The Earthworm Microbiome

Daniel Antony Pass

2015

A thesis presented for the degree of Doctor of Philosophy

Supervisory group:

Prof. P. Kille[1], Prof. A. J. Weightman[1], Prof. D. Spurgeon[2].

[1] Cardiff School of Biosciences, BIOSI 1, University of Cardiff, P.O. Box 915, Cardiff, CF10 3TL, UK.
[2] Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Wallingford, Oxfordshire OX10 8BB, UK.

Summary

Background: Host-associated microbial communities play a significant role in a species' environmental interactions, often performing functions unachievable by the eukaryotic host, and is essential in developing a comprehensive understanding of the species and its impact on the local and global ecosystem.

Earthworms (Lumbricina) habituate almost every type of soil environment globally, including sites of severe environmental stress and is an essential ecosystem engineer, central to healthy natural and agricultural soils. To date, only a singular symbiotic species (*Verminephrobacter sp.*) has been identified, but the earthworm impact on transient microbial communities and the surrounding soil microbiome is profound.

Methods: Previous culture and molecular based studies found earthworm-associated microbiota unlikely however, this has not been explored using High Throughput Sequencing. Utilisation of Illumina, 454 and Ion Torrent sequencing has enabled production of the highest resolution microbial analysis of host-associated bacteria of any single eukaryotic species to date, including spatial bacterial localisation of the entire *Lumbricus rubellus* organism and impact analysis of a wide range of anthropogenic contaminants and environmental stressors on the basal microbiomic community.

Results: A core bacterial community has been described which is distinct from the surrounding soil. A number of novel species have been associated with the earthworm crop, body wall and hindgut, contravening claims that the earthworm has limited or no impact on ingested soil bacteria. This demonstrate that the host properties impart significant effects on the transient population, demanding further analysis to determine potential symbiotic functionality. However, while a biologically important community has been described, the significant impact of anthropogenic contamination on the host microbiome must be considered given the observed eradication of the *Verminephrobacter* symbiont during the host's exposure to arsenic and the potential subsequent implications on host health.

Declaration

This work has not previously been accepted in substance for a degree and is not concurrently submitted in candidature for any degree.

Signed…… (Candidate) Date……

STATEMENT 1
This thesis is being submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Signed…… (Candidate) Date……

STATEMENT 2
This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed…… (Candidate) Date……

STATEMENT 3
I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed…… (Candidate) Date……

Acknowledgements

Gratitude goes to Pete for his time, supervision, patience and friendship.

To my family for their support throughout this thesis and the years that preceded it.

To my partner Mel, who has been alongside me from the beginning and whose relationship means more than can be expressed.

This work would not have been possible without the support of a number of people; Prof. Andy Weightman and Professor John Morgan in Cardiff, Prof. Dave Spurgeon, Dr. Dan Read and Dr. Anna Oliver in CEH and Prof. Dawn Field from further afar.

I am unquantifiably indebted to (alphabetically) Angharad Williams, Ceri Morris, Craig Anderson, Goncalo Ferreira, Gordon Webster, Luis Cunha, Pierfranchesco Sechi, Marta Novo, Rico Randall and Yaser Sahl for their friendship and scientific guidance.

Thanks must go to both Tim Booth and the whole NBAF/Biolinux team for their support and friendship.

This PhD was supported through funding from the NERC Environmental Bioinformatics Facility.

Contents

Chapter Summaries

The earthworm microbiome and the effect of anthropogenic arsenic contamination (Chapter 3).

This study describes the bacterial community associated with the common terrestrial earthworm *Lumbricus rubellus* utilising High Throughput Sequencing for the first time in Annelidia. Furthermore, the impact of anthropogenic contamination from a disused arsenic mine site on the microbiome was investigated. The stress gradient demonstrated the increasing dysbiosis imparted by the toxic environment and highlights the degree to which we are unaware of contaminant effects on microbial symbiosis.

Examining the microbiome of an extremophilic earthworm living on volcanic geothermal soils (Chapter 4).

The largest of the Azorean Islands (São Miguel, Atlantic Ocean) was formed 30,000 years ago through volcanic action and to this day there remains significant volcanic activity within selected calderas. The island maintains a human population of approximately 150,000 despite locations exhibiting soil temperatures of up to 50ºC, high CO_2 and other toxic gasses, and the increased bioavailability of metal ions.

These geogenic characteristics allowed investigation of potential association between a prominent extremophilic earthworm species (*Amynthis gracilis*) and bacterial communities in the soil environment via a reciprocal cross transplantation of individuals to volcanic and control environments.

The impact of diverse soil geochemistry and host phylogeography on the microbiomic hindgut community of *Lumbricus rubellus* (Chapter 5).

Analysis of earthworms from 17 field sites assessed the high level of variation in soil chemistries and heavy metal contamination in UK soils and validated previous observations across multiple sites. Assessment of the hindgut of *L. rubellus* identified the internal microbiota closely associated with cast material production and how the major chemical properties of soil impact on the earthworm microbiome. This has helped develop knowledge of earthworm microbiota through description of core taxa and those populations associated with environmental and contaminant factors.

The Voxel Worm: Spatial Characterisation of the *Lumbricus rubellus* microbiome (Chapter 6).

It is now understood that earthworm species host a range of microbiota, both those modified during transit from the soil and at least one known vertically transmitted symbiont. Previous studies have focussed upon the earthworm as a whole organism when *in situ* microbes were described. Organ and tissue specific communities were here defined to an unprecedented resolution in fresh and depurated individuals. A number of novel species potentially symbiotic in nature were identified for further study and the effect of *L. rubellus* on the soil microbial environment was elucidated upