link Scyllo-inositol, which is related to signal transduction and limits the formation of plaques associated with Alzheimer's disease in mice, though remains poorly characterised (Fenili et al. 2007).

3.4.3 Transcriptome

The *L. rubellus* transcriptome, derived from 80 M short-length sequence read generated in the current study, provides a valuable resource for future ecotoxicological and evolutionary biology studies and will greatly benefit the genome sequencing effort by supplementing annotation (Stuerzenbaum et al. 2009). The range of statistics derived suggest that the combination of the outputs from the Oases and Trinity assemblies provides a robust consensus as they align with those derived by Kumar and Blaxter (2010) for *C. elegans*. Further, the alignment of raw reads demonstrates that the reference transcriptome can provide an excellent basis for preliminary transcriptomic analysis, leveraging further evidence to distinguish mechanisms involved in As metabolism and toxicity. Due to the design of the experiment and limitations upon the number of samples that could be analysed, the conclusions that can be drawn are preliminary. Thus, the transcriptome has been implemented provisionally in support of potential mechanisms established among other endpoints, to resolve management of As. There were a large range of gene bodies identified that weren't annotated, highlighting the lack of resources applicable to non-model

organisms. The glut of sequencing studies will better inform metabolic models in future, while tools exist to better aid hypothetical annotation based upon sequence, such tools lie outside of the remit of this specific investigation.

Using the *de novo* transcriptome, a preliminary analysis of gene expression variation relative to As exposure was conducted. The exposure concentrations selected for transcriptomic analysis provide insight into the effects of As among more readily observable life-history parameters and can elaborate upon metabolic endpoints. At the intermediate concentration of 36 mg/kg As, relatively subtle effects that remain ambiguous among other molecular and life-cycle endpoints were evident. Thus, the effects of low level exposure are potentially better reflected at the transcriptional level. At 36 mg/kg, up-regulation of both the generalised antioxidative enzyme, SOD1, and SQSTRM1, a transcript which encodes a protein responsible for the ubiquitination of proteins, are likely indicators of increased levels of cellular AsIII and the resulting oxidative stress and associated damage. The networks derived from IPA demonstrate the degree of variation in gene-regulation and reflect biological activity relative to exposure, with the most prominent networks from 36 mg/kg As have much lower content than those exposed to 125 mg/kg (Figure 3.9 A, B and C).

Significant reduction of cocoons produced by earthworms exposed to 125 mg/kg As as a result of toxicity is reflected through a clearer and stronger signal of As induced toxicity using transcriptomic analysis. In terms of viable mechanisms for reducing As toxicity, preventing entry into the cell remains highly effective in reducing toxicity. Previously, increased expression of an aquaporin in mouse heptaocytes has been found to enhance the uptake of AsIII (Shinkai et al. 2009). At the highest concentration, down-regulation of an aquaporin, commonly used to transport water and neutral small molecules into the cell, is observed. This may highlight AQP12 as a candidate for the potential regulation of arsenic transport.

Upon entry of the cell, the role of genes relating to the methionine cycle (and by association, the folate cycle) as well as GSH production are of considerable interest following As exposure (Figure 3.10). Within the 125 mg/kg sample, there were no significant differences in the expression of genes relating to methionine adenosyltransferase (responsible for the conversion of methionine to SAM), glutathione synthase, reductase or peroxidise. Additionally serine hydroxymethyltransferase, which is responsible for the conversion of tetrahydrofolate to 5,10 tetrahydrofolate, did not show evidence of variation in expression, as is also the case for serine oxidase, which

is responsible for the conversion of tetrahydrofolate to 5N-tetrahydrofolate. Finally all genes annotated with function in the folate cycle were unchanged.

Although many genes associated with the methionine and folate cycles and GSH production undifferentiated, a range of genes within these pathways were up-regulated, implying a pathway specific response to As exposure. A pathway for glutathione metabolism and genes considered representative among model organisms derived from IPA is shown in Appendix B. Notably, glutaredoxin 2 (GLRX2) is a GSH-dependent oxidoreductase (also known as arsenate reductase), which facilitates the reduction of AsV to AsIII and has been determined to play a central role in the response to both redox signals and oxidative stress (Beer et al. 2004). GSTO1, which has a known glutaredoxin activity and is also known as glutathione: dehydroascorbate reductase as well as monomethylarsenate reductase is also up-regulated (Mukherjee et al. 2006). This may reflect other functions, including the reparative mechanisms of glutaredoxins for proteins damaged through oxidative stress (Yoshitake et al. 1994), and their potential to offer reductive mechanisms for not only the oxidation, but also the reduction of GSH (Washburn and Wells 1999). They have also been shown to bind and deliver Fe-sulphur clusters and play an important role in Fe metabolism (Lillig et al. 2008). Another up-regulated gene within this set of associated pathways include adenosyl homocysteinase (AHCYL2), which is responsible for the conversion of S-adenosyl-homocysteine to homocysteine, the precursor of methionine.

In addition, a gene encoding glutathione gamma-glutamylcysteinyltransferase (TaPCS1), considered analogous to PC1 (a phytochelatin synthase), was up-regulated in earthworms exposed to 125 mg/kg As (Clemens et al. 1999). PCs are well established as a means of chelation for AsIII (Song et al. 2010) and up-regulation can be putatively linked to the dose dependent increase in glycine relative to As exposure as a product of its activity. Further analysis of tissue from adults analysed here, using an HPLC system coupled to an API 2000 Q-TRAP mass spectrometer, has shown that PCs within tissues derived from earthworms in this experiment demonstrate a dose response, increasing relative to As exposure (Figure 3.11) (Garcia-Perez et al. In preparation). The occurrence of PCs found within the transcriptome and up-regulation of one specifically relative to As exposure confers an important development in evolutionary biology whereby the occurrence and use of phytochelatins varies among taxa in which it is documented (Clemens 2006). These results suggest *L. rubellus*, as a candidate for further analysis of PCs in metazoans, along with *Eisenia fetida* (Brulle et al. 2008).

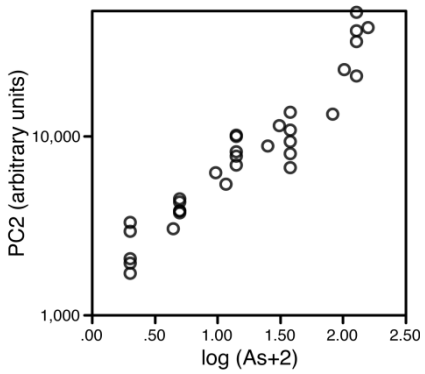


Figure 3.11 The molecular response of adult *L. rubellus* from this analysis to As through observation of Phytochelatin (PC_2) concentrationsplotted against soil arsenic concentration. Reproduced and modified with permission (Garcia-Perez et al. In preparation).

Observation that tissue Fe concentrations were significantly increased in worms exposed to the highest As concentrations reflect potential changes in the homeostasis of this essential metal. In relation to this, FRRS1 (ferric-chelate reductase 1) was found to be up-regulated in both 36 mg/kg As and 125 mg/kg As. The gene encodes a protein that is responsible for reducing Fe^{3+} to Fe^{2+}. The Fe-sulfur cluster assembly 1 homolog (ISCA1) is also up-regulated in earthworms at the highest concentration. This enzyme is responsible for the incorporation of Fe^{2+} into Fe-containing proteins in humans, and so supports the basis for the assembly of several important enzymes involved in a number of key cellular processes, including the electron transport chain, Krebs cycle and heme biosynthesis, as well as glutaredoxin activity (Rouault and Tong 2005) (Lillig et al. 2005). Fe^{2+} has also been found to act as a cofactor in the formation of alanine and is potentially associated with the depression of succinate relative to As exposure. PHYHIPL (phytanoyl-CoA 2-hydroxylase interacting protein-like) also includes an Fe^{2+} co-factor and is also found to be down-regulated. This enzyme interacts with phytanoyl-CoA dioxygenose and alpha-ketoglutarate in the Krebs cycle as a co-substrate, derived from L-glutamate with leucine. It works to take alpha-ketoglutarate, O2 and phytanoyl-CoA to produce 2-hydroxyphytanoyl-CoA, CO2 and succinate. Together these changes in Fe containing enzymes have the potential to explain the observed decreases of succinate in relation to As exposure (Mukherji et al. 2001).

Iron has been further implicated in As metabolism and also GSH through its role in Fe-sulphur cluster formation (Kumar et al. 2011; Lam et al. 2006; Paulose et al. 2010). At a physiological level, Meunier et al. (2010) focussed upon bioavailability of As in

sediment via a semi-standardised physiologically-based extraction test method that simulates gastro-intestinal absorption. In their analysis, As bioaccessibility was negatively correlated with total Fe. Further, the authors posited that glycine may increase As bioaccessibility through acting as a ligand for Fe. Fe^{2+} has been found to bind directly to As (Thoral et al. 2005) and investigations have been made to assess the capability of Fe for chelating As in drinking water (Nieto-Delgado and Rangel-Mendez 2012), which is a major issue in south central Asia. Iron may even have a role in the adaptation of earthworms at DGC, who present a yellowing pigmentation that may be a consequence of bilirubin and is derived from normal heme metabolism (Piearce et al. 2002), potentially linked to haemoglobin with which As binds (Winski and Carter 1995). Fe, GSH and glycine should therefore be considered as prime candidates for further investigation.

In response to oxidative stress, which is a known pathology associated with As exposure, transcription rates of genes capable of limiting production of reactive oxygen species is a vital component of the cellular defence system. Such systems are frequently supported by protein-handling enzymes capable of repairing or recycling proteins damaged by oxidative stress. Supplementary to the role of GSH, the aldo-keto reductases, AKR7A2 and AKR1B1 are seen to be up-regulated following As exposure and represent pathways classically associated with arsenic toxicity and the generalised response to reactive oxygen species. SQSTM1 was found to progressively increase in expression relative to As exposure, likely reflecting the levels of AsIII occurring relative to exposure. SQSTM1 is responsible for binding ubiquitin, promoting protein aggregation and engulfment by autophagosomes for degradation (Kerscher et al. 2006). This is now synonymous with arsenic exposure, as demonstrated in other investigations in both plants and animals (Lam et al. 2006; Paulose et al. 2010). Deubiquitination is also moderated among H2A histone proteins at the highest concentration by USP16, cell cycle progression and gene expression, while marking itself as an excellent candidate for further study of epigenetic variation in this species (Joo et al. 2007).

As well as affecting cellular signal transduction cascades (Kumagai and Sumi 2007), AsIII has also been found to inhibit the notch pathway (Zhen et al. 2010), a highly conserved cell signalling system. Genes associated with this pathway are down-regulated at 125 mg/kg As including NOTCH1, sno and a NOTCH3 homologue, the latter of which is also reduced in expression among worms exposed to 36 mg/kg As. There are also 2 genes that differ at the top concentration relative to controls in relation to the role of glutamate as a neurotransmitter; a GABA enzyme (glutamate

decarboxylase) and a glutamate receptor are both down-regulated. This may be evidence for a reaction to varying levels of glutamate which has already been implicated with methionine, glycine and the Krebs cycle. This, in conjunction with previously mentioned affectors of the Krebs cycle may support the suggestion of reduced cocoon production observed among adult earthworms exposed to 125 mg/kg As (Chapter 2).

3.5 Conclusion

Sublethal concentrations of As were previously found to adversely affect earthworm life-history parameters. Here, earthworms exposed to the sub-lethal concentrations that were associated with these effects were analysed using a combination of techniques to assess metabolic and transcriptomic profiles for insight into the mechanisms of toxicity and metabolism in *L. rubellus* previously naive to exposure. Arsenic speciation ensured that this investigation was consistent withprevious analyses, agreeing that AsIII and AsV represented the majority of As in tissue in earthworms. The proportions of these inorganic species were significantly increased at the highest exposure concentrations and are known to affect a range of cellular processes through disruption of proteins and increasing levels of oxidative stress, while affecting signalling mechanisms and energy sources.

Evidence for the effects of As upon these molecular processes was observed via RNAseq data, which also demonstrated a regulation of genes associated with production of GSH and PC as well as identifying a potential membrane protein (an aquaporin) that may act to limit intracellular As concentrations. Transcriptomic analysis linked many similar responses seen in the literature, anchoring the other endpoints and potentially linking responses across taxa. Metabolomic analysis further underpinned the transcriptomic data and firmly implicated the methionine cycle and glutathione production as a response to As exposure, as supported by exposure concentration related changes in methionine and glycine levels. The inter-relationship between metabolite and variation in gene expression seen demonstrates the benefit of a systems biology approach, which has allowed for a better resolved analysis in this instance. An example of the implications includes the unexpected implication of Fe at all levels of analysis, supporting its role in managing As and highlighting it as a candidate for inclusion in future analyses relative to As exposure. This investigation has also delivered a range of pathways in which genetic variation among adapted populations may reflect mechanisms responsible for resistance. In light of these findings, future analyses would strongly benefit from the work of Raab et al. (2005),

who has developed methods for reverse phase analysis of sulfhydroxl-bound As. This would reveal the association of As with GSH as well as the role of PCs within earthworms exposed to As.

Chapter 4

Phenotyping Earthworms Derived From Field Populations to Determine Adaptation to Arsenic

4.1 Introduction

Adaptation of earthworms to a range of soil contaminants is a field of research that has enjoyed wide-ranging analyses to attempt to explore the mechanisms and ecology behind such phenomena (Andre et al. 2010a; Fisker et al. 2011; Langdon et al. 2003b). In particular, a great deal of attention has been focussed upon arsenic tolerance among populations of *Lumbricus rubellus* known to inhabit highly polluted soils. This includes analyses from two highly contaminated former mine sites at Devon Great Consols (DGC) and Carrock Fell (CF), in the UK. These populations have been shown to have higher fitness relative to control populations when exposed to high levels of As under laboratory conditions (Kille et al. Submitted; Langdon et al. 2001). The strongest evidence that this is based upon a genetic trait comes from populations from DGC being bred to F2, with juveniles belonging to lineages derived from the site maintaining fitness under exposure (Langdon et al. 2009).

While the fitness consequences of As tolerance have been relatively well investigated, the mechanisms for this resistance are poorly resolved. It has been speculated that the adaptation is a quantitative, rather than qualitative trait and may entail reduction of arsenic compounds towards arsenobetaine (Watts et al. 2008b) with complexation via interaction of molecules possessing sulphydryl groups (Langdon et al. 2005b). Metal tolerance among other invertebrates has previously been found to have a basis within variation of the Mt1 gene coding for metallothionein, which has been well implicated in resistance to high concentrations of Cd (Maroni et al. 1987; Tanguy et al. 2002; van Straalen et al. 2011). Arsenic resistance is derived via the *ars* operon in a range of prokaryotes, archaea and fungi, where by a number of genes are associated with transcriptional repression, arsenate reduction and arsenite translocation and efflux (Lin et al. 2006). These instances highlight the range of mechanisms potentially employable by metal resistant organisms and indicate that resistance to As may be somewhat more complex given valency-specific toxicity. A similar range of adaptive mechanisms may be implicated by findings from Chapter 3, wherein an aquaporin, glutathione/phytochelatin, and iron metabolism were implicated with As exposure. The means by which an organism metabolises a chemical and the variation in which that

implicates resistance to high concentrations will therefore require a range of cascading analyses representing the different levels of biological organisation.

Determining As tolerance in earthworms has proven to be complex. Previous work has led to the suggestion of a quantitative trait (Langdon et al. 2009) and the extent of the adaptation may be difficult to determine on account of the laborious nature of isolating individuals and breeding them in the laboratory for comparative analysis with unadapted populations. Additionally, the soil environment is extremely heterogenous and geochemical profiles are highly dimensional, making determination of the nature and extent of specific adaptations to a selective environmental variable complex.Moreover, it may be extremely difficult to recognise complex genetics underlying a trait with basic toxicological endpoints, therefore a suitable, quantitative measure would benefit analyses, especially if it can be married with classical ecotoxicological endpoints. It is likely that population specific variation, mutually beneficial to what may be considered polygenic traits, will fail to be recognised among consensus candidates if molecular phenotypes go unexplored.

Recombinant inbred lineages (RILs) provide a basis for resolving genetic variation responsible for phenotypes and provide insight towards recognising the number of loci involved in adaptive traits (Walker et al. 2006). The method involves pairing an individual that is homozygous for the adaptive trait with another that is homozygous susceptible. Their heterozygous offspring are then inbred to produce F2s that are likely to consist of both homozygous positive and negative individuals, as well as heterozygotes, varying relative to the complexity of the trait. Suitable methods of phenotyping individuals are therefore necessary to link end points with genetic variation responsible for adaptive traits in RILs. What's more, understanding variation in life-history endpoints, such as those described in Chapter 2, may elucidate subsequent costs associated with adaptation.

Analyses that focus primarily upon metal handling have been widely implemented to investigate potential variation relating to adaptation and so may be valuable as a means of discriminating between phenotypes. Arsenic speciation has been widely used to infer metabolism and speculate upon adaptive mechanisms in earthworms (Langdon et al. 2002; Watts et al. 2008b). This set of analyses has shown that earthworms from DGC tend to only have higher levels of inorganic As, with concentrations of arsenosugars, methylated As and arsenobetaine potentially only accumulating as by-products of metabolic activity or via environmental uptake. Previous work focussing upon metal compartmentalisation has demonstrated basal biological operations within

an organism in relation to exposure and would be better suited to determine phenotypes and adaptive processes (Arnold et al. 2008; Vijver et al. 2007). Vijver et al. (2007) used differential centrifugation to describe Cu handling in the earthworm, *Aporrectodea caliginosa*. For their analysis, the authors described 3 definable cellular compartments suitable for analysis, termed C, D and E. These reflect the cytosolic, soluble protein and microsomal fraction (C) the granular fraction (D) and tissue fragment, cell membrane, and mitochondrial fraction (E). Such end points are capable of defining management of metals between discrete cellular fractions and are therefore potentially indicative of variation relating to adaptation to metal toxicity. Specifically, resolving As compartmentalisation among earthworms derived from polluted populations may provide insight into potential mechanisms associated with adaptation. This is true particularly when these metals are analysed in conjunction with elements such as Fe and P, which have a defined role in As metabolism (Lam et al. 2008; Lee and Kim 2008; Winski and Carter 1995).

This investigation will therefore focus upon the ability to resolve phenotypes among RILs through analysis of cellular compartmentalisation following exposure to As. This will demonstrate potential phenotypes relative to As resistance among related natural populations, which may be used in association with future genetic analyses. Variation in the mass of earthworms will also permit an insight into the role of adaptation upon development, relative to As accumulation. A novel method of exposure via liquid medium will allow for constant observation in a low stress, minimal handling system that is based upon classical aquatic ecotoxicological methodology (Rand 1995).

4.2 Method

4.2.1 Collection, Culture and Breeding

To determine the number of loci that may confer resistance in potential genetic analyses, it was necessary to design a classic Mendelian “full-sib” breeding experiment that began with field-collected homozygous individuals, as is summarised in Figure 4.1. Putative tolerant earthworms were sampled from Devon Great Consols (DGC), a Cu and As mine near Tavistock, Devon, (SX 423736) and also from Carrock Fell (CF) an As and W mine (NY 324330) located on the edge of the Cumbrian Lakes. Sites are definable in Figure 5.4. Homozygous non-resistant individuals were collected from the control site, Dinas Powys (DPA) (ST 149723), in south Wales. Juveniles (individuals lacking a clitellum) were preferentially selected, as adult *L. rubellus* are capable of

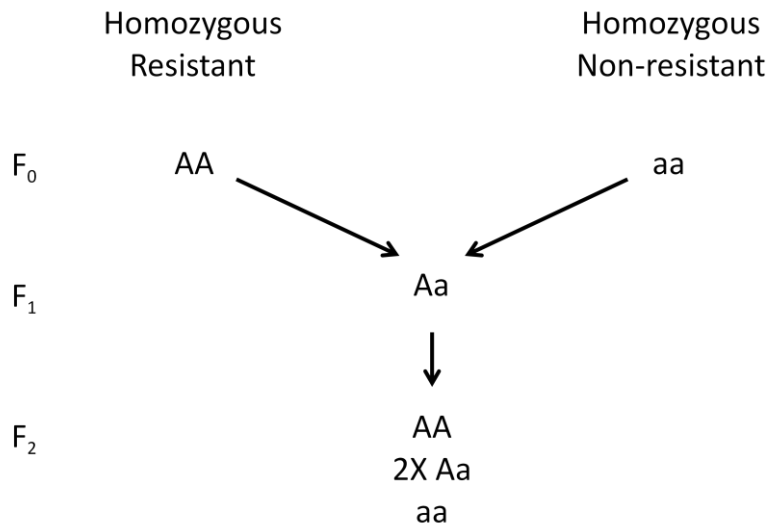


Figure 4.1 Breeding strategy used to produce recombinant inbred lineages of *L. rubellus*, derived from natural populations to establish phenotypes relating to As resistance. Specific details are given in 4.2.1. Alleles are shown for a monogenic trait.

storing spermatozoa for up to 6 months following hermaphroditic copulation (Edwards and Bohlen 1996).

Throughout the breeding experiment, all juveniles were maintained individually at 12 °C in constant darkness to maximise growth and reduce mortality by disease (Klok 2007). Upon relocation to the laboratory, earthworms were kept individually in 250 g of culture soil consisting of a 1:1:1 ratio of commercially available clay loam soil (Broughton Loam, Kettering, UK), composted bark and peat moss (LBS Horticultural, Colne, UK), that was further wet with 20% deionized water. Earthworms were fed sterilized, uncontaminated horse manure as required.

Upon development of a clitellum, individuals were considered to be adult, at which point putative homozygous resistant individuals from the two contaminated sites (F0) were crossed with homozygous non-resistant individuals originating from DP to produce heterozygous individuals (F1). The basis for crossing earthworms was established through genotypic characterisation of sites via sequencing of the mitochondrial cytochrome oxidase II (COII) gene (2.2.4 and 5.2.6). Individuals were preferentially paired with others assumed and later determined (5.3.3) to be of the same clade. In order to identify parents for future analyses, those from control sites within pairs were subcuticularly injected with a visible implant elastomer (Northwest Marine Technology, Inc., Washington, USA) as described by Butt and Lowe (2007). All tags remained in place over the duration of the experiment and only a single mortality

was observed. After 60 days, F0 earthworms were removed from the soil, which was then sieved using low pressure flowing water through a 2 mm sieve to extract cocoons. These were kept in water at 14 °C and regularly assessed for hatched individuals, which were removed and placed in deionised water at 4 °C to produce a “day 0” cohort. F0 worms were frozen in liquid nitrogen before being stored at -80 °C for further analysis. Hatched juveniles were transferred to culture soil and reared to adulthood as previously stated. When mature, earthworms were coupled with siblings for the production of cocoons as with F0 parents. Emergent F2 juveniles were reared in clean soil as described previously until 150 days in culture. At this time the F2 individual were of a suitable size to be used in the experiment. F1 individuals were kept in pairs at 4 °C, forming a resource available for future analyses.

4.2.2 Water exposure

Earthworms (n=27) from either edge of a body mass range of available RILs were selected for exposure, representing 9 from 3 lineages originating from both DGC and CF to resolve potential costs associated with adaptive phenotypes. These earthworms were prepared for exposure to As and subsequent fractionation and As analysis. To provide a baseline against which to compare tissue concentrations in exposed individuals, RILs (n=4) randomly selected from each population were left unexposed. These earthworms were used to determine normal tissue compartmentalisation strategies employed following exposure to the background metal concentrations naturally present in soils used for rearing.

Individuals were depurated for 48 hours in McCartney bottles lined with filter paper, rinsed in deionised water (Arnold and Hodson 2007). Filter paper was changed every 12 hours to prevent coprophagy. At this point, unexposed controls and immediately processed for analysis. For the water exposures, 40 ml of distilled water was spiked with 125 mg/L sodium arsenate ($Na_2HAsO_4 \cdot 7H_2O$) (Sigma Aldrich, Dorset, UK), in a 100 ml flat-bottomed glass jar covered with perforated foil to prevent escape but allow gas exchange. The As contaminated solution was replaced every 7 days to ensure that As concentration was maintained and kept in solution. Every other day, worms were checked for condition. Any worms that died (unresponsive, enlarged clitellum), were immediately washed, weighed, frozen and removed from the analysis. After 28 days, the remaining earthworms were weighed and a tissue sample (~10 mg) was removed for future genetic analyses. Andre et al. (2010b) determined that this period was sufficient for cellular Lead concentrations to reach equilibrium, and was used without data suggesting otherwise for As. In preparation for As accumulation in cellular fractions, Individuals were then placed in covered McCartney vials and oven dried for

48 hours at 100 °C. All glassware and consumables throughout the analysis were acid washed prior to use.

4.2.3 Cell Fractionation

Both exposed and unexposed RILs were individually homogenized in 4ml 0.01 M Tris–HCl, pH 7.5, for 2 minutes using a SilentCrusher M (Heidolph, Essex, UK). The homogeniser blade was well rinsed using 0.01 M TRIS after each sample was processed. All washings were added to the tube representing each respective sample and subject to differential centrifugation for tissue fractionation, following methods described by Vijver et al. (2007), with modifications by Andre et al. (2010b). Samples were centrifuged at 6 °C and 10,000 g for 30 minutes before the supernatant was poured into a McCartney bottle as the C fraction. 4 ml HPLC water was added to the pellet before it being vortexed and heated at 100 °C in a water bath for 2 minutes. Following, 1M NaOH (4 ml) was added before being briefly vortexed and placed in a water bath at 70 °C for 1 hour. Samples were then subject to centrifugation at 10 °C, at 10,000 g for 15 minutes. The supernatant was decanted into a second McCartney bottle as the E fraction. The remaining pellet was resuspended in 2 ml 0.01 M TRIS before being well vortexed and decanted into a third McCartney bottle as the D fraction. The bottle was rinsed with 0.01M TRIS with all washings going into the D fraction. All fractions were then dried using a sand bath in a fume cupboard at 100 °C. Individual fractions were digested in boiling 2 ml 16 M HNO_3 and solutions were evaporated dry as above and repeated once more before 0.5 M NHO3 (5 ml) was added to each McCartney bottle. All chemicals were purchased from Sigma Aldrich (Dorset, UK).

4.2.4 Metal Analysis

Concentrations of metals and As in fractionated samples were measured by inductively coupled plasma optical emission spectroscopy (ICP-OES) (Optima 7300, Perkin Elmer, Cambridge, UK) using matrix matched standards. The C and D fractions were diluted by a factor of 5 and the E fraction by a factor of 10 in order to reduce the acidity and total dissolved solids to a level that would not damage the ICP-OES. Detection limits were calculated as 10X the standard deviation of the concentrations in 10 reagent blanks. Precision was estimated as the median of the difference between two repeat analyses on each sample expressed as a percentage of the mean value. No certified reference materials are available for the fractionation method so accuracy can not be unambiguously stated. However, the analysis of a custom reference standard (University of Reading, UK) gave recoveries in the range 90 - 110 %. The resulting concentrations for fractions from individuals are expressed as mg of metal per kg (wet weight) of earthworm.

4.2.5 Statistics

Concentrations of As in cellular fractions observed among exposed and unexposed RILs were compared using t-tests to distinguish significant variation in accumulated As between groups. Analysis of variance (ANOVA) was used better resolve variation in metal concentrations observed between cohorts identified among exposed RIL, as well as to assess the influence of earthworm mass on As concentrations (Minitab v15). Where significant differences were found within the ANOVA, Tukey's post-hoc test was used to resolve significant variation between treatments.

4.3 Results

Mortality remained low among exposed earthworms (n=2 and n=1 for CF and DGC, respectively). These individuals were removed from the exposure and statistical analysis and kept at -80 °C for future analyses.

4.3.1 Cellular Metal Concentration

In RILs of *L. rubellus* derived from DGC, As was widely spread among exposed individuals inthe C fraction, found to be significantly higher than unexposed controls (T-test, $P < 0.01$). Concentrations of As in the D fraction exposed and unexposed earthworms were generally lower than those found in the E fraction(Figure 4.2). Arsenic concentrations of D and E fractions remained similar between both groups, relative to controls (for both, $P > 0.05$). The presence of a similar As concentrations in these fractions between the exposed and unexposed worms suggests that factions D and E play no discernible role in metabolism. In exposed RILs derived from CF, As concentrations in the C and E fractions were more variable than those in unexposed earthworms, accumulating significantly higher concentrations ($P < 0.001$ and $P < 0.01$ for C and E, respectively). In contrast, analyses of the D fraction failed to identify any significant difference between the unexposed and exposed groups ($P > 0.05$). Dotplot graphs were used to distinguish cohorts relative to quantitative variation in As concentrations of cellular fractions throughout the RILs, considered to be indicative of genotypes (Figure 4.2). Cohorts were determined using only As concentrations, selecting candidate cohorts via comparative analysis relative to unexposed individuals. Thus, variation in P and Fe within discrete fractions may be identified relative to variation in As concentration.

To identify putative cohorts in the DGC population, data for the C fraction was used as the main source of information, since this fraction was shown as the main site of

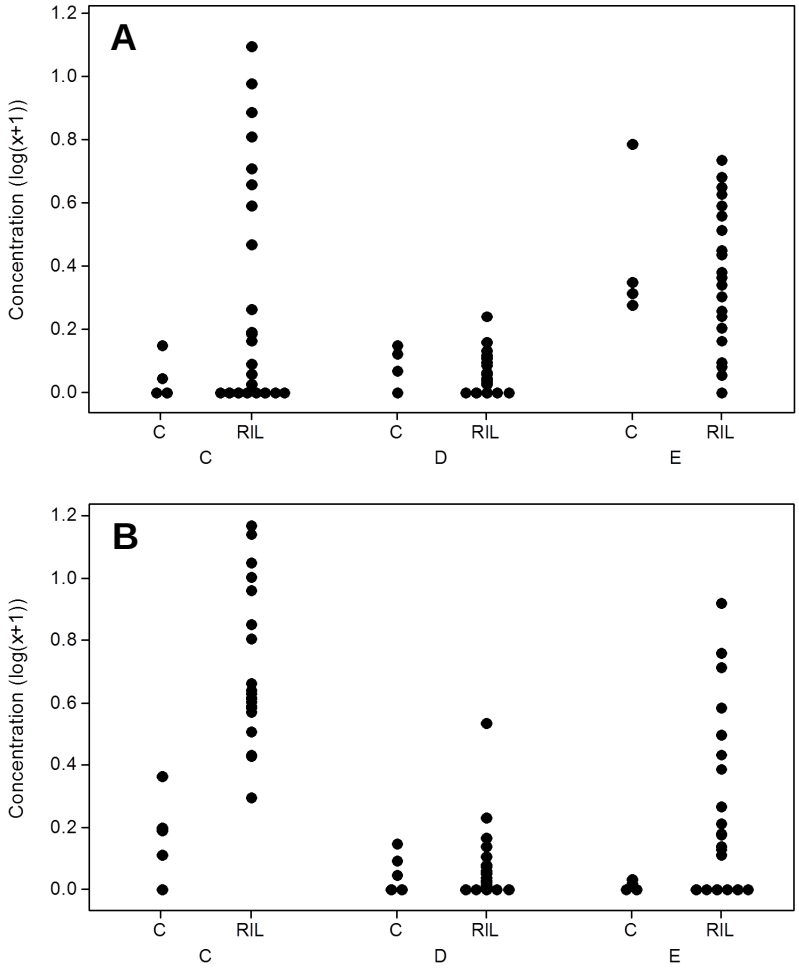


Figure 4.2 Dotplot of As concentrations in cellular fractions as a measure of phenotype in recombinant inbred lineages (RILs) of *L. rubellus* exposed to As for 28 days, compared to unexposed controls. RILs were derived from **A**) DGC and **B**) CF. Concentrations in cellular fractions represent the cytosol, soluble proteins and microsomal fraction (C); the granular fraction (D); and tissue fragments, cell membranes, and mitochondrial fraction (E).

accumulation in exposed worms (i.e. the D and E fraction did not significantly vary for this population). In the C fraction, 8 individuals had distinctly greater quantities of As when compared to the other cohort of 14 individuals. A significantly greater concentration of As was found in the high As cohort relative to that in the low As cohort and controls ($F = 27.38$, $P < 0.001$) (Figure 4.3). Consequently, this group may represent a non-resistant phenotype. Arsenic concentrations among earthworms belonging to the low As cohort were not significantly different from that in unexposed earthworms ($P > 0.05$), which is suggestive of reduced toxicity in what may be represented as an adaptive phenotype.

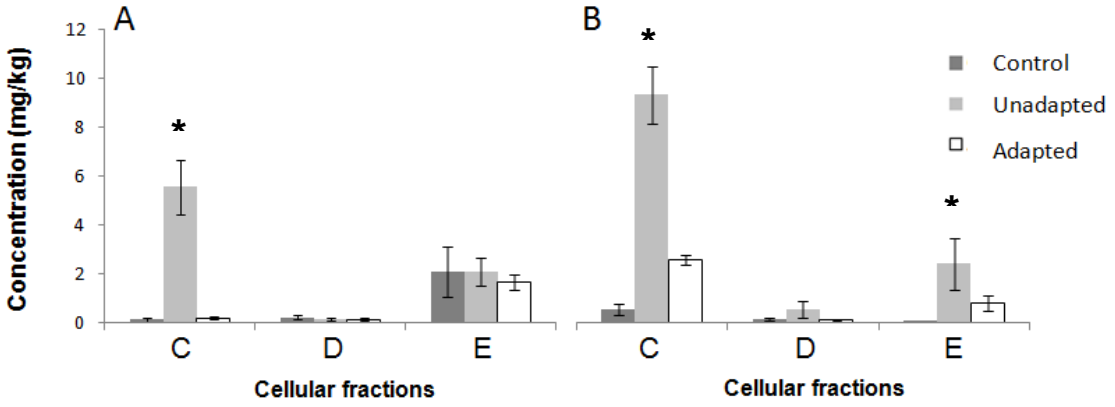


Figure 4.3 Intracellular concentrations of As in *L. rubellus* recombinant inbred lineages derived from populations at DGC (**A**) and CF (**B**) exposed to 125 mg/L As. Concentrations in cellular fractions represent the cytosol, soluble proteins and microsomal fraction (C); the granular fraction (D); and tissue fragments, cell membranes, and mitochondrial fraction (E). * = significant, where $P < 0.05$.

Resolving putative adapted cohorts relative to cellular As concentrations was more complex for RILs originating from CF, whereby the concentration of As in the C fraction of all exposed individuals was significantly greater than that found in controls. This signifies that the C fraction is a major site of As accumulation and suggests that the adaptive phenotypes capable of excluding As, as in RILs from DGC, may not be present among populations at CF. Thus, cohorts were distinguished among earthworms relative to the C fraction, though both high and low As cohorts had accumulated significantly higher concentrations ($F = 4829$, $P < 001$). Potentially definable cohorts were also discernible among concentrations of As in the E fraction, with a distinction between what may be considered adapted and control individuals having significantly lower concentrations than unadapted earthworms, as demonstrated by ANOVA ($P < 0.001$, $F = 48.29$). The high As cohort of exposed CF RILs derived from variation in the C fraction are not the same as those derived from variation in the E fraction, but do occur at a similar frequency, therefore it is possible that this reflects site specific cellular mechanisms that relate to As metabolism.

Analysis of As, Fe and P concentrations were analysed across the two populations to provide insight into potential variation in As assimilation strategies. Comparison of C fraction As concentrations demonstrated that there was more in both unadapted ($F = 5.33$, $P = 0.038$) and adapted ($F = 122.4$, $P < 0.001$) earthworms from CF compared to

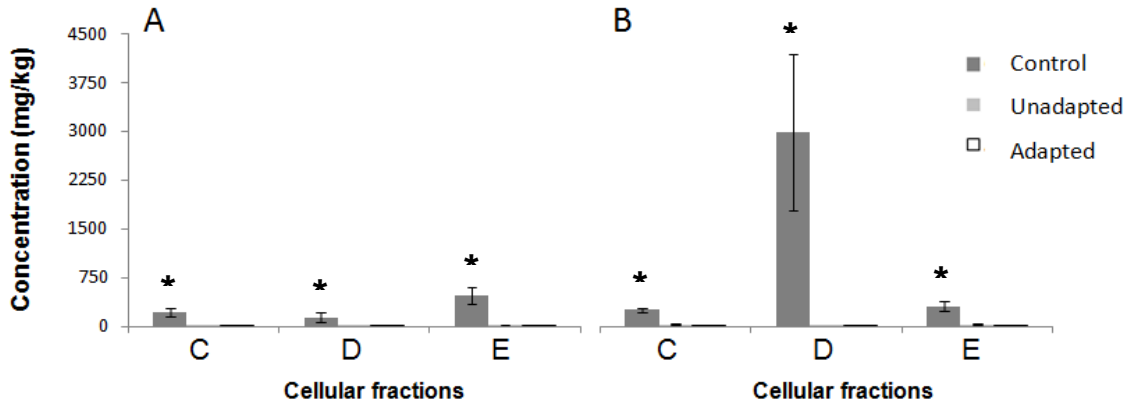


Figure 4.4 Intracellular concentrations of Fe in *L. rubellus* recombinant inbred lineages derived from populations at DGC (**A**) and CF (**B**) exposed to 125 mg/L As. Concentrations in cellular fractions represent the cytosol, soluble proteins and microsomal fraction (C); the granular fraction (D); and tissue fragments, cell membranes, and mitochondrial fraction (E). * = significant, where $P < 0.05$.

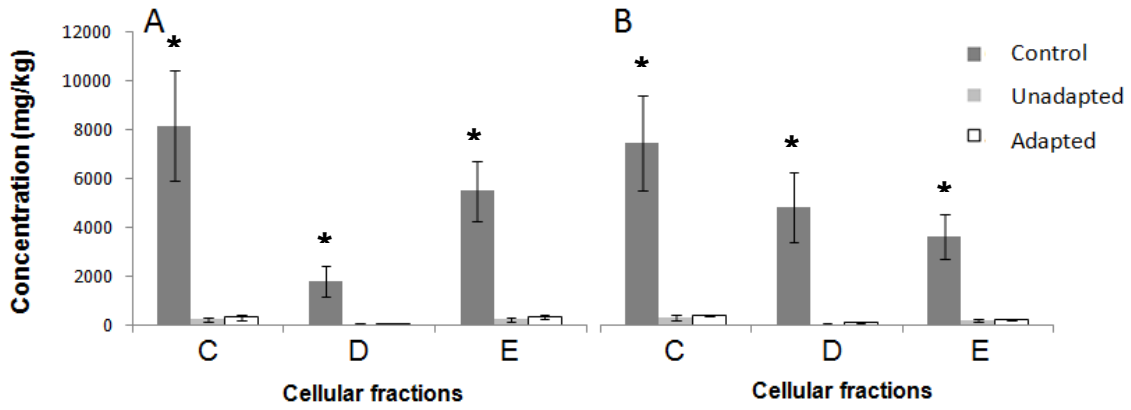


Figure 4.5 Intracellular concentrations of P in *L. rubellus* recombinant inbred lineages derived from populations at DGC (**A**) and CF (**B**) exposed to 125 mg/L As. Concentrations in cellular fractions represent the cytosol, soluble proteins and microsomal fraction (C); the granular fraction (D); and tissue fragments, cell membranes, and mitochondrial fraction (E). * = significant, where $P < 0.05$.

DGC. This may be related to mechanisms associated with adaptive phenotypes. No significant difference between As in fraction E (F = 0.82, P = 0.371) and D fraction (F = 0.52, P = 0.599) were observed. There was no significant difference between unexposed controls for RILs from each of the sites for P and Fe in all analysed fractions, except P in the D fraction, for which CF controls had significantly higher concentrations (F = 4.8, P = 0.037). Fe in the D fraction differed slightly between CF and DGC, but this was not significant (P = 0.076). Fe and P were greater among controls relative to exposed earthworms across all fractions from both DGC and CF (P < 0.001). There was also no family specific grouping within the analysis, indicating that the trait mechanism for handling is shared among RILs.

4.3.2 Mass

There was a general decrease in mass of the exposed earthworms by 0.29 ± 0.07 and 0.02 ± 0.04 g, for the DGC and CF populations respectively. No significant difference was observed between the mass before and mass after exposure among individuals from DGC (P = 0.738) and CF (P = 0.378). There was also no correlation between mass of the earthworm at the beginning of the experiment (after depuration) and the concentration of As in the C fraction of the among RILs derived from earthworms derived from DGC (r = 0.106, P = 0.664) or CF (r = 0.383, P = 0.086). This suggests that As exposure did not have a detrimental effect on the energy metabolism and resource allocation of the exposed earthworms. This finding confirms that any changes seen in concentrations of As and other trace elements in worms will not be the result of difference in the extent of concentration of endogenous metals as a result of weight loss between individuals.

4.4 Discussion

During this investigation, 3 discrete cellular fractions were assessed for As accumulation for implication of mechanisms of As handling. Fractions C and D reflect primary points of impact for metals entering the cell, whereas fraction E is more likely to represent long-term cellular stores. In this instance, the most consistent elevated As concentrations are associated with the C fraction of the exposed worm, suggesting that this is a major As pool following exposure. Further, variation in concentrations of As found in the C fraction is suggestive for discrimination between adapted and unadapted individuals among RILs derived from heavily contaminated former mine sites. The C fraction represents cytosolic and microsomal As, as well as that associated with soluble proteins and can therefore be considered to be readily accessible. Arsenic here has been documented to interact with glutathione, which is capable of reducing arsenate to

arsenite (seen previously at this concentration in soil, 3.3.2) and can also bind to the latter via sulphydryl groups to reduce cytotoxicity (Langdon et al. 2005b), properties shared by phytochelatins (Raab et al. 2004). Arsenic in this fraction has also been found to bind to iron hydroxides and iron rich-proteins such as haemoglobin (Bang and Meng 2004; Winski and Carter 1995).

The low concentrations and general absence of variation in As in the D fraction of RILs demonstrates that the inorganic matrix of granules that includes metal-rich granules is relatively unimportant in terms of As trafficking or accumulation. Among individuals from both populations, the total concentration of As in the D fraction was relatively low and was not significantly different from that of unexposed earthworms. This is in agreement with Vijver et al. (2006), who observed partitioning of a suite of metals among subcellular fractions of *A. caliginosa*. In contrast to the results presented here, the study of Vijver et al. (2006) identified that the majority of accumulated As occurs in the E fraction, which may signify species specific mechanisms for As handling. The E fraction represents a more stable store for As in cells, as it consists of cell membranes, mitochondria and tissue fragments. In this investigation, As in the E fraction shows a continuous distribution among earthworms originating from DGC. Similar concentrations are observable among exposed and unexposed earthworms, which likely reflects long-term accumulation of As from background soil concentrations (17.5 mg/kg As) naturally present within the culture medium.

Measurements of As in the C fraction highlight accumulation in some although not all worms with the measured concentrations showing two cohorts. This bimodal distribution was observed in both of the RILs, although this was more pronounced in the DGC population. The variation seen in the C fraction potentially signifies the key distinguishing feature between adapted and unadapted earthworms residing at DGC and CF and provides further supporting evidence, complementary to Langdon et al (2009), that earthworms at these sites are genetically adapted. The biomodal distribution and the nature of the difference suggests that within RILs, genotypes may be present that both allow and prevent As reaching the biological targets that are associated with the C fraction. Thus, it can be hypothesised that adapted earthworms may be able to significantly reduce the amount of As entering cells. The number of individuals occurring within the adapted group can offer a novel insight into the genetics of such potentiated adaptive mechanisms. Thus, two-thirds of individuals exposed demonstrated low concentrations of As in the C fraction, suggestive of a codominant trait or epistasis between 2 biallelic loci (Hallgrimsdottir and Yuster 2008). The trait may be more complicated yet, based upon variation seen between the 2 sites

as the proficiency of reducing the levels of cellular As appears reduced among earthworms originating from CF relative to those from DGC, as confirmed by Langdon et al. (2001). Although similar mechanisms maybe responsible for adaptation to As, metabolism maybe affected by differing basal biochemical management derived from adaptation to local geochemistry. The observed variation between concentrations of P seen in the D fraction among controls enforces this preliminary insight, given that P has been previously found to mitigate As toxicity (Lee and Kim 2008).

The affects of As exposure have previously been linked to P and Fe (lee and Kim 2008; Lam et al. 2006). Here, both P and Fe were found to occur at low concentrations among all fractions in exposed worms. This may signify that these elements have been scavenged for incorporation into biological processes following the lengthy period of starvation. Specifically for P, we may hypothesise that that this element may occur at lower cellular concentrations among adapted individuals given its chemical similarity to arsenate resulting in competition for uptake by membrane carrier proteins (Lee and Kim 2008). Concentrations of Fe may be expected to increase during soil exposure and would likely be maintained throughout a water exposure. Previous evidence in naive worms demonstrated that Fe handling genes are up-regulated and total tissue Fe increases in exposed worms (Chapter 3).

The reduced fractional concentrations of Fe and P seen in exposed RILs are likely to be a product of the conditions related to micronutrient supply and loss during exposure. The exposure method in this instance is novel, given that it was not previously acknowledged that earthworms were capable of lasting for extensive periods in water (n.b. unpublished work has seen a day 0 hatchlings kept at 4 °C for 5 months, and enforces the long-term suitability of this exposure regime for earthworms). The method was implemented because it holds a number of advantages over soil exposure. For example, exposure via a liquid medium is not only significantly easier to set up than that through soil matrices, but may also allow for better interpretation of classic ecotoxicological endpoints relative to measurements of bioavailability in soil exposures. The method also allows for recovery of material for subsequent chemical or genetic analysis if the organism should die (earthworms decompose readily in soil- pers.obs.). What's more, the method means that earthworms can be constantly assessed for vitality without handling the animal. In the experiment, it is not possible to feed the worms, a factor potential linked to the reduced P and Fe in exposed earthworms relative to controls, however, previous evidence suggesting that absorption of metals is largely dermal (Vijver et al. 2003) indicates that this exposure regime may be sufficient

not only to understand toxicity, but may also be ecologically relevant to field exposure scenarios.

Finally, there was no variation in mass associated with concentration of As in the C fraction in unadapted and adapted individuals, indicating that size has no relation to quantities of As here and is therefore not indicative of costs related to adaptation. It is also not clear as to how variation in As concentration across the fractions reflects toxicity. Therefore the neutral red retention assay may be a suitable candidate to partner with the earthworm water toxicity test (Weeks and Svendsen 1996).

4.5 Conclusion

The work here implicates RILs as an important biological resource for determining adaptive traits in natural populations, specifically signifying that that they are available for contemporary genetic analysis to link phenotypes to genotypes. What's more, this work determines that there are quantifiable methods already described in the literature for discriminating between the range of phenotypes that occur. Analyses within this investigation have demonstrated that the C fraction of homogenised earthworms, representing cytosolic, soluble protein and microsomal fraction, is a major site of As accumulation in exposed earthworm tissues. Variation in As concentrations of the C fraction can also permit for differentiation of adapted and unadapted individual *L. rubellus* from RILs originating from DGC and CF. This is suggestive of a genetic mechanism for adaptation to As exposure among earthworms at these sites, which prevents uptake or facilitates elimination of As. The proportion of adapted individuals and variation in the quantities of As, P and Fe are indicative that not only are traits complex, but may be reliant on other population specific environmental adaptations. Mass, indicative of the rate of development, was not correlated with the C fraction and therefore a cost of As resistance in these populations remains to be determined. This analysis was facilitated by a novel exposure regime that will have beneficial implications for phenotyping in earthworms. Ultimately, the most prolific means of distinguishing adaptive mechanisms would be to directly determine genetic variation relative to phenotype.

Chapter 5

The Effect of Metal Pollution on Genetic Variation Among Populations of the Earthworm, *Lumbricus rubellus*, Using Mitochondrial and Genetic Markers.

5.1 Introduction

The rich geological nature of the United Kingdom means that a wide variety of mineral ores are available, and have been heavily exploited for over 4,000 years (2008). Currently, there are no operating metal mines in England or Wales, yet thousands of abandoned former mine sites remain, leaving a legacy of soil contamination cited as one of the most significant pollution threats (Hudson-Edwards et al. 2008). For instance, soils at the former arsenic mine, Devon Great Consols (DGC) in southwest England, have an average of 8,081 mg/kg As, which is 2 orders of magnitude above background concentrations (Klinck et al. 2005). Similar sites exist that are simultaneously contaminated with copper, nickel, zinc, cadmium and lead. These mine sites are often characterised by the significant lack of biodiversity relative to uncontaminated sites. However, a range of organisms are found on highly contaminated soils. The populations that inhabit these sites are subject to the persistent and strong selective pressure imposed by metal and metalloid contamination. Therefore, adaptive traits that mitigate toxicity are often assumed to develop among populations inhabiting the highly disturbed soils at former mine sites.

The earthworm, *Lumbricus rubellus* is commonly found across highly contaminated former mine sites (Langdon et al. 2001; Marino et al. 1998; Morgan and Morgan 1990). In particular, the populations at DGC have been well studied and evidence for genetically based resistance to As toxicity found (Langdon et al. 2009). Analysis of mitochondrial and nuclear genetic markers have recently shown that *L. rubellus* consists of a complex of 2 cryptic clades, differing by 13% as determined by mitochondrial and nuclear markers (Andre et al. 2010a). Recent work at DGC has demonstrated that there is reduced genetic variation among populations relative to soil As (Kille et al. Submitted) and that the extent of this selection differs between the two clades. Such stringent selection, such as that seen among *L. rubellus* clades at heavily contaminated mine sites can lead to a loss of genetic variation termed “genetic erosion”, which can play a significant role in further destabilisation or even extinction.

Van Straalen and Timmermans (2002) critically assessed the attempts made to explore the occurrence of genetic erosion among natural populations exposed to stressors derived from anthropogenic sources, citing mutation rate, directional selection, population bottlenecks and migration, by which populations suffer from depression in genetic variation. The importance of recognising these events stems from the effects of reduced genetic variation upon effective population size and thus adaptive potential. The small population sizes associated with strong selective pressures will, in turn, significantly exaggerate the effects of demographic (reproductive and mortality) and environmental stochasticity, which can additionally limit adaptive response and increase the risk of extinction (Hoffmann and Sgro 2011). For example, Athrey et al. (2007) composed a study whereby selection for cadmium resistance occurred among populations of the Lesser Killifish (*Heterandria formosa*). Using microsatellite markers, adaptation in this instance lead to a greater loss of genetic diversity, relative to control populations, which also reduced variation in response to the small effective population size.

Phylogeographic analysis of affected populations relative to unaffected populations can provide an indication as to the ecological effects of metal and metalloid contamination in the terrestrial environment. Such analysis using tools including characterisation of mitochondrial markers and assessment of nuclear genotype and genetic variation has the potential to inform on the association of genetic characteristics to soil conditions. Conducting an extensive phylogeographic analysis of *L. rubellus* from sites across the UK will provide a foundation for acknowledging the level of variation seen in the complex throughout the UK. Sites relative to a former boundary that occurred during the Devensian glaciation may also be indicative of recent selection events among the clades following recent (< 10,000 years) expansion into previously uninhabited environments (Andre et al. 2010a). Thus, comparing the genetic variation that occurs both on site and off over a range of highly contaminated former mine sites will exemplify the adaptive potential of specific clades and the effects of such extensive pollution upon an important terrestrial sentinel over what is likely to be different time scales.

Mitochondrial sequencing is often employed for phylogenetic analyses because mitochondrial markers are considered to be neutral in that they are not subject to recombination or selection (Hurst and Jiggins 2005). However, recent research suggests that mitochondrial DNA can be unreliable for inferring recent evolutionary history and therefore should be coupled with multiple nuclear markers (Ballard and Whitlock 2004). A recent development in the field of molecular ecology is the advent of

Restriction Associated DNA sequencing (RADseq); a technique capable of deriving thousands of nuclear markers with the potential to profoundly escalate what may be understood about speciation and adaptation in natural populations. RADseq incorporates restriction enzymes to break up genomic DNA (gDNA), before adaptors are ligated onto size selected fragments, which are then selectively amplified (Baird et al. 2008) before sequencing with contemporary high-throughput technology. Emerson et al. (2010) were able to recently elucidate the phylogeography of the pitcher plant mosquito, by initially establishing mitochondrial genotypes before pioneering the use of RADseq to better resolve adaptive radiation. In a similar vein, this investigation sets out to explore genetic variation relative to metal adaptation in *L. rubellus* by sampling from populations inhabiting a range of former mine sites across the UK, as well as proximal uncontaminated sites. Mitochondrial markers will then be used to provide initial genotypes and a basis to pool samples from sites for analysis by RADseq. This data can therefore better inform us as to the extent of variation seen among the *L. rubellus* species complex and provide insights as to the effects of adaptation to metal contamination.

5.2 Methods

5.2.1 Sampling

Adult *L. rubellus* were collected from eight field sites from across the UK. Six of these sites were abandoned mine sites with a history of sampling and metal analysis: DGC a Cu and As mine near Tavistock, Devon, (SX 423736); South Caradon (CAR) a Cu mine, north of Crow's Nest, Cornwall (SX 265700); Shipham (SH) a Pb and Zn mine noted for high levels of Cd, Avon (ST 448573); Ecton (ECT) a Cu mine, Staffordshire (SK 098581); Carrock Fell (CF) an As and W mine (NY 324330) located on the edge of the Cumbrian Lakes; and Cwmystwyth (CWM) Pb, Zn and Cd mine (SN 803748). The distribution of the selected sites is relevant to the key contaminants that occur at each. Thus pairs of Pb/Zn and As mines were identified in which one of the mines in each pair was located on either side of a glacial boundary as proposed by Bowen et al. (2002). Ecton was used as a former Cu mine for pairing with Caradon in the same manner, however both sites occur in geographically distinct but unglaciated locations. At 5 sites a single population was collected from an area at which it was established from previous work that soil metal concentrations were substantially above background. At DGC, sampling of 2 distinct areas within the site was undertaken on account of its large size and as it has been subject to many previous investigations. In addition to sampling on the contaminated areas, proximal "clean" sites located in the same geographic/geological region as each of the mines were also sampled. Finally

two distal control populations from uncontaminated sites that were know to be dominated by clades A and B respectively following genotyping by Andre et al. (2008), were also sampled. These sites were located at Dinas Powys (DPA) (ST 149723) and in the area of Clydach (CLAD) (SN 696013) South Wales. Geographic location of sampled sites is apparent in Figure 5.4. The determining characteristics of these control sites included similar soil consistency in an area that lacked evidence of contamination such as spoil heaps or bare soil. At each of the sample sites, worms were dug from the soil by hand and were maintained on respective soil until being processed. The posterior (< 10 mg) of each earthworm was removed using a sterile scalpel blade for DNA extraction. Remaining tissue was frozen using liquid nitrogen and stored at -80 °C.

5.2.2 Soil Metal Analysis

Characterisation of the soil was conducted to confirm the exposure histories of all of the sampled populations. For metal analysis, soils were processed as in 2.2.5. Certified values for reference materials corroborated well with data in this analysis, being within 10% of mean certified concentration for As concentrations.

The relationship between key contaminants and the occurrence of site specific clades as determined in a number of ways. Correlation for single metals was initially investigated as a simple approach. This analysis, however, was challenged by complex co-correlations that are found between trace metals. For example, in polluted soils Cu and As frequently show a co-distribution (for example both DGC and CF both contain elevated levels of these two metals. Similarly Zn and Pb are common co-contaminants. To address this issues the association of population genetic characteristics were also related to the total “toxic pressure” of metals present in site soils. This was assessed by estimating sum toxic units (TU) in the soil as a single metric that reduces data dimensionality, where:

$$TU = \frac{\Sigma \text{ Soil concentration}}{EC50}$$

in which EC50 is the effective concentration for reducing the rate of cocoon production by 50% of that observed in controls. The values used for this parameter were 993, 118, 599, 327, 209 and 163 mg/kg for Pb, As, Zn, Cu, Cd and Ni, respectively (Davies et al. 2003; Lister et al. 2011; Spurgeon et al. 2004a; Spurgeon et al. 2000). The correlation of the sum of TU’s for each site was assessed against arcsine transformed proportions of clades occurring at that site, using Pearson’s correlation coefficient.

5.2.3 Sanger Sequencing

The gene selected for analysis of mitochondrial genotype was a region of the earthworm mitochondrial cytochrome oxidase II (COII) gene. COII has previously been established as a marker for inferring the phylogenetic relationship between British *L. rubellus* populations (Andre et al. 2010b) and offers a basic assessment of phylogenetic history. Sequencing of COII was performed as described in 2.2.4.

5.2.4 RADseq

The RADseq libraries were produced in accordance to Etter et al., (2011). This method works to provide thousands of markers by selectively amplifying restriction sites that are ligated to adapter sequences, each containing primers for sequencing using high density contemporary technology. Within this study, the following modifications were applied to the initial method.

Archived material was requantified using a Qubit 1.0 (Life Technologies, Paisley, UK), which measures double-stranded DNA. DNA from 4-6 (relative to availability) individuals were equally pooled based upon mitochondrial genotype, representative of the 3 key clades. Samples were digested with RNase A (1 µl/ 50 µl of sample) (Qiagen), for 30 minutes at 54 °C, prior to purification using a Qiagen PCR purification kit. 300 ng gDNA was digested using 2.5 µl SbfI (NEB, Hitchin, UK) and 1.9 µl NEB buffer 4 (rounded to 40 µl reaction size with HPLC water (Sigma Aldrich, Dorset, UK)) at 37 °C for 1 hour, and deactivated at 65 °C for 20 minutes. P1 adapters were supplied by The Genepool (Edinburgh, UK) and based upon the design:

Top:
5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
ATCTXXXXXTGC*A3'

Bottom:
5'/Phos/XXXXXAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTG
GTCGCCGTATCAT*T3'

(* = Phosphorothioate bond, /Phos/= 5' phosphorylation, XXXXX=barcode)

and were ligated to samples individually, using 0.75 µl of 100 nM P1 adapter, 1 µl of 100 mM rATP (NEB), 1 µl NEB buffer 2 (contains 50 mM salt), 0.5 µl T4 ligase (NEB), made up to 10 µl with HPLC water. Samples were incubated at room temperature for 30 minutes before heat inactivation at 65 °C for 20 minutes. Samples were then pooled

into libraries of 12 sites per pool. 2 µg gDNA per pool was randomly sheared using a Covaris S series (KBiosciences, Herts, UK); set to achieve 300-800 bp fragments following the regime: duty cycle 10%, intensity 4, cycles per burst 200, time (s) 80, performed in 120 µl tubes. The 3 libraries were then concentrated using a vacuum concentrator (VacuFuge 5301, Eppendorf, Stevenege, UK) at 60 °C until at a volume of 20-40 µl before being run on 1% ultrapure low melting point agarose gel (Life Technologies) and excision of desired bands using a sterile scalpel. Samples were purified using via Qiagen MinElute Reaction Cleanup Kit, eluted in 20 µl EB. Samples were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter, High Wycombe, UK) as according to the manufacturer's instructions to remove unligated P1 adaptors. Fragments were blunt-end repaired using the Quick Blunting Kit (NEB) to blunt end the DNA. 19 µl of sample was mixed with 2.5 µl buffer, 2.5 µl dNTP and 1 µl enzyme, incubated at room temperature for 30 minutes, prior to purification via Qiagen MinElute Reaction Cleanup Kit and elution in 24 µl buffer EB. dATP overhangs were ligated to fragments by mixing 23 µl of sample with 15 U klenow exonuclease (NEB), 1 µl dATP (100 mM) (NEB), 3 µl NEB buffer 2 , and incubated at 37 °C for 30 minutes, and then purified with Qiagen MinElute Reaction Cleanup Kit and eluted in 25 µl buffer EB.

The second (P2) adapter was similar to those used in other studies, except that it was modified to have a barcode to overcome multiplexing limitations in future experiments. This barcode is one of 4 P2s produced, using 4 of the 64 potential combinations in a 5 bp barcode that each deviate by 2 bases. The modified P2 was assembled from these oligos, produced and HPLC purified by Integrated DNA Technologies (Glasgow, UK):

Top:
5'/Phos/XXXXXAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCAGAACAA3'

Bottom:
5'CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTXXXXX*T3'

(* = Phosphorothioate bond, /Phos/= 5' phosphorylation, XXXXX=barcode)

100 µM stocks of the P2 for each single stranded oligonucleotide in 1X Elution Buffer (EB: 10mM Tris-Cl, pH 8.5). Complementary adapter oligos were combined at 10 µM in 1X AB (10X AB: 500 mM NaCl, 100 mM Tris-Cl, pH 7.5-8.0 (Sigma Aldrich)) and

placed in a Veriti thermocycler (Applied Biosystems, Warrington, UK) at 98 °C, before slowly being ramped down to 4 °C over 2 hours. Cooling rate was slowed around the annealing temperature, with 2 minute intervals at each 0.5 °C between 75 and 70 °C.

The P2 Adapter (10 µM) was combined with 1 µl 100 mM rATP, 3 µl NEB buffer 2, 0.5 µl T4 ligase and 24 µl DNA, before incubation at room temperature for 30 minutes. Each library was then selectively amplified for fragments containing both P1 and P2 adaptors, performed in 8 20 µl PCR reactions using 2 µl of P2 ligated library with 10 µl Phusion Master Mix (NEB), 1 µl P1-PCR primer (10uM) AATGATACGGCGACCACCGA, 1 µl P2-PCR primer (10 µM) CAAGCAGAAGACGGCATACGA and 7 µl HPLC water. Cycling conditions were as according to manufacturer's instructions; 98 °C for 30 s then 18 cycles of 98 °C for 10 s, 65 °C for 30 s, 72 °C for 30 s, and a final extension at 72 °C for 5 minutes. PCR amplicons 300–800 bp long were then size extracted using gel electrophoresis (1X TBE, 1% agarose), and purified using with Qiagen gel extraction kit. Purified samples were quantified using a Qubit 1.0 fluorometer before being pooled in equal measure. Samples were assessed for quality using qPCR by The GenePool (Edinburgh, UK) before being sequenced on a HiSeq 2000 (Illumina, Essex, UK).

5.2.5 SNP analysis

The software package, Stacks, was used to assess the resultant sequences for quality, before sorting them relative to barcode, calling SNPs and determining output suitable for phylogenetic analyses (Catchen et al. 2011). Uninformative duplicates derived from PCR were removed using the script clone_filter.pl from stacks and resulted in 23,152,831 reads suitable for analysis. Sequences were processed further using the software stacks to interrogate populations, requiring a minimum of 3 raw reads to create a stack, allowing 2 mismatches between loci per population, and 3 mismatches between loci when building a catalogue for comparison across populations. Homologous RAD tag loci consensus sequences were aligned among populations for clade specific comparison using the populations programme of stacks. Any locus that was present in at least two populations at 15X coverage was retained and used in subsequent phylogenetic analysis. SNPs from putative loci were concatenated for phylogenetic analysis relative to clade as determined using the COII mitochondrial marker. The organism specific source of putative loci containing SNPs used for phylogenetic analyses were identified using BLASTN (Altschul et al. 1997), to ensure that there was no representation of the microbiome (particularly of *Verminephrobacter sp.* and *Acidovorax sp.* which inhabit nephridea), variation of which can deviate relative to soil chemistry and bias phylogeographic analyses (personal observation). A

population map for the populations software was created to establish the relationship between all sites. A proxy for genetic erosion was determined by comparing the number of unique RADseq loci (or “stacks”) from 400,000 reads between control and mine site populations of the same clade.

5.2.6 Phylogenetic Analysis

jModelTest was used to define appropriate models for phylogenetic analyses. Using Akaike’s information criterion, it was determined that GTR+G and HKY+G be used for the SNP and mtDNA data, respectively (Posada 2008). 392 bases were used for the COII analysis and Fst values were determined using Mega 4.0 (Kumar et al. 2004). *L. eisenei* and *L. castaneus* were used as outlier sequences to ensure correct identification. Alignment Transformation EnviRonment (ALTER) was used to collapse common sequences into haplotypes for analysis (Glez-Pena et al. 2010). Maximum likelihood analyses were conducted using the PhyML 3.0 server, with 1,000 bootstrap replicates (Guindon et al. 2010). Bayesian analysis of COII data was conducted using MRBAYES v3.2 over 2 independent runs, whereby four Markov Chains were run over 2 million iterations and sampled every 1,000 generations, with the first 5000 trees discarded as burn-in (Huelsenbeck and Ronquist 2001). Bayesian analysis of SNP data occurred over 2 independent runs, whereby four Markov Chains were run over 500,000 iterations and sampled every 100 generations, with the first 100 trees discarded as burn-in. 50% majority-rule consensus trees were used to determine agreement in phylogenetic models, using the program Consense included in the PHYLIP package (Felsenstein 1997), running it with a strict consensus setting and treating the trees as unrooted. Phylogenetic analysis for determination of F0 clades in the breeding analysis were determined as above.

5.2.7 Analysis of Population Structure

The software BAPS (Corander and Martinen 2006) was used to infer the best partition of the mitochondrial DNA sequences into groups. Three independent runs of the analysis were made to assess convergence of the results. Each analysis used uniform prior distributions with random values between 8 and 20 in the prior upper bound vector under the clustering with linked data method.

Analysis of Molecular Variance (AMOVA) were performed with the software Arlequin (Excoffier et al. 2005) on different groupings of the data to explore how the genetic variation was partitioned between and within groups. Three different population structure models were examined with AMOVA: Populations were grouped relative to

major soil contaminants, discrete site locations and also relative to mine and control sites. Further details are shown in Table 5.3.

5.2.8 Breeding Analysis

The breeding success of in earthworms from sampled sites in relation to the production of viable F1 and F2 offspring was used to provide details of clade compatibility. The ability of different clades to successfully mix can provide a better understanding of mechanisms for resultant genetic variation relative to geographic location. Earthworms sampled from mine site were paired with those originating from control locations with single, site dominant clades to produce recombinant inbred lineages, described fully in 4.2.1. As well as the pairing of earthworms sampled from DGC and CF with those from the control site DP, juvenile *L. rubellus* (individuals lacking a clitellum) collected from CWM were paired with earthworms from CLAD. Earthworms were not genotyped before pairing because of high levels of mortality observed following coelomic extraction or posterior tissue sampling. Pairings were therefore based upon the dominant clade occurring at respective sites. Following inbreeding of F1s to produce recombinant inbred lineages, F2 cocoons were collected and assessed regularly for confirmation of hatching. Sequencing of COII was performed as described in 2.2.4 for F0 earthworms to determine the potential causes for unsuccessful pairings.

5.3 Results

5.3.1 Soil Metal Analysis

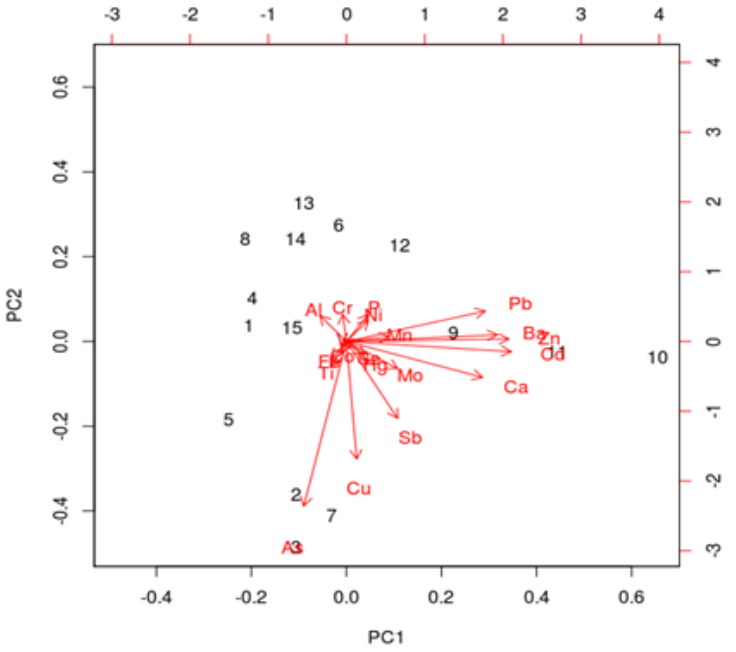
Table 5.1 gives details of the measured concentrations of trace metals in soil samples collected from each of the sampling sites. The two distal control sites each contained concentrations of metals that are within the range that are typically found in uncontaminated soils across the UK (Spurgeon et al. 2008). At each of the 6 mine sites, the concentration of selected metals were higher in the soil that was collected from the site located on the area subject to pollution from mining associated waste materials than those found in the associated reference plot soil. Although reference plot metal concentrations were lower than those found in the mining areas, these soils occasionally contained concentrations that were above background concentrations. For example, As and Cu are higher than average quantities at both DGC and CAR control sites, respectively. This reflects the fact that the soils at these control sites are derived from underlying geology that is relatively rich in the minerals that were the ultimate target of metal extraction activities. Figure 5.1 demonstrates that certain sites group together as a result of such soil chemistry. Soils highly contaminated with Zn, Pb, Ba and Cd cluster along PC1 (SH and ECT), while those highly contaminated with Cu and

As along PC2 (DGC, CF and CAR), which describe 45.7% and 24.6% of the variation in the data, respectively. This reflects the co-correlation of metals associated with the minerals that underlie the three mine type ground, with in particular the Pb/ Zn mines differing from the As/ Cu mines. The soils from the control sites form a distinctive group separated from the mine site soils, likely demonstrating reduced contamination at these sites, although in some cases they reflect basal regional geology to a certain extent by being orientated along the same axis as their associated mine site (e.g. Ecton Control). Soils at CWM cluster among control sites, demonstrating reduced concentrations of contaminants. Trace metal concentrations and their bioavailability at these sites can be expected to be higher than at control sites, due to the relatively low pH and recent disturbance of mine site soils (Morgan and Morgan 1988).

5.3.2 MtDNA Phylogeny

The mitochondrial phylogeny of *L. rubellus* from multiple sites across the UK clearly demonstrates that there is a separation of the sampled worms into two distinct clades (Figure 5.3). This division is seen in almost all of the sampled populations whether collected from the mine or proximal control site (although not the distal controls which were deliberately chosen as sites known to be dominated by a single clade from previously unpublished work). Bayesian and ML analyses essentially derives a certainty of calling the two clades as species, however resolution is mostly lost below this level with the exception of a split in clade B. Specifically, individuals from clade A and B diverge on average by 12.4%, with outlier sequences from *L. eisenei* and *L. castaneus* differing on average by 20% and 21.5%, respectively, demonstrating correct identification of individuals. The maximum inner divergence of clade A haplotypes belonged to a single individual (12.1%), with the next highest from any others as 6.5%. Lineage B observed a maximum intra-specific variation of 5.3%.

There are some differences between sites, with the proportions of earthworms from each site and clade that contributed to the population differing between locations, as illustrated in (Figure 5.2). Substantial differences of greater than 20% in clade frequencies were observed between mine and proximal control site populations at discrete sites, which may reflect lineage specific selection events. Populations on contaminated soil in CF consisted entirely of lineage A worms, whereas lineage B is present among control soils. Earthworms residing at DGCS2 and SHM were predominantly of lineage A and B respectively, with control soils containing a more even representation. The frequency of clades at DGC sites are proportional to those found previous in a survey of this site conducted a year previous to this study (Kille et al., submitted) suggesting that these differences in clade frequency are temporally



DGCC	1
DGCP	2
DGCS2	3
CARC	4
CARM	5
DPA	6
CFM	7
CFC	8
ECTC	9
ECTM	10
SHM	11
SHC	12
CWMC	13
CWMM	14
CLAD	15

Figure 5.1 PCA biplot demonstrating geochemical profiles of former mine sites and proximal control sites. The loadings plot for individual constituents of soils, generally defining the key contaminants at sites recognisable through respective chemical symbols. The scores plot demonstrates the relationships between soil geochemical profiles at sites, definable through use of the table to the right.

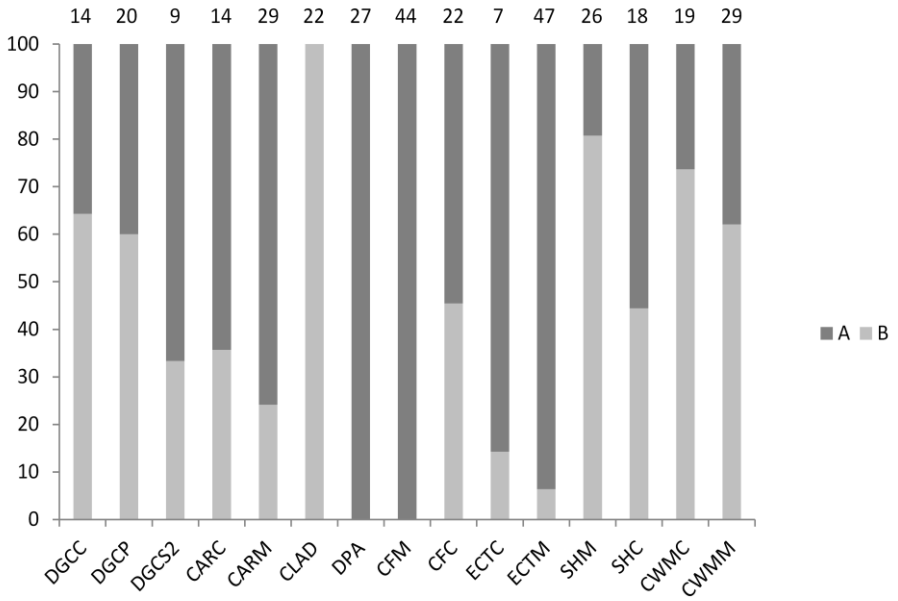


Figure 5.2 The proportion of *L. rubellus* individuals from sites across the UK relative to clades, defined by Maximum Likelihood phylogenetic tree of the cytochrome oxidase II mitochondrial gene. Numerical values above bars represent the number of individuals sequenced from respective sites.

	Cd		Sb		Ba		Hg		Pb	
	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD
CAR (C)	0.3	0.0	1.2	0.1	53.8	8.3	0.0	0.0	71.7	10.2
CAR (M)	0.3	0.2	3.0	0.5	64.9	22.4	0.0	0.0	43.4	38.2
CF (C)	0.3	0.0	1.5	0.1	49.4	3.8	0.0	0.0	136.0	19.3
CF (M)	4.8	1.3	16.3	5.9	64.1	23.7	6.8	0.3	178.0	56.3
CLAD (C)	0.7	0.4	3.9	0.8	51.6	32.8	0.2	0.0	212.0	36.6
CWM (C)	0.8	0.2	1.5	0.1	38.8	4.4	0.1	0.0	1110.0	36.1
CWM (M)	0.3	0.1	2.5	0.4	61.2	21.8	0.0	0.0	1390.0	301.4
DGC (C)	0.5	0.3	1.4	0.1	47.8	11.9	0.1	0.0	75.8	12.2
DGC (P)	0.7	0.1	19.7	5.6	58.4	16.2	0.0	0.0	226.0	29.5
DGC (S2)	0.8	0.2	31.7	8.6	48.1	22.5	0.0	0.0	111.8	55.0
DP (C)	1.2	0.0	0.9	0.1	85.4	1.7	0.1	0.0	108.7	7.6
ECT (M)	197.7	133.0	17.8	5.9	5406.7	2737.2	0.0	0.0	11483.3	9375.3
ECT (C)	6.6	1.6	6.8	1.0	596.0	173.9	0.5	0.6	478.5	217.1
SH (C)	8.9	8.0	1.8	1.3	519.3	230.3	0.2	0.1	447.7	369.7
SH (M)	105.3	35.2	18.3	3.6	1225.3	305.9	5.0	1.8	3620.0	334.1

	Cr		As		Al		V		Mn	
	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD
CAR (C)	28.5	3.8	109.1	13.1	16433.3	981.5	40.6	1.0	677.7	60.3
CAR (M)	4.5	3.2	383.3	137.2	7350.0	2830.0	9.7	5.1	532.0	177.5
CF (C)	17.1	3.1	65.5	14.6	15200.0	2128.4	43.1	3.6	728.7	279.4
CF (M)	18.8	4.3	7290.0	2263.8	11963.3	2276.8	104.0	17.7	1520.0	235.8
CLAD (C)	16.8	3.5	130.2	49.3	6633.3	1773.8	37.3	6.9	568.7	233.5
CWM (C)	23.6	2.1	20.1	1.0	18166.7	1415.4	25.7	1.6	1057.0	159.9
CWM (M)	22.3	1.3	75.9	24.6	19375.0	434.9	29.8	1.6	5687.5	2613.6
DGC (C)	28.2	2.3	315.7	8.5	21466.7	2916.0	40.4	3.7	642.0	48.1
DGC (P)	15.7	3.3	3580.0	1361.1	10126.7	1726.3	29.0	3.1	861.7	328.4
DGC (S2)	9.8	6.5	4343.3	2748.2	6330.0	2650.7	24.0	10.4	864.3	480.4
DP (C)	39.0	2.7	23.0	2.6	17166.7	1616.6	45.7	2.2	1120.0	26.5
ECT (M)	13.5	5.2	48.6	7.0	5030.0	3998.7	24.1	18.5	3890.0	3741.2
ECT (C)	16.0	0.1	36.9	2.1	4710.0	42.4	32.5	7.5	609.0	5.7
SH (C)	27.6	3.3	43.6	9.3	16100.0	2078.5	36.1	3.2	619.7	433.7
SH (M)	13.0	10.0	355.3	302.3	9360.0	3145.7	25.6	4.6	2796.7	302.9

Table 5.1 Soil metal and metalloid concentrations at former mine sites (M, P or S2) and proximal control sites (C) from across the UK. N = 3.

	Fe		Cu		Zn		P		Ca	
	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD
CAR (C)	33400.0	721.1	119.7	17.7	155.7	2.1	1326.7	70.2	1943.3	245.4
CAR (M)	24300.0	3659.2	1416.0	1214.2	66.9	22.3	253.2	320.2	1567.7	1606.2
CF (C)	33633.3	5672.2	20.2	2.2	103.8	9.2	2153.3	228.1	735.3	207.1
CF (M)	44600.0	10911.9	355.3	117.6	431.7	167.5	1393.3	259.3	4913.3	600.4
CLAD (C)	43200.0	9959.4	311.0	53.1	152.3	51.6	961.0	201.5	1607.3	2409.9
CWM (C)	38333.3	2437.9	47.8	3.5	287.3	28.0	1290.0	252.4	1237.0	503.2
CWM (M)	63200.0	10088.9	31.6	4.1	158.3	14.0	785.0	51.5	701.0	356.1
DGC (C)	43066.7	1040.8	92.9	4.3	152.0	36.1	1113.3	102.1	2086.7	220.3
DGC (P)	55733.3	14785.6	869.0	209.9	201.7	31.6	843.7	147.0	6200.0	1973.2
DGC (S2)	88833.3	50966.5	2040.0	497.3	266.7	58.7	555.7	175.8	7713.3	1140.0
DP (C)	24933.3	1050.4	42.6	2.9	797.7	58.4	2070.0	55.7	12766.7	450.9
ECT (M)	20633.3	16772.1	706.0	112.6	15180.0	10471.4	1072.0	196.3	108666.7	78117.4
ECT (C)	21550.0	3464.8	406.5	98.3	638.0	113.1	1745.0	530.3	112000.0	35355.3
SH (C)	26900.0	3751.0	42.8	29.0	968.7	747.0	1963.3	1002.9	6753.3	2556.4
SH (M)	52100.0	9796.4	51.8	12.5	13136.7	7094.6	2210.0	977.5	14756.7	12029.3

	Ti		Co		Ni		Se		Mo	
	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD	mg/kg	± SD
CAR (C)	88.8	23.6	8.6	1.6	15.4	0.8	0.9	0.0	1.5	0.1
CAR (M)	23.6	18.5	7.3	8.0	4.5	3.3	0.0	0.0	1.5	0.6
CF (C)	67.7	10.8	5.9	2.1	9.6	0.2	1.7	0.2	2.9	0.4
CF (M)	212.3	45.5	16.5	3.5	7.6	2.5	1.2	0.3	13.4	2.5
CLAD (C)	20.9	5.3	53.2	28.2	1385.0	613.2	13.2	5.8	3.6	0.9
CWM (C)	33.9	6.9	15.5	0.3	25.8	2.2	0.3	0.0	0.9	0.1
CWM (M)	14.6	0.8	64.5	24.0	23.6	1.5	1.3	0.5	2.0	0.4
DGC (C)	59.1	16.6	12.8	0.4	21.8	0.2	1.5	0.1	1.2	0.1
DGC (P)	54.8	3.4	21.3	5.3	19.1	3.2	1.9	1.3	1.8	0.3
DGC (S2)	97.9	43.1	37.9	17.7	24.5	5.9	4.0	2.4	3.3	0.5
DP (C)	41.5	3.3	16.1	0.9	45.1	15.3	1.2	0.2	1.7	0.1
ECT (M)	22.7	3.0	15.4	7.8	41.0	20.3	5.6	2.2	21.6	13.0
ECT (C)	23.4	6.9	9.1	0.8	29.5	0.1	2.0	0.3	7.1	1.3
SH (C)	99.7	25.8	8.5	2.7	21.5	4.9	0.6	0.1	1.1	0.7
SH (M)	43.3	8.4	14.0	4.9	32.3	10.0	1.1	0.2	4.3	0.8

Table 5.1 (continued) Soil metal and metalloid concentrations at former mine sites (M, P or S2) and proximal control sites (C) from across the UK. N = 3.

		As	Zn	Cd	Ba	Cu	Ca
A	pearson's	0.045	0.126	0.119	0.146	0.321	0.444
	P	0.878	0.668	0.686	0.619	0.263	0.112
B	pearson's	-0.045	-0.126	-0.119	-0.146	-0.321	-0.444
	P	0.878	0.668	0.686	0.619	0.263	0.112

Table 5.2 Correlation of toxic units (TUs) with the proportion of earthworms belonging to clades as defined by phylogeny via mitochondrial COII sequences as depicted in Figure 4.3. α = 0.05.

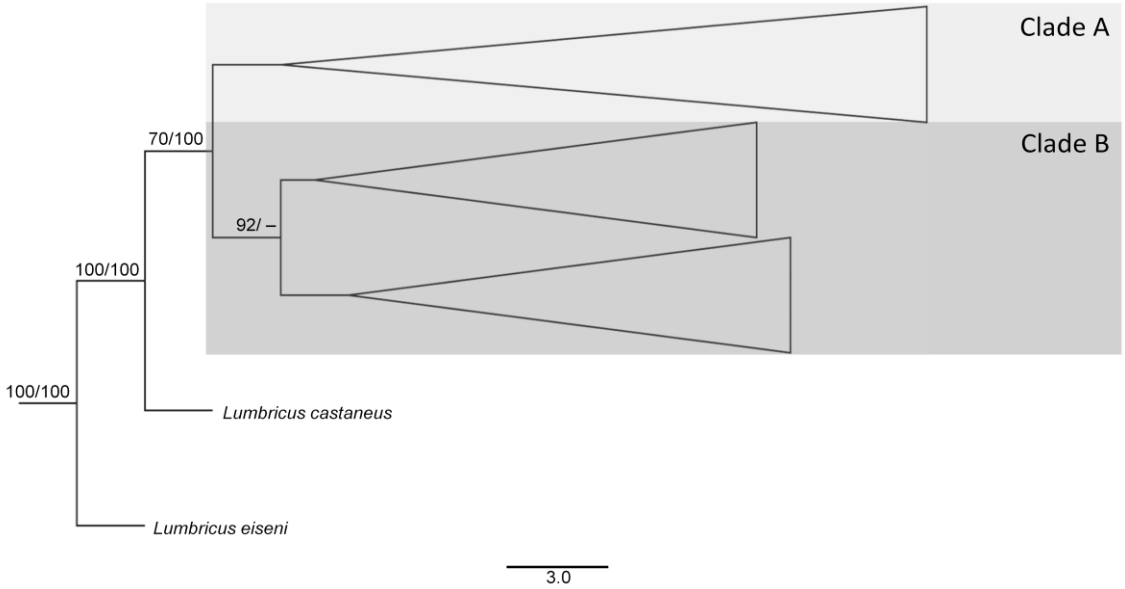
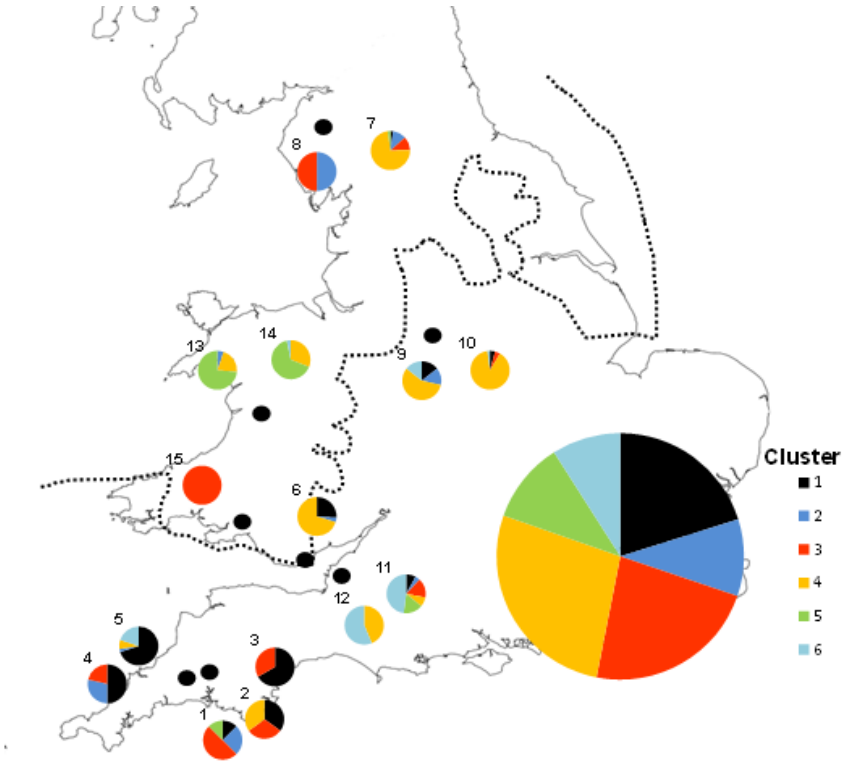


Figure 5.3 Consensus phylogenetic tree (50% majority rule) derived from Bayesian and Maximum Likelihood (ML) models describing the relationship of *L. rubellus* found at former mine and control sites across the UK, using COII mitochondrial gene sequences. Two clades with greater than 7.5% sequence variation were identified belonging to 2 previously identified genotypes, A and B (Andre et al. 2010). Sequences from *L. eiseni* and *L. castaneus* were used as outliers and values lower than 70% were condensed. Bootstrap values are located above respective nodes and define ML and Bayesian methods, respectively. Hyphens indicate where a node was not recovered in respective discrete analyses.



DGCC	1
DGCP	2
DGCS2	3
CARC	4
CARM	5
DPA	6
CFM	7
CFC	8
ECTC	9
ECTM	10
SHM	11
SHC	12
CWMC	13
CWMM	14
CLAD	15

Figure 5.4 The proportion and distribution of haplotype clusters representing genetic variation in populations of *L. rubellus* across former mine sites and proximal control sites. Clusters derived from BAPS v5 analysis of the mitochondrial COII gene assigned to populations are represented using pie charts. Numbers correspond with each sampling location as defined in the table to the right and are observable specifically relative to nearby circular black location points. The large pie chart represents all 347 earthworms defined in Figure 5.2, where clade A represents clusters 1, 2 and 4, while clade B represents clusters 3, 5 and 6.

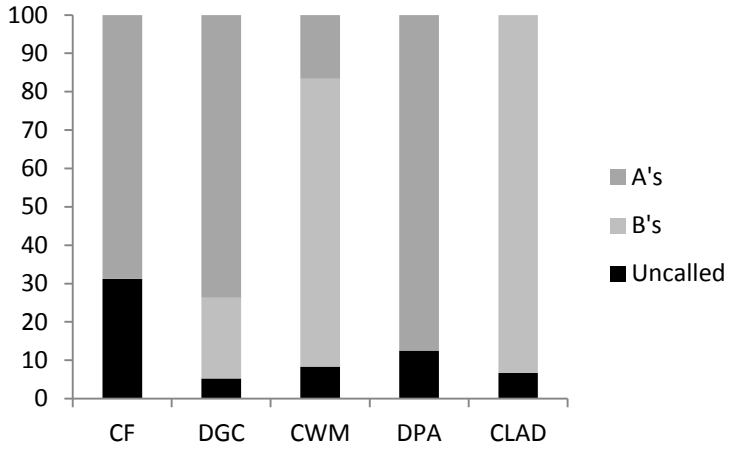


Figure 5.5 The proportion of clades that adult *L. rubellus* taken from former mine and control sites from the UK belong, defined by Maximum Likelihood phylogenetic tree of the COII mitochondrial gene. Uncalled individuals are those for which sequencing failed.

robust. Of the split seen in clade B, 7 of the mine and control sites are inhabited exclusively by B1 (top), whereas both CWM sites have only B2. Average distance within these groups is 0.011 and 0.006, respectively, with an average (0.22) and maximum difference (0.028) indicating that these groups are extremely similar. This variation is not associated with contaminant (r = 0.193, P = 0.492).

There was no correlation between total TUs and the proportion of earthworms relative to clade at each site (r = 0.111, P = 0.695). Moreover, there was no correlation between clade occurrence specific contaminant (Table 5.2), indicating that neither clade is associated with a specific contaminant. Further analysis using BAPS presented sequences representing populations to fall within 6 discrete clusters, of which no single cluster was associated with a single discrete site or contaminant (Figure 5.4). Only a single site, CLAD, was composed of one cluster. Overall, clusters representing lineages A and B accounted for roughly half or all earthworms collected. Sites beyond the border of the former glacial boundary demonstrated no significant difference in the number of clusters occurring at sites relative to previously unglaciated territories ($P > 0.05$). It is also observable that no single cluster is associated with a specific contaminant as defines sites as in the PCA.

When considering the genetic variation of individuals relative to geochemical profiles, it was observed that variation seen among populations, considered to have similar profiles, explained 35% of the variance, but the most variation (63%) was significant within populations ($P < 0.001$) (Table 5.3). This is echoed across the other comparisons, whereby intraspecific variation among populations significantly described the majority of observable genetic variation at discrete sites (62%), as well as among mine and control sites (71%) ($P < 0.001$). The variation among the populations of groups also explained a significant proportion of the genetic variation, though reduced relative to that within populations. A significant level of variation was also observable among groups when considering discrete sites (32%, $P < 0.001$).

5.3.3 Breeding Analysis

There were a total of 46 pairs during inbreeding experiment that incorporated 5 of the sites from the phylogeographic study in an attempt to produce pairs of the same lineage. The sites for which pairs were established were geographically distinct including the most northerly site (CF) and also a south-westerly (DGC) and mid-westerly (CWM) location. Analysis of COII sequences conducted post pairing, indicated that in fact 6 of the pairs were incorrectly matched (Figure 5.5). Missing genotypes (n=12, Uncalled” in Figure 5.5) were assumed to be the most common for each site

Table 5.3 Partitioning of genetic variation between groups by AMOVA, as determined via grouping relative to **A**) 3 groups, where populations are defined as “CFC, CFM. CARC, CARM, DGCC, DGCP and DGCS2”; “ECTC, ECTM, SHC and SHM”; “CWMC, CWMM, DPA, CLAD” **B**) X groups, where populations are defined as “CFC and CFM”; “CARC and CARM“; “DGCC, DGCP and DGCS2”; “ECTC and ECTM“; “SHC and SHM”; “CWMC and CWMM“; “DPA“; “CLAD” **C**) 2 groups, where populations are defined as “CFC, CARC, DGCC, ECTC, SHC, and CWMC”; “CFM, CARM, DGCP, DGCS2, ECTM, SHM, and CWMM”. D.F. is degrees of freedom, α = 0.05 and permutations = 1023.

	Source of variation	D.F.	Sum of squares	Variance	Variance (%)	*P*
A	Geochemistry					
	Among groups	2	151.59	0.1149	1.65	0.32845
	Among pops	15	757.231	2.45813	35.32	0
	Within pops	336	1474.02	4.38696	63.03	0
	Total	353	2382.842	6.95999		
B	Discrete sites					
	Among groups	7	791.445	2.23107	31.51	0
	Among pops	10	117.377	0.46311	6.54	0.00196
	Within pops	336	1474.02	4.38696	61.95	0
	Total	353	2382.842	7.08115		
C	Mines vs. controls					
	Among groups	1	60.085	0.1611	2.32	0.24438
	Among pops	14	545.33	1.82833	26.28	0
	Within pops	289	1435.458	4.96698	71.4	0
	Total	304	2040.872	6.95642		

among, where 10 represent sites with 100% single clade occupancy. Of the 40 correct pairings, all produced viable cocoons to F2. Of the 6 incorrect pairs (n=3 from both CWM and DGC), only 3 produced cocoons, none of which hatched. The data suggests that the lineages are sexually incompatible and so support the identification of the two *L. rubellus* clades as separate species.

5.3.4 RADseq SNP phylogeny

A total of 71,624 and 53,284 concatenated SNPs (including no-calls) were found when among populations for both clades A and B, respectively. These SNPs were used to infer the phylogeography of populations inhabiting former mine sites and the associated proximal control sites. An analysis that concatenated almost 8,000 SNPs from loci derived from all populations at 10X coverage, shared by at least 12 of those populations, resulted in a tree with an almost identical topology from that constructed using over 84, 000 SNPs derived from markers shared between at least 2 populations with 20X coverage, using the maximum likelihood method.

In Figure 5.6, it can be seen that populations cluster in a very similar manner to chemical profiles. This occurs relative to site primarily, and then by contaminant, with the exception of the DGCP site which clusters with the worms from the sites dominated by Zn and Pb in soil. ECT and SH, each heavily contaminated with Pb, Zn, Ba and Cd cluster together on one side of the tree. Populations from CWM clusters more closely with the DPA and CLAD control sites with these sites all closely embedded within the group of sites that are associated with the ECT and SH populations. Another group of sites appear to cluster relative to As and Cu contamination that occurs at DGC, CAR and CF. Within this grouping the worms collected from DGC show the greatest divergence from the remaining populations. This includes the worms collected from both polluted location at this site, as well as the proximal reference site for DGC.

The fact that the control sites cluster closely with proximal populations inhabiting contaminated land reflects the long-term influence of basal geochemical properties on the genome as reflected in tree topology. This may also be observed in global Fst values of control populations and relative mine site populations, which are an average of 7.9% (maximum of 9%), both within and between clades. The different clades that occupy mine sites are more similar, reflected by an average Fst of 5.4% (maximum of 7%). The greatest variation observed occurred between populations at DGC and those at DP and SH (up to 42.8%). Several control sites from clades A and B have relatively low Fst values, demonstrating relatively low genetic variation despite large geographical distances. SNPs with an Fst of 1 and occurrence of at least 50% in both

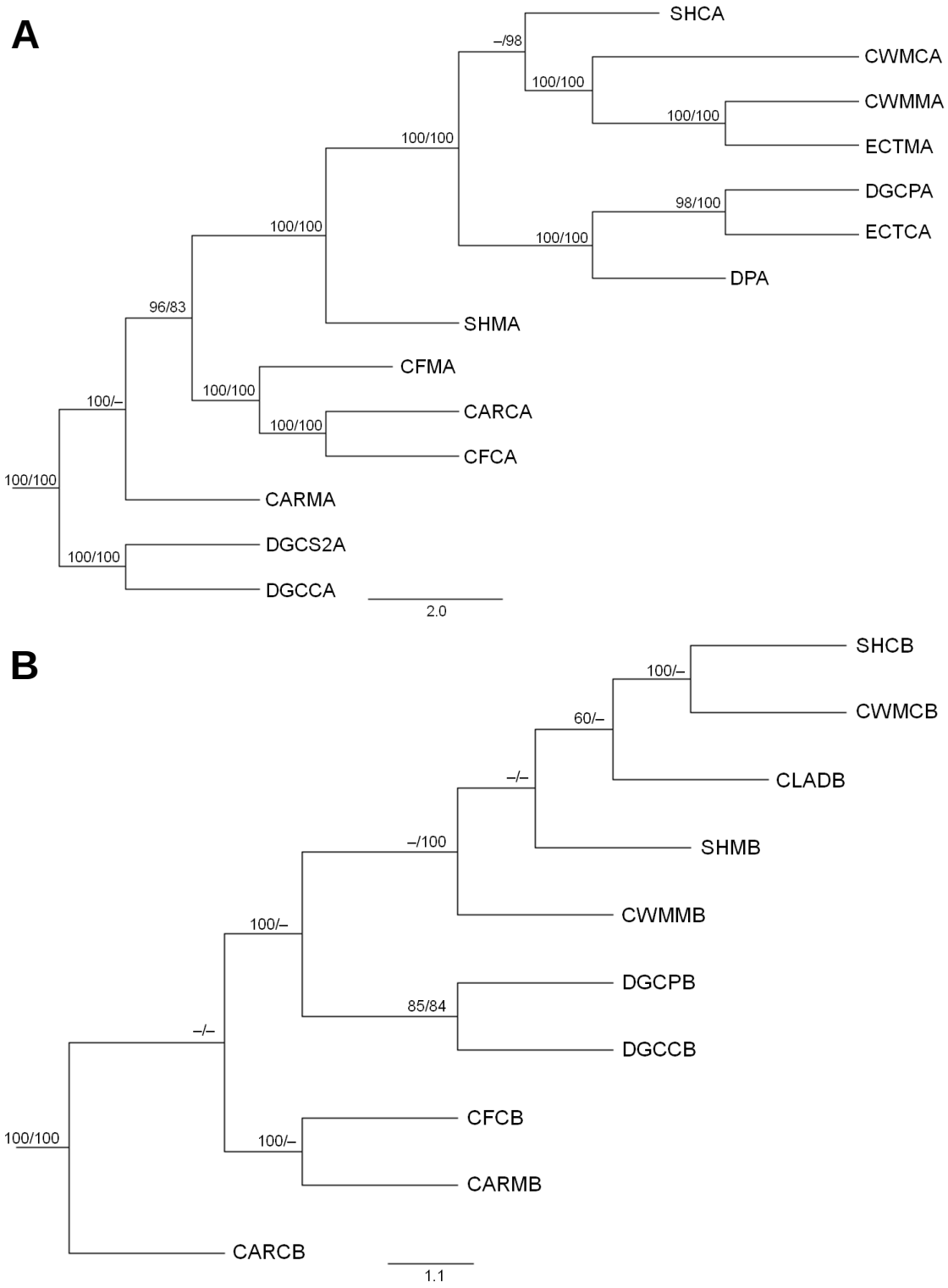


Figure 5.6 Consensus phylogenetic tree (50% majority-rule) derived from Bayesian and Maximum Likelihood (ML) models describing the relationship of *L. rubellus* found at former mine and control sites across the UK. Trees were constructed from pooled individuals belonging to clade A (**A**) and clade B (**B**), using 71,624 and 53,284 concatenated SNPs, respectively. Bootstrap values are located above respective nodes and define ML and Bayesian methods, respectively. Hyphens indicate where a node was not recovered in respective discrete analyses.

clades were assessed as lineage specific markers. A total of 16 SNPs, representing 11 contigs were valid, though none represented all populations. Only 4 contigs had matches in BLASTX, 3 were hypothetical and one was apportion of a gene for acetyl-CoA, though all pertained to E-values greater than 1 (Altschul et al. 1997).

SNPs and the number of unique loci are strongly correlated among clades A ($r= 0.924$, $P < 0.001$) and B ($r = 0.831$, $P = 0.001$). It was found using ANOVA that there was a significant difference between the number of unique loci found in clade A earthworms at mine and control sites, ($P = 0.048$, $F = 4.84$), though there was no significant difference between those in lineage B ($P = 0.863$, $F = 0.03$). Moderately more loci were observed in control populations relative to those inhabiting contaminated soil at DGC, among earthworms of clade B and more prolifically, clade A. Diversity is similarly varied among populations of clade A at SH, however the converse is true of clade B populations at CWM and those from clade A from ECT, whereby more loci were recorded among those in contaminated soil. The difference between clade B earthworms at CWM is substantial, with a difference of 8,576 loci occurring between control and mine site populations, whereas clade A earthworms from ECT differed by 2,113.

5.4 Discussion

Here, use of a common mitochondrial marker firmly establishes the occurrence and distribution of clades that contribute to the *L. rubellus* species complex across the UK. Genetic variation across the genome relative to populations of each of the clades further eludes to selection relative to soil geochemistry, which has important evolutionary and ecological implications.

Initially, COII sequencing provides a general overview that there are two distinct clades of *L. rubellus* that occur frequently at contaminated and uncontaminated former mine sites. Such cryptic species are therefore common-place in this species and supports previous work that identifies the diversity that occurs among morphologically definable earthworm species (King et al. 2008). Conserved tree topology using a mitochondrial marker remained mostly undefined beyond clade groupings as a result of the high degree of intra-specific clade similarity. Historically recent diversification of the clades would provide markers that lack resolution in this way, warranting the use of genomic markers. Li et al. (2010) incorporated whole mitochondrial genome sequencing of *Bombyx mori* and were similarly restrained in that they observed a lack of resolution between distinctive, but recently diverged, natural wild Chinese and domesticated

populations. Sexual incompatibility between the two clades enforces the extent of crypsis and is demonstrated in the breeding analysis, suggesting that clades A and B are discrete species that are unable to produce hybrids at F1. Andre et al. (2010a) found that hybrids occur at CWM, but further analysis would need to demonstrate that they are able to contribute towards the gene pool. Their role in adaptation may be vital, as described in butterflies of the genus *Heliconius*, whereby hybridisation, even at low levels, can confer adaptive traits (Dasmahapatra et al. 2012).

Furthermore, no geochemically specific lineage occurrence was observed using mitochondrial markers. Andre et al. (2010a) previously used AFLP markers to demonstrate the occurrence of clades A and B independently among similarly contaminated soils at CWM, while they occurred simultaneously across proximal sites that were less contaminated. Specific sites did see dominance by a single clade, which may be demonstrative of alternative selective pressures at such locations. BAPs analysis was able to derive a more strongly deterministic Bayesian analysis of mitochondrial variation, whereby all individuals belonged to 6 clusters, however, there was no distinctive pattern between the clusters occurring at each site. In addition, RADseq was used to derive a wide-range of genomic markers to further resolve variation in populations of *L. rubellus* relative to soil geochemistry. No lineage specific SNPs were discernible in analysis of markers via RADseq, as even the most prominent candidate loci did not occur among all populations. Candidate markers tended towards what may be assumed to be non-coding regions. These parts of the genome may be considered to be less prone to recent selection events, however, there is mounting evidence that they may be under selection (Nevo 2011).

Previously, RADseq has allowed Emerson et al. (2010) to distinguish the fine-scale genetic relationships of pitcher plant mosquito (*Wyeomyia smithii*) populations following receding glacial boundaries. In this analysis, distinction of populations relative to a former glacial boundary was not observed. Instead, RADseq has revealed that most populations under similar geochemical profiles cluster phylogenetically and is true for both worms belonging to both clades discernible through COII sequencing. Specifically, sites where underlying geology has provoked mining of lead and zinc has resulted in SNP distributions among populations of earthworm that differ from those inhabiting soils established from geologies that support Cu and As extraction. This is indicative of wide-spread genomic variation that occurs relative to local environments and is synonymous with parallel evolution across the clades. This observation is most compelling among earthworms at DGC and CF, which endure high concentrations of As and occur over 600 km from each other. Intriguingly, proximal contaminated and

control site populations were generally closely related. This suggests the major component of genetic variation is driven by the long-term influence of variation in natural soil chemistry and that anthropogenic influences related to pollution are overlaid on this background of genetic divergence.

Analysis of variation observed in mitochondrial COII gene sequences supports the potential for parallel evolution within the species complex, whereby geographically isolated populations appear to show similar genetic variation indicative of the same adaptive traits. Better resolved analysis of individuals at sites studied here may demonstrate the potential for similarities in genetic variation between clades, relating to adaptation. An analysis similar to this has already demonstrated similarities between species of Cameroonian mosquito that are adapted to climactic variation. Analysis of *Anopheles gambiae* populations has shown that the species is definable through specific chromosomal inversions that cover megabases and are the product of selection relative to aridity (Cheng et al. 2012). A closely related species (*A. funestus*) that occupies many of the same areas has been shown to demonstrate co-localisation of similar sets of genes among inversions and implicates the importance of genome architecture towards the development of adaptation.

Populations can also be seen to cluster relative to soil chemical profiles across analysis of SNPs in each of the clades, which is indicative of adaptation to geochemistry rather than population history. Different discrete populations of the same clade are able to inhabit sites that impose a strong selective pressure via similar prolific toxicants, which is suggestive of parallel evolution. Other analogous **e**xamples of these phenomenainclude inhabitance of low nutrient, highly contaminated soils by *Arabidopsis lyrata* and populations of *Gadus morhua* occupying temperature clines. A wide-variety of unlinked and linked markers defining candidates for adaptation were observed in each example. The likelihood of such parallel or convergent evolution among populations of *L. rubellus* is reduced with respect to the number of possible solutions to impositions by selective pressure (Wood et al. 2005). When a single trait is the most efficient or only means of adaptation, it is increasingly likely to occur between highly divergent lineages (Bollback and Huelsenbeck 2009). Research using the terrestrial arthropod, *Orchesella cincta*, has found a genetic basis for adaptation to Cd, which is largely considered to occur as the result of variation in a metallothionein (Mt) promoter, leading to increased excretion through shedding of the midgut epithelium (van Straalen et al. 2011). In other cases, metal resistance in *Drosophila melanogaster* has been attributed to Mt duplication (Maroni et al. 1987), whereas polymorphisms within the gene have been attributed to degrees of tolerance in *Crassostrea gigas*

(Tanguy et al. 2002). With respect to these points, future analyses would benefit from assembling linkage maps and through determination of specific regions of variation throughout the genome, acting to elucidate evolutionary and functional mechanisms for adaptation, both within and between clades (Hohenlohe et al. 2010).

Based upon the nature of the adaptation, one would expect to see selection that makes use of underlying genetic material, as classically described by Alhiyaly et al. (1993) who observed tolerance of grass occupying soil contaminated with metals following the installation of electric pylons. Here, phylogenetic topology is maintained between local populations when the number of markers is massively reduced, suggesting that a number of loci are responsible for genetic differentiation between populations, which is synonymous with polygenic adaptation mentioned in Chapter 1 (Pritchard et al., 2010). If this is found following a more highly resolved analysis of individuals in natural populations, one would expect to see mixing of tolerant and non-tolerant individuals on both contaminated and non-contaminated land. One may also expect to see a range of genotypes if the trait is polygenic, reflected in recent research by Kille et al. (Submitted) where different populations from discrete locations at DGC are examined for resistance to As. Certain populations that reside in soils with the largest concentrations of As are more resistant than those that inhabit less contaminated soil, possibly suggesting less stringent selection for adaptive traits. What's more, the highly heterogeneous terrestrial environment potentiates the occurrence of local refugia that may maintain unadapted earthworms. This should be particularly effective among epigeic species that can occur throughout leaf litter (Per. Obs.). The effects of such are exemplified in control measures for the populations of the pest *Helicoverpa armigera*, which are developing resistance to genetically modified crops with inceticidal properties (Wu et al. 2008). Refugia, i.e. normal crops that don't impose selection pressure for insecticidal resistance, allow unadapted individuals to persist nearby to modified crops, acting to reduce the frequency of adaptive alleles within the metapopulation. This is proving particularly effective for a recessive trait and would strongly affect a polygenic trait.

Phylogenetic analysis of COII sequences using BAPS demonstrated that higher levels of intraspecific variation were observed at sites. Although the reasons for this are unclear, it may possibly reflect the highly heterogeneous environment inhabited by earthworms, generally. A minor, yet significant degree of genetic variation is found between populations relative to contaminated and uncontaminated sites as well as relative to geochemistry, potentially signifying that certain selective pressures may more strongly affect genetic variation than others, though site history may also have a role. Similar findings have been observed in *C. elegans* and *D. melanogaster*, which

see recent selective sweeps relative to human movement that acts as basis for global genetic variation (Andersen et al. 2012; Haddrill et al. 2005). In this instance, earthworms have been found to be affected through anthropogenic activity, but may be considered to be a more reliable model for assessing adaptation in natural populations as low migratory rates of this species should not impact analysis of genetic variation across isolated communities. There may, however, be evidence for movement within populations seen at the heavily contaminated site DGC. Earthworms belonging to clade A at DGCP were observed to cluster more closely with populations occurring at sites contaminated with Pb, Cd and Zn. DGCP is proximal to a path used by walkers and has been extensively visited for earthworm sampling by researchers. Other sites have roads and paths nearby; therefore, it is surprising that potential evidence for movement of populations is not more frequently observed. It is therefore possible that this may reflect alternative selective pressures unrecognised in this study, which could also have a role in the rare occurrence of sites where only a single clade is found (Figure 5.2). For the most part, populations occur relative to underlying geochemistry and reflect that human movement has had little observable effect upon major genetic variation observable in highly resolved phylogeographic analysis.

Genetic erosion was potentiated following the reduction of haplotypes that occurred at in a previous analysis of populations at CWM, suggesting that there is a high degree of spatially localised genetic differentiation (Andre et al. 2010a). Loss of genetic variation has been observed in other studies through use of microsatellites, whereby the presence of alleles were quantified, reflecting genetic erosion among communities exposed to pollution (Athrey et al. 2007) and subjected to habitat fragmentation (Srikwan and Woodruff 2000). Within this investigation, the number of distinct loci occurring in a standardised number of raw sequences was quantified, representing population-wide genetic variation. Within this analysis, it was found that populations belonging to clade A, which inhabited highly contaminated former mine sites, demonstrated less variation relative to proximal control sites. This is supported by a similar analysis of sites across DGC by Kille et al. (Submitted), who synonymously demonstrated a loss of genetic variation in clade A populations relative to soil As concentrations. The authors also demonstrated that there is no significant variation among earthworms belonging to clade B, citing instead, a loss of epigenetic variation. This is further suggestive of parallel evolution and provides a compelling scenario for further investigation.

The similarity of earthworms occupying contaminated land to proximal control populations based upon adaptation to local geochemistry is suggestive that

earthworms at former mine sites are derived from proximal populations. What is viewed as genetic erosion in this instance may in fact reflect an open niche for adaptive genotypes that occur naturally in populations at low frequency. Phylogenetic analysis of a large number of individuals using a large number of genomic markers, as provided by RADseq, would provide insight regarding whether individuals from control sites cluster with those on mine sites. In addition, preliminary data here informs the field of the importance of more broadly defining the occurrence of hybrids among natural populations, as this would greatly influence our understanding of ecological and evolutionary mechanisms underlying adaptation to metal contamination in an important environmental sentinel.

5.5 Conclusion

British populations of *L. rubellus* can confidently be defined as a species complex composed of two clades. Breeding analysis has indicated that these species likely represent isolated breeding cohorts and that the two are not obviously differentially sensitive to metal pollution. Local soil conditions are shown to impose a sufficiently strong selective pressure upon populations to affect genome-wide variation. Clustering of soil geochemical profiles of geographically distant sites (at both control and contaminated sites) that are synonymous with resident populations of the same sites clustering in phylogeographic analyses is indicative of convergent evolution. The support of mitochondrial phylogeny distinguishing clades as distinct further supplements this conclusion. A high degree of intraspecific genetic variation in mitochondrial sequences was also observed and may relate to the heterogeous nature of terrestrial environments or site history.

Genetic erosion was determined through the number of distinct loci occurring in a standardised number of raw sequences from RADseq analysis. This distinguishes that populations from clade A suffer from a reduced genetic diversity as a result of inhabiting highly contaminated sites, relative to populations inhabiting proximal control sites. Genetic variation among individuals, which putatively reflects adaptation to local environments, could better resolve ecological mechanisms underlying adaptation to soil geochemical profiles. Identification of adaptation to specific contaminants and associated candidate loci will require better resolution at the level of individuals from specific sites.

Chapter 6

General Discussion

The aim of this thesis was to recognise the effects of arsenic (As) upon *Lumbricus rubellus*, as this species is a key indicator of ecosystem health. Chapter 1 established that selection pressures, such as those from long-term anthropogenic metal contamination, can cause populations to develop genetic bases for resistance, which allows for maintained fitness. Basal mechanisms affected by environmental perturbation must be sufficiently described to differentiate unadapted populations from those that are resistant. Within the current body of work, a hierarchical approach was taken towards determining the range of phenotypically plastic responses that contribute towards mitigating As toxicity among earthworms that were previously naive to exposure. Primarily, life-history endpoints were used to distinguish phenotypic variation associated with exposure as well as to anchor a range of holistic molecular endpoints able to elaborate upon underlying mechanisms associated with As uptake, metabolism and toxicity. Collectively, these analyses acted as a basis for implicating variation seen relative to that seen among genetically resistant natural populations.

Specifically, short-term responses to As exposure were observed in Chapter 2, whereby variation in life-history parameters were derived from adult and juvenile *L. rubellus*. These results were used to inform a model for defining the ecological implications of As contamination on unadapted populations, distinguishing a relatively low soil As concentration in which they would fail to persist. This is relative to natural populations that have previously been documented to occur among soils contaminated with extremely high concentrations of As. The low concentration found to affect unadapted populations infers that those at highly contaminated sites have adopted mechanisms that act towards better mitigation of As toxicity.

In conjunction with understanding variation in life-history parameters in response to As exposure, suitably anchored molecular end points can provide a more highly resolved appreciation of As metabolism and toxicity in an environmental sentinel such as *L. rubellus*. A systems biology approach was used for this assessment, interweaving -omic analyses to derive a highly resolved overview capable of inferring a range of molecular mechanisms associated with phenotypic variation relating to As exposure. The findings of these endpoints are well defined in 3.4, but are summarised herein.

Speciation analysis using HPLC-ICP-MS concluded that As primarily remains in its inorganic forms. A significant lack of alternative organic species was noted, potentially highlighting the role of gut and soil microbiota in providing an alternative means for As transformation (Button et al. 2012; Ritchie et al. 2004). Arsenic speciation was able to highlight potential pathways of interest involved in As metabolism and was used to support mechanisms described by transcriptomic and metabolomic analyses. Analysis using HPLC-ICP-MS also ensured that data were comparable to previous investigations in which it was implemented.

Metabolomic analysis by ^{1}H-NMR was further able to distinguish specific metabolites affected by As exposure in adult *L. rubellus*. Specifically, glycine and methionine were strongly correlated with exposure and are known to be associated with the production of glutathione, implicated previously in the management of As (Hayakawa et al. 2005; Nemeti et al. 2012). Transcriptomic analysis was implemented to elaborate upon metabolomic variation, beginning initially with assembly of the *L. rubellus* transcriptome. This has acted as an extremely useful tool, forming the basis for gene discovery while providing valuable information that can be used to support the assembly of the draft *L. rubellus* genome**.** Transcriptomic assembly was followed by an initial analysis of the effects of As exposure upon differential expression, supporting insights derived from metabolomic as well as tissue chemistry data. Primarily, the combined analyses support the role of glycine and methionine in the reduction of pentavalent arsenate to trivalent arsenite, as well as highlighting the potential role of glutathione as a prominent mechanism for managing cellular As. The results also support the novel finding that phytochelatin, derived from glutathione, is actively involved in As metabolism. Iron metabolism was also implicated by metabolomic and transcriptomic analyses. Although this finding is supported within the literature, the role of iron in As metabolism has yet to be explored in earthworms. However, its role may actually be implicated in resistant phenotypes found across Devon Great Consols (DGC), where earthworm colouration is suggestive of products derived from heme (Piearce et al. 2002), which is known to bind As (Winski and Carter 1995).

Use of a range of –omic analyses provided significant power for discriminating the mechanisms implicated in short-term phenotypic response to As toxicity. A systems biology analysis of As exposure, although technically challenging, was able to identify wide-ranging molecular variation to those observed in life-history parameters. The approach also helps to establish the repercussions of exposure at a level that is relevant to populations. Conversely, molecular endpoints were also capable of distinguishing signals of exposure at concentrations that did not manifest in deviation of

life-history parameters. The potential for identifying low level effects encourages the further implementation of –omics endpoints in future ecotoxicological studies, as this can better respect levels of contamination that can affect populations and supplement investigations that aim to identify the synergistic toxic effects of mixtures (Spurgeon et al. 2010). Ultimately, this portion of work has demonstrated the effects of As exposure upon unadapted *L. rubellus*, reflecting that it is a useful environmental model that is largely comparative against conventional model organisms in at least describing metabolism of a widely-spread environmental toxicant.

Analyses deriving highly resolved life-history and molecular endpoints form the basis for recognising variation that may be associated with As adaptation in natural populations such as those from DGC and Carrock Fell (CF). Populations inhabiting these sites demonstrate maintained fitness in spite of exposure to As, relative to unadapted populations (Langdon et al. 2001). Previous evidence, mostly in the form of chemical analysis of As speciation occurring in tissues of earthworms directly collected from contaminated sites have been used to suggest genetic bases for adaptation (Langdon et al. 2009). However, acknowledgment of the occurrence of specific genetic variation is necessary to resolve specific mechanisms associated with adaptation to As. Recombinant inbred lineages (RILs) were therefore created to establish a range of phenotypes capable of relating to genotypes responsible for resistance. Previous attempts to breed earthworms from DGC and CF have proven unsuccessful, ultimately demonstrating that this is a time consuming and complex enterprise (Langdon et al. 2003b). However, with the successful establishment of RILs, as described within this body of work, a vital resource for distinguishing the affects of population specific phenotypes derived from genomic variation is now available. F2 RILs are currently being maintained at 4 °C, while F0 earthworms sampled from natural populations have been archived at -80 °C, providing opportunities for further genetic analyses.

Recent development of techniques capable of deriving thousands of genome-wide markers provide a means by which analysis of genetic variation, as is necessary for contemporary environmental genomics. Available techniques are reviewed by Davey et al. (2011), however restriction associated DNA sequencing (RADseq) has been specifically implemented to derive genetic variation associated with phenotypic variation. Previous investigations that have analysed genetic variation among RILs using RADseq have enabled the production of linkage maps, which describe the proximity of markers relative to recombination frequency. Linkage maps therefore enable for discrimination of genetic signatures of selection, and identify regions associated with variation relative to phenotypes (Baxter et al. 2011). Specifically,

RADseq has been used to explore genetic variation among butterflies of the genus *Heliconius* (Reed et al. 2011). These species are able to share the cost of associating wing patterns with unplatability among predators through mimicry across species. However, the phenotype varies immensely over short geographical distances in response to selection pressure. Linkage mapping via RADseq has led to the identification of two specific genomic regions that are associated with phenotypic variation (Dasmahapatra et al. 2012). Fine mapping by long-range PCR established that that the these traits are the result of continued gene flow between the species, offering a unique perspective on adaptation to a selective pressure, as well as exploring the defining ecology of the species. Such work has set a standard that future investigations aiming to resolve genetic bases for phenotypic variation should aim to achieve.

Analysis of genetic variation that endows As resistance seen among populations inhabiting highly contaminated former mine sites at DGC and CF, would benefit from implementation of techniques such as RADseq. In previous investigations, analyses implementing the technique have made use of clearly distinguishable morphological variation (Baird et al. 2008; Dasmahapatra et al. 2012; Hohenlohe et al. 2010). As little is known about As resistance in earthworms, such as the costs relating to specific phenotypes or morphological variation associated, it was necessary to determine phenotypic variation among RILs so that genetic variation may be distinguished relative to the ability to mitigate As toxicity. A potentially observable phenotype that clearly describes sensitivity to As exposure is mortality. This endpoint can, however, prove to be difficult to implement in terms of defining suitable exposure concentrations that fully discriminate between resistant and non-resistant individuals. This is especially true if a polygenic trait is responsible for adaptation. What's more, basal variation in size and activity, potentially not affected by variation associated with resistance, may increase susceptibility to toxicity and incorrectly implicate individuals for genetic analysis. Thus it is important to investigate sub-lethal endpoints of As exposure that identify cellular mechanisms suitable for defining adaptation.

A suitable discriminatory method is described in Chapter 4, wherein differential centrifugal fractionation was used to distinguish phenotypes relative to resistance. In F2 earthworms derived from DGC and CF, a range of phenotypes were apparent relative to the distribution of accumulated As in different cell fractions. Cohorts were defined among RILs through identification of cytosolic As concentrations as the main site of As sequestration. The range of phenotypes is suggestive of a multilocus trait

that may differ in its nature among natural populations at DGC and CF, but ultimately describe potential mechanisms for intracellular management of As, maintained to F2. In addition to their use for recognising phenotypes relating to As resistance, RILs were subsequently used to infer the sexual compatibility of clades A and B. Breeding analysis within Chapter 5 was able to provide insight that suggests the two clades are incapable of mixing to produce offspring, although previous evidence by Andre et al. (2010b) is suggestive that hybrids occur. Following recent work in the *Heliconius* butterfly system (Dasmahapatra et al. 2012), where the evolutionary role of hybrid fertility is recognised, further evidence of species mixing may prove useful in recognising the ecology underlying metal resistance among populations of *L. rubellus*. Establishing clade specific cultures and rearing them to F2 will likely be a difficult but rewarding endeavour.

Further in Chapter 5, phylogeographic analysis of a range of populations of *L. rubellus* inhabiting former mine sites, as well as proximal control sites, was used to demonstrate genetic variation associated with that conferring adaptation to metal toxicity. This work was based upon that described by Emerson et al. (2010), who used mitochondrial genotypes to inform further analysis by RADseq to derive thousands of genomic markers from across the genome. The sequencing of a mitochondrial marker across several populations in the UK firmly establishes the occurrence of two distinct clades of *L. rubellus*. The relatively restricted clade compositions of British populations is in contrast to recent work that has determined four additional clades across mainland Europe, each differing by 10 – 15% in cytochrome oxidise I and II mitochondrial gene sequences (Sechi, unpublished). Reduced diversity of clades in the UK has ramifications for evolutionary biology, as one of the British clades (B) has not been found elsewhere, yet is highly divergent from clade A (~13%), which it occurs along-side. Within the UK, it might be expected that certain lineages may dominate particular soil conditions, establishing alternative adaptive capabilities or preferences relating to prevailing soil condictions. However, no underlying relationship between soil geochemical profiles and clade occurrence among British *L. rubellus* was found, as defined in Chapter 5. Although this may be indicative that other environmental factors may partition clades, the fact that they both occur together at many sites enforces that soil physiochemical characteristics are not selective for either clade. Furthermore, preliminary analyses of an effort to more widely sample populations across the UK has yet to distinguish environmental parameters associated with clade occurrence (Spurgeon, unpublished.). This analysis additionally incorporates a wider range of discriminatory methods, encompassing metabolomic endpoints that are used in tandem with a mitochondrial marker and more highly resolved soil and climate variables. Life-

history endpoints in juveniles (Chapter 2) were able to provide initial discriminatory analyses of variation among clades belonging to the *L. rubellus* species complex. Specifically, one clade was shown to mature more quickly than the other, however, there was no significant difference in the affect of As upon endpoints. This may provide insight as to how both of the British clades can be found across the same sites, even when highly disturbed.

Genomic markers were further used to resolve phylogeographic variation in populations of *L. rubellus* across the UK, as described fully in Chapter 5. Briefly, genomic DNA from individuals that were genotyped using a mitochondrial marker were pooled by clade, to derive a genome-wide SNP analysis for field collected *L. rubellus* populations. Using RADseq, it was observed that different discrete populations of the same clade inhabited sites relative to underlying soil geochemistry. This is suggestive of parallel evolution, as geographically distinct but genetically similar populations were found to occur across sites that were similarly contaminated. Earthworms found at polluted sites most often clustered phylogenetically with those from proximal control sites, suggesting that adapted individuals are likely derived from background genetic variation in unadapted populations. There was also an indication of genetic erosion among earthworms from mine sites belonging to clade A at mine when compared to control populations, though the effects of this upon population stability are unknown. Observations of genetic erosion are supported by a similar analysis of sites across DGC by Kille et al. (Submitted), who also demonstrated a loss of genetic variation in clade A populations in correlation with soil As concentrations. This work was conducted in parallel with this study and also suggested varying degrees of resistance, as demonstrated by the range of RIL phenotypes described in Chapter 4. In addition, the authors also demonstrated that there was no significant variation among earthworms belonging to clade B, citing instead a loss of epigenetic variation. Current methodological advances permit analytical methods for probing the role of epigenetic variation in discrete populations relative to As exposure. As suggested in Chapter 1, the flexible nature of epigenetic modifications means that their transference into subsequent generations may also occur – although this aspect remains controversial. -Omics endpoints, such as the methylome of *L. rubellus* under As exposure, may have important implications for ecotoxicological research as they permit better resolution of toxicity, capable of establishing the temporal significance of exposure (Reichard and Puga 2011). However, determining epigenetic variation within natural metazoan populations remains an as yet unconquered challenge. Future experiments that link epigenetic and transcriptomic variation are an attainable primary research objective that would benefit from transcriptomic resources developed in Chapter 3.

Preliminary analysis of individuals inhabiting DGC and CF, as well as adjacent control sites, using genomic markers derived from RADseq has demonstrated that the most genetically similar earthworms are not necessarily from the same site, nor share the same sensitivity to As. The only variation realised relative to geography was observed among artefact sequences originating from microbial communities inhabiting the nephridia, which are excretory organs. This ultimately demonstrates the power of holistic endpoints in ecotoxicology, providing evidence that they can give insight into phenomena that are otherwise unrecognisable. With respect to the findings of Chapter 5, as well as more broadly within this field, future analyses would benefit from assembling linkage maps and through determination of specific regions of variation throughout the genome, acting to elucidate evolutionary and functional mechanisms for adaptation, both within and between clades.

Broadly, **t**he work presented here recognises the paradigm shift in ecotoxicology, whereby molecular methodology is suitably applicable for non-model organisms as well as for recognising the effects of exposure at multiple levels of biological complexity. Although not definitively resolved within these analyses, this work has laid the foundations for demonstrating genetic variation associated with As resistance, with respect to metabolic processes recognised in unadapted earthworms. This research has the potential to establish *L. rubellus* as a model not only in ecotoxicology, but also evolutionary biology. Future investigations should aim to better resolve genetic variation in individuals derived from natural populations, while also working to uncover the role of epigenetic and microbial variation in As resistance. However, the lack of morphological variation between clades, as well as between adapted and unadapted phenotypes will hinder progress. Slow generation time of many earthworms, including *L. rubellus*, means that the field may lose out on funding to more readily definable systems with more easily available biological resources.

References

Adam O, Degiorgi F, Crini G, Badot P-M. 2010. High sensitivity of *Gammarus sp.* juveniles to deltamethrin: Outcomes for risk assessment. Ecotoxicology and Environmental Safety 73(6):1402-1407.

Agrawal AA. 2001. Ecology - Phenotypic plasticity in the interactions and evolution of species. Science 294(5541):321-326.

Alberch P. 1991. From genes to phenotype - dynamic systems and evolvability. Genetica 84(1):5-11.

Alhiyaly SAK, McNeilly T, Bradshaw AD, Mortimer AM. 1993. The effect of zinc contamination from electricity pylons- genetic constraints on selection for zinc tolerance. Heredity 70:22-32.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25(17):3389-3402.

Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, Felix M-A, Kruglyak L. 2012. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. Nature Genetics 44(3):285-U83.

Andre J. 2008. Earthworm adaptations to metals: Inorganic speciation, biochemical fingerprinting and molecular genetics. Reading: University of Reading.

Andre J, King RA, Stürzenbaum SR, Kille P, Hodson ME, Morgan AJ. 2010a. Molecular genetic differentiation in earthworms inhabiting a heterogeneous Pb-polluted landscape. Environmental Pollution 158(3):883-890.

Andre J, Sturzenbaum SR, Kille P, Morgan AJ, Hodson ME. 2010b. Metal bioaccumulation and cellular fractionation in an epigeic earthworm (*Lumbricus rubellus*): The interactive influences of population exposure histories, site-specific geochemistry and mitochondrial genotype. Soil Biology & Biochemistry 42(9):1566-1573.

Andrew AS, Bernardo V, Warnke LA, Davey JC, Hampton T, Mason RA, Thorpe JE, Ihnat MA, Hamilton JW. 2007. Exposure to arsenic at levels found in US drinking water modifies expression in the mouse lung. Toxicological Sciences 100(1):75-87.

Andrew AS, Jewe DA, Mason RA, Whitfleld ML, Moore JH, Karagas MR. 2008. Drinking-water arsenic exposure modulates gene expression in human lymphocytes from a US population. Environmental Health Perspectives 116(4):524-531.

Andrew AS, Warren AJ, Barchowsky A, Temple KA, Klei L, Soucy NV, O'Hara KA, Hamilton JW. 2003. Genomic and proteomic profiling of responses to toxic metals in human lung cells. Environmental Health Perspectives 111(6):825-838.

Argos M, Kibriya MG, Parvez F, Jasmine F, Rakibuz-Zaman M, Ahsan H. 2006. Gene expression profiles in peripheral lymphocytes by arsenic exposure and skin lesion status in a Bangladeshi population. Cancer Epidemiology Biomarkers & Prevention 15(7):1367-1375.

Arnold BE, Hodson ME, Charnock J, Peijnenburg WJGM. 2008. Comparison of subcellular partitioning, distribution, and internal speciation of Cu between Cu-tolerant and naive populations of *Dendrodrilus rubidus* Savigny. Environmental Science & Technology 42(10):3900-3905.

Arnold RE, Hodson ME. 2007. Effect of time and mode of depuration on tissue copper concentrations of the earthworms *Eisenia andrei*, *Lumbricus rubellus* and *Lumbricus terrestris*. Environmental Pollution 148(1):21-30.

Athrey NRG, Leberg PL, Klerks PL. 2007. Laboratory culturing and selection for increased resistance to cadmium reduce genetic variation in the least killifish, *Heterandria formosa*. Environmental Toxicology and Chemistry 26(9):1916-1921.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. Plos One 3(10).

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. Molecular Ecology 13(4):729-744.

Bang S, Meng XG. 2004. A review of arsenic interactions with anions and iron hydroxides. Environmental Engineering Research 9(4):184-192.

Barreto FS, Moy GW, Burton RS. 2011. Interpopulation patterns of divergence and selection across the transcriptome of the copepod *Tigriopus californicus*. Molecular Ecology 20(3):560-572.

Barriere A, Felix MA. 2005. High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. Current Biology 15(13):1176-1184.

Barton NH. 2000. Genetic hitchhiking. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 355(1403):1553-1562.

Barton NH, Charlesworth B. 1984. Genetic revolutionsm founder effects and speciation. Annual Review of Ecology and Systematics 15:133-164.

Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, Blaxter ML. 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. Plos One 6(4).

Beckonert O, Coen M, Keun HC, Wang Y, Ebbels TMD, Holmes E, Lindon JC, Nicholson JK. 2010. High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues. Nature Protocols 5(6):1019-1032.

Beer SM, Taylor ER, Brown SE, Dahm CC, Costa NJ, Runswick MJ, Murphy MP. 2004. Glutaredoxin 2 catalyzes the reversible oxidation and glutathionylation of mitochondrial membrane thiol proteins - Implications for mitochondrial redox regulation and antioxidant defense. Journal of Biological Chemistry 279(46):47939-47951.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate- a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Methodological 57(1):289-300.

Bernardo J. 1996. The particular maternal effect of propagule size, especially egg size: Patterns, models, quality of evidence and interpretations. American Zoologist 36(2):216-236.

Blaxter M. 2011. Nematodes: The worm and its relatives. Plos Biology 9(4).

Bollback JP, Huelsenbeck JP. 2009. Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. Genetics 181(1):225-234.

Bowen DQ, Phillips FM, McCabe AM, Knutz PC, Sykes GA. 2002. New data for the last glacial maximum in Great Britain and Ireland. Quaternary Science Reviews 21(1-3):89-101.

Briffa M, Rundle SD, Fryer A. 2008. Comparing the strength of behavioural plasticity and consistency across situations: animal personalities in the hermit crab *Pagurus bernhardus*. Proceedings of the Royal Society B-Biological Sciences 275(1640):1305-1311.

Brulle F, Cocquerelle C, Wamalah AN, Morgan AJ, Kille P, Lepretre A, Vandenbuicke F. 2008. cDNA cloning and expression analysis of *Eisenia fetida* (Annelida : Oligochaeta) phytochelatin synthase under cadmium exposure. Ecotoxicology and Environmental Safety 71(1):47-55.

Bundy JG, Sidhu JK, Rana F, Spurgeon DJ, Svendsen C, Wren JF, Sturzenbaum SR, Morgan AJ, Kille P. 2008. 'Systems toxicology' approach identifies coordinated metabolic responses to copper in a terrestrial non-model invertebrate, the earthworm *Lumbricus rubellus*. Bmc Biology 6.

Burgess SC, Marshall DJ. 2011. Temperature-induced maternal effects and environmental predictability. Journal of Experimental Biology 214(14):2329-2336.

Burke MK. 2012. How does adaptation sweep through the genome? Insights from long-term selection experiments. Proceedings of the Royal Society B-Biological Sciences doi: 10.1098/rspb.2012.0799.

Butt KR, Lowe CN. 2007. A viable technique for tagging earthworms using visible implant elastomer. Applied Soil Ecology 35(2):454-457.

Button M, Jenkin GRT, Harrington CF, Watts MJ. 2009. Arsenic biotransformation in earthworms from contaminated soils. Journal of Environmental Monitoring 11(8):1484-1491.

Button M, Koch I, Reimer KJ. 2012. Arsenic resistance and cycling in earthworms residing at a former gold mine in Canada. Environmental Pollution 169:74-80.

Calow P. 1996. Variability: Noise or information in ecotoxicology? Environmental Toxicology and Pharmacology 2(2-3):121-123.

Caswell H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. Sunderland, MA, USA: Sinauer Associates.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: Building and genotyping Loci *de novo* from short-read sequences. G3 1(3):171-82.

Challenger F. 1945. Biological methylation. Chemical Reviews 36(3):315-361.

Cheng C, White BJ, Kamdem C, Mockaitis K, Costantini C, Hahn MW, Besansky NJ. 2012. Ecological genomics of *Anopheles gambiae* along a latitudinal cline: A population-resequencing approach. Genetics 190(4):1417.

Clemens S. 2006. Evolution and function of phytochelatin synthases. Journal of Plant Physiology 163(3):319-332.

Clemens S, Kim EJ, Neumann D, Schroeder JI. 1999. Tolerance to toxic metals by a gene family of phytochelatin synthases from plants and yeast. Embo Journal 18(12):3325-3333.

Cole LC. 1954. The population consequences of life history phenomena. Quarterly Review of Biology 29:103-137.

Corander J, Martinen P. 2006. Bayesian identification of admixture events using multilocus molecular markers. Molecular Ecology 15:2833-2843.

Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, Nadeau NJ, Zimin AV, Hughes DST, Ferguson LC, Martin SH and others. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487(7405):94-98.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12(7):499-510.

Davies NA, Hodson ME, Black S. 2003. Is the OECD acute worm toxicity test environmentally relevant? The effect of mineral form on calculated lead toxicity. Environmental Pollution 121(1):49-54.

de Jong G. 2005. Evolution of phenotypic plasticity: patterns of plasticity and the emergence of ecotypes. New Phytologist 166(1):101-117.

Derby CD, Sorensen PW. 2008. Neural processing, perception, and behavioral responses to natural chemical stimuli by fish and crustaceans. Journal of Chemical Ecology 34(7):898-914.

Dolgin ES, Felix MA, Cutter AD. 2008. Hakuna Nematoda: genetic and phenotypic diversity in African isolates of *Caenorhabditis elegans* and *C. briggsae*. Heredity 100(3):304-315.

Donker MH, Zonneveld C, van Straalen NM. 1993. Early reproduction and increased reproductive allocation in metal-adapted populations of the terrestrial isopod *Porcellio scaber*. Oecologia 96(3):316-323.

Doroszuk A, Wojewodzic MW, Kammenga JE. 2006. Rapid adaptive divergence of life-history traits in response to abiotic stress within a natural population of a parthenogenetic nematode. Proceedings of the Royal Society B-Biological Sciences 273(1601):2611-2618.

Ebisch IMW, Thomas CMG, Peters WHM, Braat DDM, Steegers-Theunissen RPM. 2007. The importance of folate, zinc and antioxidants in the pathogenesis and prevention of subfertility. Human Reproduction Update 13(2):163-174.

Edwards CA, Bohlen PJ. 1996. Biology and Ecology of Earthworms. London: Chapman and Hall.

Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM. 2010. Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences of the United States of America 107(37):16196-16200.

Emmett BA, Reynolds B, Chamberlain PM, Rowe E, Spurgeon D, Brittain SA, Frogbrook Z, Hughes S, Lawlor AJ, Poskitt J and others. 2010. 2010 Countryside survey: Soils report from 2007. Technical Report No. 9/07 NERC/Centre for Ecology & Hydrology. 192 p.

Engel K, Tollrian R. 2009. Inducible defences as key adaptations for the successful invasion of *Daphnia lumholtzi* in North America? Proceedings of the Royal Society B-Biological Sciences 276(1663):1865-1873.

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods in Molecular Biology. 772:157-78.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics 1:47-50.

Felsenstein J. 1997. An alternating least-squares approach to inferring phylogenies from pairwsie distances. Systematic Biology 46(1):101-111.

Fenili D, Brown M, Rappaport R, McLaurin J. 2007. Properties of scyllo-inositol as a therapeutic treatment of AD-like pathology. Journal of Molecular Medicine 85(6):603-611.

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 473(7347):398-U589.

Finnegan EJ. 2002. Epialleles - a source of random variation in times of stress. Current Opinion in Plant Biology 5(2):101-106.

Fischer E, Koszorus L. 1992. Sublethal effects, accumulation caoacities and elimination rates of As, Hg and Se in the manure worm, *Eisenia foetida* (Oligochaeta, Lumbricidae). Pedobiologia 36(3):172-178.

Fisker KV, Sorensen JG, Damgaard C, Pedersen KL, Holmstrup M. 2011. Genetic adaptation of earthworms to copper pollution: is adaptation associated with fitness costs in *Dendrobaena octaedra*? Ecotoxicology 20(3):563-573.

Flatt T. 2005. The evolutionary genetics of canalization. Quarterly Review of Biology 80(3):287-316.

Forbes VE, Calow P, Grimm V, Hayashi TI, Jager T, Katholm A, Palmqvist A, Pastorok R, Salvito D, Sibly R and others. 2011. Adding value to ecological risk assessment with population modeling. Human and Ecological Risk Assessment 17(2):287-299.

Forbes VE, Calow P, Sibly RM. 2008. The extrapolation problem and how population modeling can help. Environmental Toxicology and Chemistry 27(10):1987-1994.

Francesconi KA. 2010. Arsenic species in seafood: Origin and human health implications. Pure and Applied Chemistry 82(2):373-381.

Francesconi KA, Edmonds JS. 1998. Arsenic species in marine samples. Croatica Chemica Acta 71(2):343-359.

Galletly BC, Blows MW, Marshall DJ. 2007. Genetic mechanisms of pollution resistance in a marine invertebrate. Ecological Applications 17(8):2290-2297.

Gao H, Granka JM, Feldman MW. 2010. On the classification of epistatic interactions. Genetics 184(3):827-U351.

Garcia-Perez I, Liebeke M, Anderson CJ, Lawlor AJ, Bennett MH, Morris CA, Kille P, Spurgeon DJ, Bundy JG. In preparation. Arsenic exposure in earthworms leads to phytochelatin production.

Garcia-Reyero N, Perkins EJ. 2011. Systems biology: leading the revolution in ecotoxicology. Environmental Toxicology and Chemistry 30(2):265-273.

Geiszinger AE, Goessler W, Kosmus W. 2002. An arsenosugar as the major extractable arsenical in the earthworm *Lumbricus terrestris*. Applied Organometallic Chemistry 16(8):473-476.

Glez-Pena D, Gomez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D. 2010. ALTER: program-oriented conversion of DNA and protein alignments. Nucleic Acids Research 38:W14-W18.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q and others. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29(7):644-U130.

Gresser MJ. 1981. ADP arsenate - formation by sub-mitochondrial particles under phosphorylating conditions. Journal of Biological Chemistry 256(12):5981-5983.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology 59(3):307-321.

Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. 2005. Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Genome Research 15(6):790-799.

Hall SE, Beverly M, Russ C, Nusbaum C, Sengupta P. 2011. A cellular memory of developmental history generates phenotypic diversity in *C. elegans*. Current Biology 20(2):149-155.

Hallgrimsdottir IB, Yuster DS. 2008. A complete classification of epistatic two-locus models. Bmc Genetics 9(17):1-15.

Hamilton MA, Russo RC, Thurston RV. 1977. Trimmed Spearman-Karber method for estimating median lethal concentrations in toxicity bioassays. Environmental Science & Technology 11(7):714-719.

Harvey SC, Shorto A, Viney ME. 2008. Quantitative genetic analysis of life-history traits of *Caenorhabditis elegans* in stressful environments. Bmc Evolutionary Biology 8(15):1-16.

Hayakawa T, Kobayashi Y, Cui X, Hirano S. 2005. A new metabolic pathway of arsenite: arsenic-glutathione complexes are substrates for human arsenic methyltransferase Cyt19. Archives of Toxicology 79(4):183-191.

Hazlett BA. 1995. Behavioral plasticity in crustacea: Why not more? Journal of Experimental Marine Biology and Ecology 193(1-2):57-66.

Head JA, Dolinoy DC, Basu N. 2012. Epigenetics for ecotoxicologists. Environmental Toxicology and Chemistry 31(2):221-7.

Hermisson J, Pennings PS. 2005. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. Genetics 169(4):2335-2352.

Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42(2):182-192.

Ho DH, Burggren WW. 2010. Epigenetics and transgenerational transfer: a physiological perspective. Journal of Experimental Biology 213(1):3-16.

Hoelzel AR, Halley J, Obrien SJ, Campagna C, Arnbom T, Leboeuf B, Ralls K, Dover GA. 1993. Elephant seal genetic variation and the use of sumulation models to investigate historical population bottlenecks. Journal of Heredity 84(6):443-449.

Hoffmann AA, Sgro CM. 2011. Climate change and evolutionary adaptation. Nature 470(7335):479-485.

Hoffmann AA, Willi Y. 2008. Detecting genetic responses to environmental change. Nature Reviews Genetics 9(6):421-432.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. Plos Genetics 6(2).

Hooper HL, Jurkschat K, Morgan AJ, Bailey J, Lawlor AJ, Spurgeon DJ, Svendsen C. 2011. Comparative chronic toxicity of nanoparticulate and ionic zinc to the earthworm *Eisenia veneta* in a soil matrix. Environment International 37(6):1111-1117.

Huang XQ, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome Research 9(9):868-877.

Hudson-Edwards K, Macklin M, Brewer P, Dennis I. 2008. Assessment of metal mining-contaminated river sediments in England and Wales. Environment Agency. Report nr SC030136/SR4.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17(8):754-755.

Hughes MF. 2002. Arsenic toxicity and potential mechanisms of action. Toxicology Letters 133(1):1-16.

Hurst GDD, Jiggins FM. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proceedings of the Royal Society B-Biological Sciences 272(1572):1525-1534.

Izutsu M, Zhou J, Sugiyama Y, Nishimura O, Aizu T, Toyoda A, Fujiyama A, Agata K, Fuse N. 2012. Genome features of "dark-fly", a *Drosophila* line reared long-term in a dark environment. Plos One 7(3).

Jablonka E, Raz G. 2009. Transgenerational epigenetic inheritance: Prevalence, mechanisms, and implications for the study of heredity and evolution. Quarterly Review of Biology 84(2):131-176.

Jager T, Heugens EHW, Kooijman SALM. 2006. Making sense of ecotoxicological test results: Towards application of process-based models. Ecotoxicology 15(3):305-314.

Janssens TKS, Marien J, Cenijn P, Legler J, van Straalen NM, Roelofs D. 2007. Recombinational micro-evolution of functionally different metallothionein promoter alleles from *Orchesella cincta*. Bmc Evolutionary Biology 7.

Johnston D, Potter H, Jones C, Rolley S, Watson I, Pritchard J. 2008. Abandoned mines and the water environment. Environment Agency.

Joo H-Y, Zhai L, Yang C, Nie S, Erdjument-Bromage H, Tempst P, Chang C, Wang H. 2007. Regulation of cell cycle progression and gene expression by H2A deubiquitination. Nature 449(7165):1068-U9.

Joshi N. 2011.Sickle. ver. 1.0 [Online]. Available: https://github.com/najoshi/sickle [2011, February 10].

Kammenga JE, Spurgeon DJ, Svendsen C, Weeks JM. 2003. Explaining density-dependent regulation in earthworm populations using life-history analysis. Oikos 100:89-95.

Karasov T, Messer PW, Petrov DA. 2010. Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. Plos Genetics 6(6).

Karatolos N, Pauchet Y, Wilkinson P, Chauhan R, Denholm I, Gorman K, Nelson DR, Bass C, Ffrench-Constant RH, Williamson MS. 2011. Pyrosequencing the transcriptome of the greenhouse whitefly, *Trialeurodes vaporariorum* reveals multiple transcripts encoding insecticide targets and detoxifying enzymes. Bmc Genomics 12.

Kerscher O, Felberbaum R, Hochstrasser M. 2006. Modification of proteins by ubiquitin and ubiquitin-like proteins. Annual Review of Cell and Developmental Biology. p 159-180.

Kille P, Andre J, Anderson C, Ang HA, Bruford MW, Bundy JG, Donnelly R, Hodson ME, Juma G, Lahive E, Morgan AJ, Stürzenbaum, Spurgeon DJ. 2013. Patterns of mitochondrial and nuclear polymorphism and DNA methylation in an arsenic tolerant earthworm population. Soil Biology and Biochemistry 57: 524-532.

King RA, Tibble AL, Symondson WOC. 2008. Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. Molecular Ecology 17(21):4684-4698.

Klerks PL, Levinton JS. 1989. Rapid evolution of metal resistance in a benthic oligochaete inhabiting a metal polluted site. Biological Bulletin 176(2):135-141.

Klinck B, Palumbo B, Cave M, Wragg J. 2005. Arsenic dispersal and bioaccessibilty in mine contaminated soils: a case study from an abandoned arsenic mine in Devon, UK. Nottingham: British Geological Survey. 52 p.

Klok C. 2007. Effects of earthworm density on growth, development, and reproduction in *Lumbricus rubellus* (Hoffm.) and possible consequences for the intrinsic rate of population increase. Soil Biology & Biochemistry 39(9):2401-2407.

Klok C, Zorn M, Koolhaas JE, Eijsackers HJP, van Gestel CAM. 2006. Does reproductive plasticity in *Lumbricus rubellus* improve the recovery of populations in frequently inundated river floodplains? Soil Biology & Biochemistry 38(3):611-618.

Kooijman SALM. 2000. Dynamic energy and mass budgets in biological systems. Cambridge, UK: Cambridge University Press.

Kozul CD, Hampton TH, Davey JC, Gosse JA, Nomikos AP, Eisenhauer PL, Weiss DJ, Thorpe JE, Ihnat MA, Hamilton JW. 2009. Chronic exposure to arsenic in the drinking water alters the expression of immune response genes in mouse lung. Environmental Health Perspectives 117(7):1108-1115.

Kulaeva OI, Gaykalova DA, Studitsky VM. 2007. Transcription through chromatin by RNA polymerase II: Histone displacement and exchange. Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis 618(1-2):116-129.

Kullback S, Leibler RA. 1951. On information and sufficiency. Annals of Mathematical Statistics 22:79-86.

Kumagai Y, Sumi D. 2007. Arsenic: Signal transduction, transcription factor, and biotransformation involved in cellular response and toxicity. Annual Review of Pharmacology and Toxicology. p 243-262.

Kumar C, Igbaria A, D'Autreaux B, Planson A-G, Junot C, Godat E, Bachhawat AK, Delaunay-Moisan A, Toledano MB. 2011. Glutathione revisited: a vital function in iron metabolism and ancillary role in thiol-redox control. Embo Journal 30(10):2044-2056.

Kumar S, Blaxter ML. 2010. Comparing de novo assemblers for 454 transcriptome data. Bmc Genomics 11(571):1-12.

Kumar S, Tamura K, Nei M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics 5:150 - 163.

Lacy RC. 1997. Importance of genetic variation to the viability of mammalian populations. Journal of Mammalogy 78(2):320-335.

Lam SH, Winata CL, Tong Y, Korzh S, Lim WS, Korzh V, Spitsbergen J, Mathavan S, Miller LD, Liu ET and others. 2006. Transcriptome kinetics of arsenic-induced adaptive response in zebrafish liver. Physiological Genomics 27(3):351-361.

Langdon CJ, Hodson ME, Arnold RE, Black S. 2005a. Survival, Pb-uptake and behaviour of three species of earthworm in Pb treated soils determined using an OECD-style toxicity test and a soil avoidance test. Environmental Pollution 138(2):368-375.

Langdon CJ, Meharg AA, Feldmann J, Balgar T, Charnock J, Farquhar M, Piearce TG, Semple KT, Cotter-Howells J. 2002. Arsenic-speciation in arsenate-resistant and non-resistant populations of the earthworm, *Lumbricus rubellus*. Journal of Environmental Monitoring 4(4):603-608.

Langdon CJ, Morgan AJ, Charnock JM, Semple KT, Lowe CN. 2009a. As-resistance in laboratory-reared F1, F2 and F3 generation offspring of the earthworm *Lumbricus rubellus* inhabiting As-contaminated mine soil. Environmental Pollution 157(11):3114-3119.

Langdon CJ, Piearce TG, Black S, Semple KT. 1999. Resistance to arsenic-toxicity in a population of the earthworm *Lumbricus rubellus*. Soil Biology & Biochemistry 31(14):1963-1967.

Langdon CJ, Piearce TG, Feldmann J, Semple KT, Meharg AA. 2003a. Arsenic speciation in the earthworms *Lumbricus rubellus* and *Dendrodrilus rubidus*. Environmental Toxicology and Chemistry 22(6):1302-1308.

Langdon CJ, Piearce TG, Meharg AA, Semple KT. 2001. Survival and behaviour of the earthworms *Lumbricus rubellus* and *Dendrodrilus rubidus* from arsenate-contaminated and non-contaminated sites. Soil Biology & Biochemistry 33(9):1239-1244.

Langdon CJ, Piearce TG, Meharg AA, Semple KT. 2003b. Interactions between earthworms and arsenic in the soil environment: a review. Environmental Pollution 124(3):361-373.

Langdon CJ, Winters C, Sturzenbaum SR, Morgan AJ, Charnock JM, Meharg AA, Piearce TG, Lee PH, Semple KT. 2005b. Ligand arsenic complexation and immunoperoxidase detection of metallothionein in the earthworm *Lumbricus rubellus* inhabiting arsenic-rich soil. Environmental Science & Technology 39(7):2042-2048.

Lavelle P, Bignell D, Lepage M, Wolters V, Roger P, Ineson P, Heal OW, Dhillion S. 1997. Soil function in a changing world: the role of invertebrate ecosystem engineers. European Journal of Soil Biology 33(4):159-193.

Lee B-T, Kim K-W. 2008. Arsenic accumulation and toxicity in the earthworm *Eisenia fetida* affected by chloride and phosphate. Environmental Toxicology and Chemistry 27(12):2488-2495.

Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. Cell 128(4):707-719.

Li D, Guo Y, Shao H, Tellier LC, Wang J, Xiang Z, Xia Q. 2010. Genetic diversity, molecular phylogeny and selection evidence of the silkworm mitochondria implicated by complete resequencing of 41 genomes. Bmc Evolutionary Biology 10.

Lika K, Kooijman S. 2003. Life history implications of allocation to growth versus reproduction in dynamic energy budgets. Bulletin of Mathematical Biology 65(5):809-834.

Lillig CH, Berndt C, Holmgren A. 2008. Glutaredoxin systems. Biochimica et Biophysica Acta - General Subjects 1780(11):1304-1317.

Lillig CH, Berndt C, Vergnolle O, Lonn ME, Hudemann C, Bill E, Holmgren A. 2005. Characterization of human glutaredoxin 2 as iron-sulfur protein: A possible role as redox sensor. Proceedings of the National Academy of Sciences of the United States of America 102(23):8168-8173.

Lin Y-F, Walmsley AR, Rosen BP. 2006. An arsenic metallochaperone for an arsenic detoxification pump. Proceedings of the National Academy of Sciences of the United States of America 103(42):15617-15622.

Lister LJ, Svendsen C, Wright J, Hooper HL, Spurgeon DJ. 2011. Modelling the joint effects of a metal and a pesticide on reproduction and toxicokinetics in lumbricid earthworms. Environment International 37(4):663-670.

Lomax C, Liu W-J, Wu L, Xue K, Xiong J, Zhou J, McGrath SP, Meharg AA, Miller AJ, Zhao F-J. 2012. Methylated arsenic species in plants originate from soil microorganisms. New Phytologist 193(3):665-672.

Lopes PC, Sucena E, Santos ME, Magalhaes S. 2008. Rapid experimental evolution of pesticide resistance in *C. elegans* entails no costs and affects the mating system. Plos One 3(11).

Lowe CN, Butt KR. 2007. Earthworm culture, maintenance and species selection in chronic ecotoxicological studies: A critical review. European Journal of Soil Biology 43:S281-S288.

Lu XW, Zhang XL. 2005. Environmental geochemistry study of arsenic in Western Hunan mining area, PR China. Environmental Geochemistry and Health 27(4):313-320.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. Nature 389(6648):251-260.

Luo J-h, Qiu Z-q, Zhang L, Shu W-q. 2012. Arsenite exposure altered the expression of NMDA receptor and postsynaptic signaling proteins in rat hippocampus. Toxicology Letters 211(1):39-44.

Maldonado Santoyo M, Rodriguez Flores C, Lopez Torres A, Wrobel K, Wrobel K. 2011. Global DNA methylation in earthworms: A candidate biomarker of epigenetic risks related to the presence of metals/metalloids in terrestrial environments. Environmental Pollution 159(10):2387-2392.

Mandrioli M. 2007. A new synthesis in epigenetics: towards a unified function of DNA methylation from invertebrates to vertebrates. Cellular and Molecular Life Sciences 64(19-20):2522-2524.

Marino F, Sturzenbaum SR, Kille P, Morgan AJ. 1998. Cu-Cd interactions in earthworms maintained in laboratory microcosms: the examination of a putative copper paradox. Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 120(2):217-223.

Maroni G, Wise J, Young JE, Otto E. 1987. Metallothionein gene duplications and metal tolerance in natural populations of *Drosophila melanogaster*. Genetics 117(4):739-744.

Marshall DJ, Keough MJ. 2007. The evolutionary ecology of offspring size in marine invertebrates. Advances in Marine Biology 53:1-60.

Marshall DJ, Uller T. 2007. When is a maternal effect adaptive? Oikos 116(12):1957-1963.

Martin JA, Wang Z. 2011. Next-generation transcriptome assembly. Nature Reviews Genetics 12(10):671-682.

Mayr E. 1963. Animal species and evolution. Cambridge: Belknap Press of Harvard University Press.

McGrath PT, Xu Y, Ailion M, Garrison JL, Butcher RA, Bargmann CI. 2011. Parallel evolution of domesticated *Caenorhabditis* species targets pheromone receptor genes. Nature 477(7364):321-U92.

Meharg AA, Shore RF, Broadgate K. 1998. Edaphic factors affecting the toxicity and accumulation of arsenate in the earthworm *Lumbricus terrestris*. Environmental Toxicology and Chemistry 17(6):1124-1131.

Meunier L, Wragg J, Koch I, Reimer KJ. 2010. Method variables affecting the bioaccessibility of arsenic in soil. Journal of Environmental Science and Health Part a-Toxic/Hazardous Substances & Environmental Engineering 45(5):517-526.

Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. 2012. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3 ' UTRs and near Stop Codons. Cell 149(7):1635-1646.

Meyers LA, Bull JJ. 2002. Fighting change with change: adaptive variation in an uncertain world. Trends in Ecology & Evolution 17(12):551-557.

Miller WH, Schipper HM, Lee JS, Singer J, Waxman S. 2002. Mechanisms of action of arsenic trioxide. Cancer Research 62(14):3893-3903.

Mitchell SE, Read AF. 2005. Poor maternal environment enhances offspring disease resistance in an invertebrate. Proceedings of the Royal Society B-Biological Sciences 272(1581):2601-2607.

Moczek AP. 2010. Phenotypic plasticity and diversity in insects. Philosophical Transactions of the Royal Society B-Biological Sciences 365(1540):593-603.

Morgan AJ, Kille P, Sturzenbaum SR. 2007. Microevolution and ecotoxicology of metals in invertebrates. Environmental Science & Technology 41(4):1085-1096.

Morgan HD, Sutherland HGE, Martin DIK, Whitelaw E. 1999. Epigenetic inheritance at the agouti locus in the mouse. Nature Genetics 23(3):314-318.

Morgan JE, Morgan AJ. 1988. Earthworms as biological monitors of cadmium, copper, lead and zinc in metalliferous soils. Environmental Pollution 54(2):123-138.

Morgan JE, Morgan AJ. 1990. The distribution of cadmium, copper, lead, zinc and calcium in the tissues of the earthworm *Lumbricus rubellus* sampled from one uncontaminated and 4 polluted soils. Oecologia 84(4):559-566.

Mukherjee B, Salavaggione OE, Pelleymounter LL, Moon I, Eckloff BW, Schaid DJ, Wieben ED, Weinshilboum RM. 2006. Glutathione S-transferase omega 1 and omega 2 pharmacogenomics. Drug Metabolism and Disposition 34(7):1237-1246.

Mukherji M, Chien W, Kershaw NJ, Clifton IJ, Schofield CJ, Wierzbicki AS, Lloyd MD. 2001. Structure-function analysis of phytanoyl-CoA 2-hydroxylase mutations causing Refsum's disease. Human Molecular Genetics 10(18):1971-1982.

Nelson GM, Ahlborn GJ, Allen JW, Ren HZ, Corton JC, Waalkes MP, Kitchin KT, Diwan BA, Knapp G, Delker DA. 2009. Transcriptional changes associated with reduced spontaneous liver tumor incidence in mice chronically exposed to high dose arsenic. Toxicology 266(1-3):6-15.

Nemeti B, Anderson ME, Gregus Z. 2012. Glutathione synthetase promotes the reduction of arsenate via arsenolysis of glutathione. Biochimie 94(6):1327-1333.

Nesbitt LM, Riessen HP, Ramcharan CW. 1996. Opposing predation pressures and induced vertical migration responses in *Daphnia*. Limnology and Oceanography 41(6):1306-1311.

Neubert MG, Caswell H. 2000. Density-dependent vital rates and their population dynamic consequences. Journal of Mathematical Biology 41(2):103-121.

Nevo E. 2011. Evolution under environmental stress at macro- and microscales. Genome Biology and Evolution 3:1039-1052.

Nieto-Delgado C, Rangel-Mendez JR. 2012. Anchorage of iron hydro(oxide) nanoparticles onto activated carbon to remove As(V) from water. Water Research 46(9):2973-2982.

Norton GJ, Lou-Hing DE, Meharg AA, Price AH. 2008. Rice-arsenate interactions in hydroponics: lwhole genome transcriptional analysis. Journal of Experimental Botany 59(8):2267-2276.

Nussey DH, Wilson AJ, Brommer JE. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. Journal of Evolutionary Biology 20(3):831-844.

OECD. 1984. Guidelines for the testing of chemicals No. 207 Earthworm acute toxicity tests. Organisation for Economic Co-operation and Development.

OECD. 2004. Guidelines for the testing of chemicals No. 222 Earthworm reproduction test. Organisation for Economic Co-operation and Development.

Park B-Y, Lee J-K, Ro H-M, Kim YH. 2011. Effects of heavy metal contamination from an abandoned mine on nematode community structure as an indicator of soil ecosystem health. Applied Soil Ecology 51:17-24.

Partridge MA, Huang SXL, Hernandez-Rosa E, Davidson MM, Hei TK. 2007. Arsenic induced mitochondrial DNA damage and altered mitochondrial oxidative function: Implications for genotoxic mechanisms in mammalian cells. Cancer Research 67(11):5239-5247.

Paulose B, Kandasamy S, Dhankher OP. 2010. Expression profiling of *Crambe abyssinica* under arsenate stress identifies genes and gene networks involved in arsenic metabolism and detoxification. Bmc Plant Biology 10.

Peijnenburg WJGM, Vijver MG. 2009. Earthworms and their use in ecotoxicological modeling. In: Devillers J, editor. Ecotoxicology modeling. London: Springer. p 177-204.

PerezLosada M, Eiroa J, Mato S, Dominguez J. 2005. Phylogenetic species delimitation of the earthworms *Eisenia fetida* (Savigny, 1826) and *Eisenia andrei* Bouche, 1972 (Oligochaeta, Lumbricidae) based on mitochondrial and nuclear DNA sequences. Pedobiologia 49(4):317-324.

Petrick JS, Ayala-Fierro F, Cullen WR, Carter DE, Aposhian HV. 2000. Monomethylarsonous acid (MMA(III)) is more toxic than arsenite in Chang human hepatocytes. Toxicology and Applied Pharmacology 163(2):203-207.

Piearce TG, Langdon CJ, Meharg AA, Semple KT. 2002. Yellow earthworms: distinctive pigmentation associated with arsenic- and copper-tolerance in *Lumbricus rubellus*. Soil Biology & Biochemistry 34(12):1833-1838.

Posada D. 2008. jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution 25(7):1253-1256.

Posthuma L, Hogervorst RF, Joosse ENG, van Straalen NM. 1993. Genetic variation and covariation for characteristics associated with cadmium tolerance in natural populations of the springtail *Orchesella cincta* (L.). Evolution 47(2):619-631.

Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. Current Biology 20(4):R208-R215.

Pulido F. 2007. Phenotypic changes in spring arrival: evolution, phenotypic plasticity, effects of weather and condition. Climate Research 35(1-2):5-23.

Qiu JW, Qian PY. 1998. Combined effects of salinity and temperature on juvenile survival, growth and maturation in the polychaete *Hydroides elegans*. Marine Ecology-Progress Series 168:127-134.

Raab A, Feldmann J, Meharg AA. 2004. The nature of arsenic-phytochelatin complexes in *Holcus lanatus* and *Pteris cretica*. Plant Physiology 134(3):1113-1122.

Raab A, Schat H, Meharg AA, Feldmann J. 2005. Uptake, translocation and transformation of arsenate and arsenite in sunflower (*Helianthus annuus*): formation of arsenic-phytochelatin complexes during exposure to high arsenic concentrations. New Phytologist 168(3):551-558.

Ramakrishnan U, Hadly EA, Mountain JL. 2005. Detecting past population bottlenecks using temporal genetic data. Molecular Ecology 14(10):2915-2922.

Rambaut A. 2009. FigTree. ver. 1.3.1. [Online]. Available: http://tree.bio.ed.ac.uk/software/figtree/ [2009, December 21].

Rand GM. 1995. Fundamentals of aquatic toxicology: Effects, environmental fate and risk assessment. London: Taylor and Francis Inc.

Reale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ. 2007. Integrating animal temperament within ecology and evolution. Biological Reviews 82(2):291-318.

Reed DH, Briscoe DA, Frankham R. 2002. Inbreeding and extinction: The effect of environmental stress and lineage. Conservation Genetics 3(3):301-307.

Reed RD, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, Jiggins CD, Chamberlain NL, Kronforst MR, Chen R and others. 2011. Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333(6046):1137-1141.

Regev A, Lamb MJ, Jablonka E. 1998. The role of DNA methylation in invertebrates: Developmental regulation or genome defense? Molecular Biology and Evolution 15(7):880-891.

Reichard JF, Puga A. 2010. Effects of arsenic exposure on DNA methylation and epigenetic gene regulation. Epigenomics 2(1):87-104.

Reichard JF, Puga A. 2011. Effects of arsenic exposure on DNA methylation and epigenetic gene regulation. Epigenomics 2(1):87-104.

Ritchie AW, Edmonds JS, Goessler W, Jenkins RO. 2004. An origin for arsenobetaine involving bacterial formation of an arsenic-carbon bond. Fems Microbiology Letters 235(1):95-99.

Roelofs D, Morgan J, Stuerzenbaum S. 2010. The significance of genome-wide transcriptional regulation in the evolution of stress tolerance. Evolutionary Ecology 24(3):527-539.

Rouault TA, Tong WH. 2005. Iron-sulphur cluster biogenesis and mitochondrial iron homeostasis. Nature Reviews Molecular Cell Biology 6(4):345-351.

Saccheri IJ, Wilson IJ, Nichols RA, Bruford MW, Brakefield PM. 1999. Inbreeding of bottlenecked butterfly populations: Estimation using the likelihood of changes in marker allele frequencies. Genetics 151(3):1053-1063.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4(4):406-25.

Schlichting CD, Smith H. 2002. Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes. Evolutionary Ecology 16(3):189-211.

Schones DE, Zhao K. 2008. Genome-wide approaches to studying chromatin modifications. Nature Reviews Genetics 9(3):179-191.

Schwartz BE, Ahmad K. 2005. Transcriptional activation triggers deposition and removal of the histone variant H3.3. Genes & Development 19(7):804-814.

Scott-Fordsmand JJ, Weeks JM, Hopkin SP. 1998. Toxicity of nickel to the earthworm and the applicability of the neutral red retention assay. Ecotoxicology 7(5):291-295.

Seong K-H, Li D, Shimizu H, Nakamura R, Ishii S. 2011. Inheritance of stress-induced, ATF-2-dependent epigenetic change. Cell 145(7):1049-1061.

Shin KH, Kim JY, Kim KW. 2007. Earthworm toxicity test for the monitoring arsenic and heavy metal-containing mine tailings. Environmental Engineering Science 24(9):1257-1265.

Shinkai Y, Sumi D, Toyama T, Kaji T, Kumagai Y. 2009. Role of aquaporin 9 in cellular accumulation of arsenic and its cytotoxicity in primary mouse hepatocytes. Toxicology and Applied Pharmacology 237(2):232-236.

Shirayama M, Seth M, Lee H-C, Gu W, Ishidate T, Conte D, Jr., Mello CC. 2012. piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. Cell 150(1):65-77.

Shirley MDF, Sibly RM. 1999. Genetic basis of a between-environment trade-off involving resistance to cadmium in *Drosophila melanogaster*. Evolution 53(3):826-836.

Sih A, Bell A, Johnson JC. 2004. Behavioral syndromes: an ecological and evolutionary overview. Trends in Ecology & Evolution 19(7):372-378.

Simes RJ. 1986. An improved Bonferroni procedure for multiple tests of significance. Biometrika 73(3):751-754.

Sims RW, Gerard BM. 1985. Earthworms. Kermack DM, Barnes RSK, editors. London: Linnean Society and the Estuarine and Brackish-Water Sciences Association. 171 p.

Slatkin M. 1996. In defense of founder-flush theories of speciation. American Naturalist 147(4):493-505.

Smith R, Pollard SJT, Weeks JM, Nathanail CP. 2005. Assessing significant harm to terrestrial ecosystems from contaminated land. Soil Use and Management 21:527-540.

Song W-Y, Park J, Mendoza-Cozatl DG, Suter-Grotemeyer M, Shim D, Hoertensteiner S, Geisler M, Weder B, Rea PA, Rentsch D and others. 2010. Arsenic tolerance in *Arabidopsis* is mediated by two ABCC-type phytochelatin transporters. Proceedings of the National Academy of Sciences of the United States of America 107(49):21187-21192.

Spurgeon DJ, Hopkin SP. 1995. Extrapolation of the laboratory-based OECD earthworm toxicity test to metal contaminated field sites. Ecotoxicology 4(3):190-205.

Spurgeon DJ, Hopkin SP. 1996. Effects of metal-contaminated soils on the growth, sexual development, and early cocoon production of the earthworm *Eisenia fetida*, with particular reference to zinc. Ecotoxicology and Environmental Safety 35(1):86-95.

Spurgeon DJ, Jones OAH, Dorne J-LCM, Svendsen C, Swain S, Stuerzenbaum SR. 2010. Systems toxicology approaches for understanding the joint effects of environmental chemical mixtures. Science of the Total Environment 408(18):3725-3734.

Spurgeon DJ, Lawlor A, Hooper HL, Wadsworth R, Svendsen C, Thomas LDK, Ellis JK, Bundy JG, Keun HC, Jarup L. 2011. Outdoor and indoor cadmium distributions near an abandoned smelting works and their relations to human exposure. Environmental Pollution 159(12):3425-3432.

Spurgeon DJ, Rowland P, Ainsworth G, Rothery P, Long S, Black HIJ. 2008. Geographical and pedological drivers of distribution and risks to soil fauna of seven metals (Cd, Cu, Cr, Ni, Pb, V and Zn) in British soils. Environmental Pollution 153(2):273-283.

Spurgeon DJ, Stürzenbaum SR, Svendsen C, Morgan AJ, Weeks JM, Kille P. 2004a. Molecular and cellular biomarkers of metal toxicity in *Lumbricus rubellus*; Validation by integration with a demographic parameter. . Comparative Biochemistry and Physiology Part C: Toxicology and Pharmacology 138:11-21.

Spurgeon DJ, Svendsen C, Kille P, Morgan AJ, Weeks JM. 2004b. Responses of earthworms (*Lumbricus rubellus*) to copper and cadmium as determined by measurement of juvenile traits in a specifically designed test system. Ecotoxicology and Environmental Safety 57(1):54-64.

Spurgeon DJ, Svendsen C, Rimmer VR, Hopkin SP, Weeks JM. 2000. Relative sensitivity of life-cycle and biomarker responses in four earthworm species exposed to zinc. Environmental Toxicology and Chemistry 19(7):1800-1808.

Spurgeon DJ, Svendsen C, Weeks JM, Hankard PK, Stubberud HE, Kammenga JE. 2003a. Quantifying copper and cadmium impacts on intrinsic rate of population increase in the terrestrial oligochaete *Lumbricus rubellus*. Environmental Toxicology and Chemistry 22(7):1465-1472.

Spurgeon DJ, Weeks JM, Van Gestel CAM. 2003b. A summary of eleven years progress in earthworm ecotoxicology. Pedobiologia 47(5-6):588-606.

Srikwan S, Woodruff DS. 2000. Genetic erosion in isolated small-mammal populations following rainforest fragmentation. In: Young AG, Clarke GM, editors. Genetics, demography and viability of fragmented populations. Cambridge: Cambridge University Press. p 149-172.

States JC, Singh AV, Knudsen TB, Rouchka EC, Ngalame NO, Arteel GE, Piao Y, Ko MSH. 2012. Prenatal arsenic exposure alters gene expression in the adult liver to a proinflammatory state contributing to accelerated atherosclerosis. Plos One 7(6).

Stroud JL, Norton GJ, Islam MR, Dasgupta T, White RP, Price AH, Meharg AA, McGrath SP, Zhao F-J. 2011. The dynamics of arsenic in four paddy fields in the Bengal delta. Environmental Pollution 159(4):947-953.

Stuerzenbaum SR, Andre J, Kille P, Morgan AJ. 2009. Earthworm genomes, genes and proteins: the (re)discovery of Darwin's worms. Proceedings of the Royal Society B-Biological Sciences 276(1658):789-797.

Szinicz L, Forth W. 1988. Effects of As2O3 on gluconeogenesis. Archives of Toxicology 61(6):444-449.

Tanguy A, Boutet I, Bonhomme FY, Boudry P, Moraga D. 2002. Polymorphism of metallothionein genes in the Pacific oyster *Crassostrea gigas* as a biomarker of response to metal exposure. Biomarkers 7(6):439-450.

Taylor AC, Sherwin WB, Wayne RK. 1994. Genetic variation of microsatellite loci in a bottlenecked species- the northern hairy-nosed wombat *Lasiorhinus krefftii*. Molecular Ecology 3(4):277-290.

Thoral S, Rose J, Garnier JM, Van Geen A, Refait P, Traverse A, Fonda E, Nahon D, Bottero JY. 2005. XAS study of iron and arsenic speciation during Fe(II) oxidation in the presence of As(III). Environmental Science & Technology 39(24):9478-9485.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7(3):562-578.

Tully T, Ferriere R. 2008. Reproductive Flexibility: Genetic Variation, Genetic Costs and Long-Term Evolution in a *Collembola*. Plos One 3(9).

van der Ploeg MJC, Baveco JM, van der Hout A, Bakker R, Rietjens IMCM, van den Brink NW. 2011. Effects of C(60) nanoparticle exposure on earthworms (*Lumbricus rubellus*) and implications for population dynamics. Environmental Pollution 159(1):198-203.

van Straalen NM, Feder ME. 2012. Ecological and evolutionary functional genomics-how can it contribute to the risk assessment of chemicals? Environmental Science & Technology 46(1):3-9.

van Straalen NM, Janssens TKS, Roelofs D. 2011. Micro-evolution of toxicant tolerance: from single genes to the genome's tangled bank. Ecotoxicology 20(3):574-579.

van Straalen NM, Timmermans M. 2002. Genetic variation in toxicant-stressed populations: An evaluation of the "genetic erosion" hypothesis. Human and Ecological Risk Assessment 8(5):983-1002.

Vandegehuchte MB, Kyndt T, Vanholme B, Haegeman A, Gheysen G, Janssen CR. 2009. Occurrence of DNA methylation in *Daphnia magna* and influence of multigeneration Cd exposure. Environment International 35(4):700-706.

Vandegehuchte MB, Lemiere F, Vanhaecke L, Vanden Berghe W, Janssen CR. 2010. Direct and transgenerational impact on *Daphnia magna* of chemicals with a known effect on DNA methylation. Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 151(3):278-285.

Vijver MG, Koster M, Peijnenburg WJGM. 2007. Impact of pH on Cu accumulation kinetics in earthworm cytosol. Environmental Science & Technology 41(7):2255-60.

Vijver MG, van Gestel CAM, van Straalen NM, Lanno RP, Peijnenburg W. 2006. Biological significance of metals partitioned to subcellular fractions within earthworms (*Aporrectodea caliginosa*). Environmental Toxicology and Chemistry 25(3):807-814.

Vijver MG, Vink JPM, Miermans CJH, van Gestel CAM. 2003. Oral sealing using glue: a new method to distinguish between intestinal and dermal uptake of metals in earthworms. Soil Biology & Biochemistry 35(1):125-132.

Walker CH, Hopkin SP, Sibly RM, Peakall DB. 2006. Principles of ecotoxicology. London, UK: Taylor & Francis. 314 p.

Walsh TK, Brisson JA, Robertson HM, Gordon K, Jaubert-Possamai S, Tagu D, Edwards OR. 2010. A functional DNA methylation system in the pea aphid, *Acyrthosiphon pisum*. Insect Molecular Biology 19:215-228.

Wang Y, Ghaffari N, Johnson CD, Braga-Neto UM, Wang H, Chen R, Zhou H. 2011. Evaluation of the coverage and depth of transcriptome by RNA-Seq in chickens. Bmc Bioinformatics 12.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10(1):57-63.

Washburn MP, Wells WW. 1999. The catalytic mechanism of the glutathione-dependent dehydroascorbate reductase activity of thioltransferase (glutaredoxin). Biochemistry 38(1):268-274.

Waterland RA, Kellermayer R, Laritsky E, Rayco-Solon P, Harris RA, Travisano M, Zhang W, Torskaya MS, Zhang J, Shen L and others. 2010. Season of conception in rural Gambia affects DNA methylation at putative human metastable epialleles. Plos Genetics 6(12).

Watts MJ, Button M, Brewer TS, Jenkin GRT, Harrington CF. 2008a. Quantitative arsenic speciation in two species of earthworms from a former mine site. Journal of Environmental Monitoring 10(6):753-759.

Watts MJ, Button M, Brewer TS, Jenkin GRT, Harrington CF. 2008b. Quantitative arsenic speciation in two species of earthworms from a former mine site. Journal of Environmental Monitoring 10:753-759.

Weeks JM, Svendsen C. 1996. Neutral red retention by lysosomes from earthworm (*Lumbricus rubellus*) coelomocytes: A simple biomarker of exposure to soil copper. Environmental Toxicology and Chemistry 15(10):1801-1805.

Widdows J, Johnson D. 1988. Physiological energetics of *Mytilus edulis* - scope for growth. Marine Ecology-Progress Series 46(1-3):113-121.

Williams TD, Turan N, Diab AM, Wu H, Mackenzie C, Bartie KL, Hrydziuszko O, Lyons BP, Stentiford GD, Herbert JM, Abraham JK, Katsiadaki I, Leaver MJ, Taggart JB, George SG, Viant MR, Chipman KJ and Falciani F. 2011. Towards a system level understanding of non-model organisms sampled from the environment: A network biology approach. Plos Computational Biology 7(8).

Williams TD, Wu H, Santos EM, Ball J, Katsiadaki I, Brown MM, Baker P, Ortega F, Falciani F, Craft JA, Tyler CR, Chipman JK and Viant MR. 2009. Hepatic transcriptomic and metabolomic responses in the stickleback (*Gasterosteus aculeatus*) exposed to environmentally relevant concentrations of dibenzanthracene. Environmental Science & Technology 43(16):6341-6348.

Winnebeck EC, Millar CD, Warman GR. 2010. Why does insect RNA look degraded? Journal of Insect Science 10:1-7.

Winski SL, Carter DE. 1995. Interactions of rat red blood cell sulfhydryls with arsenate and arsenite. Journal of Toxicology and Environmental Health 46(3):379-397.

Wolff GL, Kodell RL, Moore SR, Cooney CA. 1998. Maternal epigenetics and methyl supplements affect agouti gene expression in A(vy)/a mice. Faseb Journal 12(11):949-957.

Wood TE, Burke JM, Rieseberg LH. 2005. Parallel genotypic adaptation: when evolution repeats itself. Genetica 123(1-2):157-170.

Workman JL. 2006. Nucleosome displacement in transcription. Genes & Development 20(15):2009-2017.

Wu B, Zhang Y, Zhao DY, Zhang XX, Kong ZM, Cheng SP. 2009. Gene expression profiles in liver of mouse after chronic exposure to drinking water. Journal of Applied Toxicology 29(7):569-577.

Wu K-M, Lu Y-H, Feng H-Q, Jiang Y-Y, Zhao J-Z. 2008. Suppression of cotton bollworm in multiple crops in China in areas with Bt toxin-containing cotton. Science 321(5896):1676-1678.

Yeates GW, Orchard VA, Speir TW, Hunt JL, Hermans MCC. 1994. Impact of pasture contamination by copper, chromium, arsenic timber preservative on soil biological activity. Biology and Fertility of Soils 18(3):200-208.

Yoshitake S, Nanri H, Fernando MR, Minakami S. 1994. Possible differences in the regenerative roles played by thioltransferase and thioredoxin for oxidatively damaged proteins. Journal of Biochemistry 116(1):42-46.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328(5980):916-919.

Zerbino DR, Birney E. 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Research 18(5):821-829.

Zhai J, Liu J, Liu B, Li P, Meyers BC, Chen X, Cao X. 2008. Small RNA-directed epigenetic natural variation in *Arabidopsis thaliana*. Plos Genetics 4(4):e1000056-e1000056.

Zhang Y, Zhang X, Wu B, Cheng S. 2012. Evaluating the transcriptomic and metabolic profile of mice exposed to source drinking water. Environmental Science & Technology 46(1):78-83.

Zhen Y, Zhao S, Li L, Li Y, Kawamoto K. 2010. Arsenic trioxide-mediated Notch pathway inhibition depletes the cancer stem-like cell population in gliomas. Cancer Letters 292(1):64-72.

Zhou S, Campbell TG, Stone EA, Mackay TFC, Anholt RRH. 2012. Phenotypic plasticity of the *Drosophila* transcriptome. Plos Genetics 8(3).

Abbreviations

Uncommon abbreviations within discrete chapters are explicitly stated upon use, however, this acts as a record for all abbreviations used. Atomic symbols are frequently used. Metabolites associated with metabolic analyses are described in Table 3.1 and 3.2.2, while gene names derived from transcriptomic analyses are found in Table 3.3.

Units of measure

°C	Degree Celcius
bp	Dase pair
d	Day
g	Gram
kb	Kilo base
kg	Kilogram
L	Litre
mg	Milligram
ml	Millilitre
mM	Millimolar
ng	Nanogram
s	Second
µl	Microlitre
µM	Micromolar

Geographic Locations

CAR	Caradon
CF	Carrock Fell
CWM	Cwmystwyth
CLAD	Clydach
DGC	Devon Great Consols
-P	Pool
-S2	Site 2
DPA	Dynas Powys
ECT	Ecton
SH	Shipham
UK	United Kingdom

The addition of C or M distinguishes whether the site is uncontaminated or a former mine site, respectively. Geographic locations are listed in 5.2.1 and observable in Figure 5.4

Miscellaneous

^{1}H-NMR	Hydrogen-1 nuclear magnetic resonance
AFLP	Amplified fragment length polymorphism
AMOVA	Analysis of molecular variance
ANOVA	Analysis of variance
cDNA	Complementary DNA
COI	Cytochrome oxidase I
COII	Cytochrome oxidase II
CRM	Certified reference material
DEG	Differentially expressed gene
EC50	Half maximal effective concentration
EST	Expressed sequence tag
gDNA	Genomic DNA
GLM	General linear model
GSH	Glutathione
HPLC	High performance liquid chromatography
ICP-MS	Inductively coupled plasma mass spectrometer
IPA	Ingenuity Pathways Analysis
LC50	Half maximum lethal concentration
mRNA	Messenger RNA
ML	Maximum Likelihood
MtDNA	Mitochondrial DNA
PC	Phytochelatin
PCA	Principle component analysis
PCR	Polymerase chain reaction
qPCR	Quantitative PCR
RADseq	Restriction associated DNA sequencing
RIL	Recombinant inbred lineage
RNAseq	RNA sequencing
SNP	Single-nucleotide polymorphism
TU	Toxic unit

Appendix A Differential gene expression values for adult *L. rubellus* exposed to 36 mg/kg As with those exposed to 125 mg/kg As. Fold change beyond a threshold of above 2 or below -2 are shown and is alphabetized based upon the gene symbol referred to by IPA. Ensemble Gene ID, the location of activity within cells and gene product are also listed.

		36/125			
Fold Change	**ID**	**Symbol**	**Entrez Gene Name**	**Location**	**Type(s)**
3.254	ENS600000087085	ACHE	acetylcholinesterase	Plasma Membrane	enzyme
-2.262	ENSP00000439955	ACSBG1	acyl-COA synthetase bubblegum family member 1	Cytoplasm	enzyme
3.149	ENSP00000357665	ADAM12	ADAM metallopeptidase domain 12	Plasma Membrane	peptidase
3.821	ENSP00000363678	AGRN	agrin	Plasma Membrane	other
-2.279	ENS6000001584167	AHCYL2	adenosylhomocysteinase-like 2	unknown	enzyme
-3.516	ENSG00000101280	ANGPT4	angiopoietin 4	Extracellular Space	growth factor
2.868	ENSP00000351681	ANK1	ankyrin 1, erythrocytic	Plasma Membrane	other
-3.245	ENSP00000373287	ANKRD28	ankyrin repeat domain 23	unknown	other
2.474	ENSP00000437825	ANKRD44	ankyrrn repeat domain 44	unknown	other
-2.131	ENSP00000433836	ARSJ	arylsulfatase family, member}	Extracellular Space	enzyme
-2.971	ENSP00000343674	ASTL	astacin-like metallo-endopeptidase (M12 family)	unknown	peptidase
-2.001	ENS600000174437	ATP2A2	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2	Cytoplasm	transporter

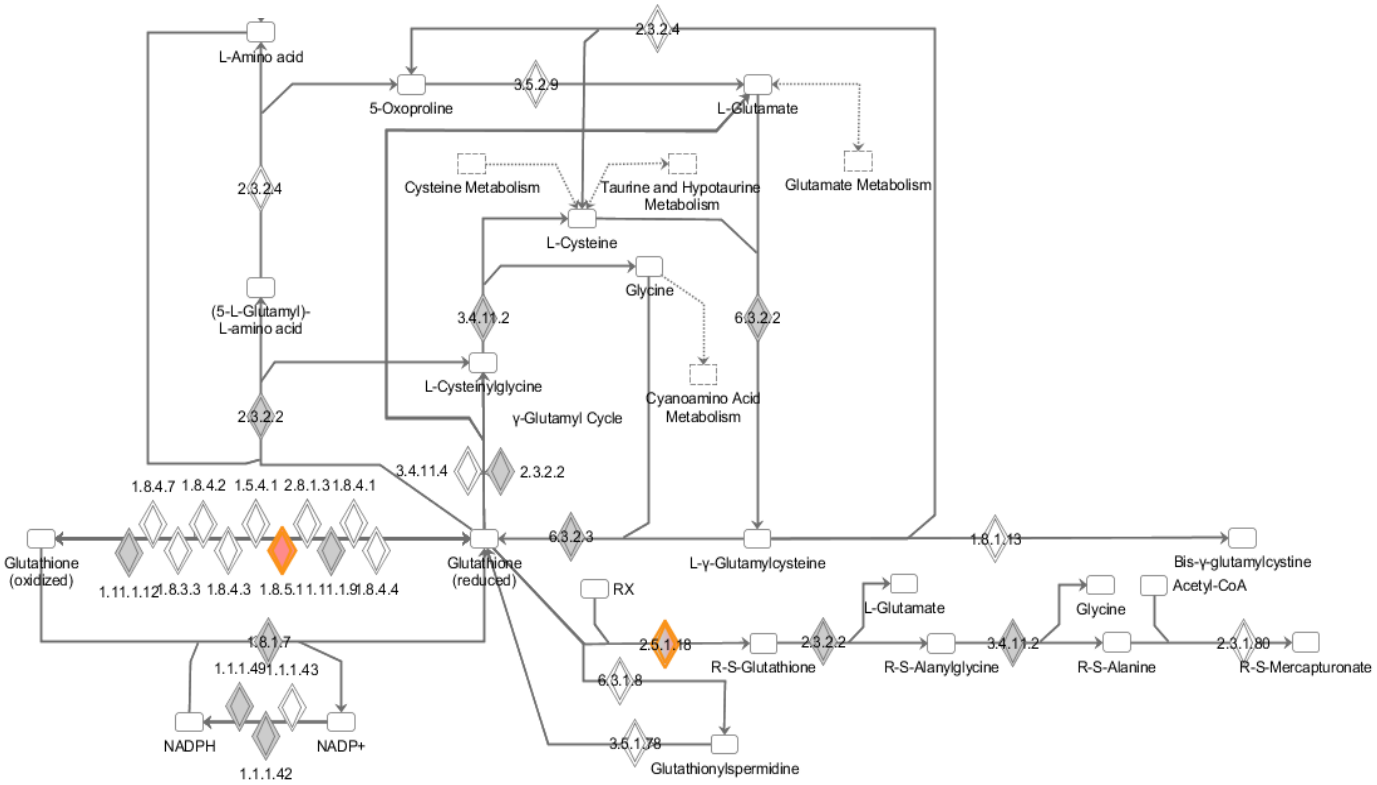
2.194	ENSP00000328087	AVL9	AVL9 homolog (S. cerevisiase)	unknown	other
2.324	ENSP00000432867	BANF1	barrier to autointegration factor 1	Nucleus	other
2.52	ENSP000004SSOS0	C16orf5	chromosome 16 open reading frame 5	Nucleus	other
3.519	ENSP00000389176	C1GALT1	core 1 synthase, glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase, 1	Plasma Membrane	enzyme
-2.315	ENSP00000371488	C1QTNF3	C1q and tumor necrosis factor related protein 3	Extracellular Space	other
2.475	ENSP00000389014	C5orf42	chromosome 5 open reading frame 42	unknown	other
-2.293	ENSG00000106144	CASP2	caspase 2, apoptosis-related cysteine peptidase	Cytoplasm	peptidase
-2.349	ENSP00000309052	CATSPER1	cation channel, sperm associated 1	Plasma Membrane	ion channel
-2.245	ENSP00000393139	CDC7	cell division cycle 7 homolog (S. cerevisiae)	Nucleus	kinase
-3.133	ENSP00000352639	CELA2A	chymotrypsin-like elastase family, member 2A	Extracellular Space	peptidase
-2.247	ENSP00000401608	CEP350	centrosomal protein 350kDa	Cytoplasm	other
2.082	ENSP00000341828	CHIA	chitinase, acidic	Extracellular Space	enzyme
-2.17	ENSG00000133063	CHIT1	chitinase 1 (chitotriosidase)	Extracellular Space	enzyme
-2.219	ENSG00000138663	COPS4	COP9 constitutive photomorphogenic homolog subunit -1 (Arabidopsis)	Cytoplasm	other
-3.216	ENSG00000109472	CPE	carboxypeptidaseE	Plasma Membrane	peptidase
-2.602	ENSP00000280527	CRIM1	cysteine rich transmembrane BMP regulator 1 (chordin-like)	Extracellular Space	kinase
-2.151	ENSP00000264474	CSTA	cystatin A (stefin A)	Cytoplasm	other
-4.337	ENSP00000365061	CTSL1	cathepsin L1	Cytoplasm	peptidase
-2.264	ENSP00000417674	CYLC2	cylicin, basic protein of sperm head cytoskeleton 2	Cytoplasm	other

2.203	ENSP00000230895	DAP	death-associated protein	Cytoplasm	transcription regulator
2.178	ENSP00000394484	DCLK3	doublecortin-like kinase 3	Cytoplasm	kinase
-3.909	ENSP00000355718	DLL1	delta-like 1 (Drosophila)	Plasma Membrane	enzyme
-2.355	ENSP00000382213	DSPP	dentin sialophosphoprotein	Extracellular Space	other
-2.832	ENSP00000381247	EFCAB9	EF-hand calcium binding domain 9	unknown	other
2.426	ENSP00000105580	ElF1	E74-like factor 1 (ets domain transcription factor)	Nucleus	transcription regulator
-3.847	ENSP00000359292	FAM178A	family with sequence similarity 178, memberA	unknown	other
-4.068	ENSP00000371262	FGL1	fibrinogen-like 1	Extracellular Space	other
3.064	ENSP00000255759	FOPNL	FGFR1OP N-terminal like	Cytoplasm	other
3.526	ENSP00000426315	FST	follistatin	Extracellular Space	other
2.521	ENSP00000440409	FSTL5	follistatin-like 5	Extracellular Space	other
2.052	ENSP00000389536	FUBP1	far upstream element (FUSE) binding protein 1	Nucleus	transcription regulator
-2.847	ENSP00000378086	FUT8	fucosyltransferase 8 (alpha (1,6) fucosyltransferase)	Cytoplasm	enzyme
-2.118	ENSP00000037243	GABARAPL2	GABA(A) receptor-associated protein-lil<e2	Cytoplasm	other
-2.815	ENSG00000114480	GBE1	glucan (1,4-alpha-), branching enzyme 1	Cytoplasm	enzyme
-2.478	ENSG00000089154	GCN1L1	GCN1 general control of amino-acid synthesis 1-like 1 (yeast)	Cytoplasm	translation regulator
-2.26	ENSP00000218516	GLA	galactosidase, alpha	Cytoplasm	enzyme
-2.471	ENSP00000349687	GM2A	GM2 ganglioside activator	Cytoplasm	enzyme
2.391	ENSP00000316598	GTPBP6	GTP binding protein 6 (putative)	unknown	other
-2.845	ENSG00000004961	HCCS	holocytochrome c synthase	Cytoplasm	enzyme
-2.529	ENSP00000441169	HCK	hemopoietic cell kinase	Cytoplasm	kinase
3.383	ENSP00000310729	HERPUD2	HERPUD family member 2	unknown	other

-2.81	ENSP00000378250	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)	Extracellular Space	growth factor
-2.257	ENSP00000389519	JAG1	jagged 1	Extracellular Space	growth factor
-4.846	ENSG00000120457	KCNJ5	potassium inwardly-rectifying channel, subfamily J, member 5	Plasma Membrane	ion channel
-3.638	ENSP00000261824	KDM2B	lysine (K)-specific demethylase 28	Nucleus	other
2.643	ENSP00000364397	KIAA0195	KIAA0195	unknown	other
2.227	ENSP00000344307	KIAA1797	KIAA1797	unknown	other
2.388	ENSP000O0313571	LGALS7/LGALS7B	lectin, galactoside-binding, soluble, 7	Extracellular Space	other
-2.439	ENSP00000409813	LRP2	low density lipoprotein receptor-related protein 2	Plasma Membrane	transporter
2.399	ENSP00000297153	MDGA1	MAM domain containing glycosylphosphatidylinositol anchor 1	Plasma Membrane	other
3.87	ENSP00000353919	MEGF11	multiple EGF-like-domains 11	unknown	other
-3.296	ENSP00000413976	MID2	midline 2	Cytoplasm	other
-2.745	ENSP00000322788	MMP1	matrix metallopeptidase 1 (interstitial collagenase)	Extracellular Space	peptidase
-3.606	ENSG00000086504	MRLP28	mitochondrial ribosomal protein L28	Cytoplasm	other
2.641	ENSP00000384815	MUC5AC/MUC5B	mucin 5AC, oligomeric mucus/gel-forming	Extracellular Space	peptidase
3.521	ENSP00000346846	MYLK	myosin light chain kinase	Cytoplasm	kinase
-2.208	ENSG00000198951	NAGA	N-acetylgalactosaminidase, alpha-	Cytoplasm	enzyme
-2.393	ENSP00000411392	NDUFS7	NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase)	Cytoplasm	enzyme
-2.087	ENSP00000446150	NOTCH3	notch 3	Plasma Membrane	transcription regulator
-2.359	ENSP00000403447	NOTCH4	notch 4	Plasma Membrane	transcription regulator
12.585	ENSG00000101405	OXT	oxytocin, prepropeptide	Extracellular Space	other

-2.842	ENSP00000367727	PANK4	pantothenate kinase 4	Cytoplasm	kinase
-2.181	ENSG00000165443	PHYHIPL	phytanoyl-CoA 2-hydroxydase interacting protein-like	Cytoplasm	other
-2.3778	ENSP00000444848	PI16	peptidase inhibitor 16	Extracellular Space	other
-2.559	ENSP00000308938	PLG	plasminogen	Extracellular Space	peptidase
2.525	ENSP00000417792	PLSCR1	phospholipid scramblase 1	Plasma Membrane	enzyme
2.522	ENSP00000234310	PPP3R1	protein phosphatase 3, regulatory subunit 8, alpha	Cytoplasm	phosphatase
2.643	ENSG00000141956	PRDMI5	PR domain containing 15	Nucleus	other
2.386	ENSP00000260045	PRKRIR	protein-kinase, interferon-inducible double stranded RNA dependent inhibitor, repressor of (P58 repressor)	Nucleus	other
2.646	ENSG00000117360	PRPF3	PRP3 pre-mRNA processing factor 3 homolog (S. cerevisiae)	Nucleus	other
-2.344	ENSG00000011304	PTBP1	polypyrimidine tract binding protein 1	Nucleus	enzyme
-2.307	ENSP00000381206	RDH16	retinol dehydrogenase 16 (all-trans)	Cytoplasm	enzyme
-2.332	ENSG00000108469	RECQL5	RecQ protein-like 5	Nucleus	enzyme
13.533	ENSP00000367590	REEP2	receptor accessory protein 2	unknown	other
2.677	ENSP00000384056	RPS24	ribosomal protein S24	Cytoplasm	other
2.333	ENSP00000357555	RPS27	ribosomal protein S27	Cytoplasm	other
-3.272	ENSG00000138674	SEC31A	SEC31 homolog A (S. cerevisiae)	Cytoplasm	other
-2.706	ENSP00000404243	SLC12A8	solute carrier family 12 (potassium/chloride transporters), member 8	unknown	transporter
2.175	ENSP00000445340	SLC5A8	solute carrier family 5 (iodide transporter), member 8	Plasma Membrane	transporter
2.073	ENSP00000339834	SNRNP48	small nuclear nbonucleoprotein 48kDa (U11/U12)	Nucleus	other
2.642	ENSP00000379621	SNX16	sorting nexin 16	unknown	transporter

2.399	ENSG00000170677	S0CS6	suppressor of cytokine signaling 6	Cytoplasm	other
2.064	ENSP00000305494	SPATA5L1	spermatogenesis associated 5-like 1	unknown	other
-4.836	ENSG00000100014	SPECC1L	sperm antigen with calponin homology and coiled-coil domains 1-like	unknown	other
3.38	ENSP00000366616	SPSB1	splA/ryanodine receptor domain and SOCS box containing 1	Cytoplasm	other
2.749	ENSP00000350630	TLL2	tolloid-like 2	Extracellular Space	peptidase
13.718	ENSP00000335094	TMEM17	transmembrane protein 17	Extracellular Space	other
-2.029	ENSP00000320239	TRPM7	transient receptor potential cation channel, subfamily M, member 7	Plasma Membrane	kinase
-2.918	ENSG00000123607	TTC21B	tetratricopepticle repeat domain 218	unknown	other
-2.105	ENSG00000260703	TTC25	tetratricopepticle repeat domain 25	Cytoplasm	other
-3.389	ENSP00000402803	URGCP	upregulator of cell proliferation	Cytoplasm	other
13.262	ENSG00000119801	YPEL5	yippee-like 5 (Drosophila)	unknown	other
-2.616	ENSP00000282007	ZC3H13	zinc finger CCCH-type containing 13	unknown	other
2.381	ENSG00000066827	ZFAT	zinc finger and AT hook domain containing	Nucleus	other
2.08	ENSP00000294258	ZFPL1	zinc finger protein-like 1	Cytoplasm	other
2.172	ENSP00000287727	ZFYVE9	zinc finger, FWE domain containing 9	Cytoplasm	peptidase
2.812	ENSG00000131115	ZNF227	zinc finger protein 227	Nucleus	transcription regulator
12.502	ENSP00000347045	ZNF43	zinc finger protein 43	Nucleus	other
3.761	ENSP00000388311	ZNF845	zinc finger protein 845	unknown	other
-2.164	ENSG00000182318	ZSCAN22	zinc finger and SCAN domain containing 22	Nucleus	transcription regulator



Appendix B

The role of glutathione is generally considered to be important for metabolism of As. This pathway from IPA demonstrates that a range of transcripts were annotated and included in the analysis (grey diamonds), however only 2 were up-regulated (highlighted in orange). Of these, only a single gene was found to be up-regulated over the 2-fold threshold; 1.8.5.1, or GSTO1. The lack of inclusion within other gene bodies within the pathway insinuates that other, unannotated contigs may be associated within pathways that are either unrecognised or novel among model species. Enzyme references are based upon KEGG database nomenclature, while metabolites directly associated are also shown.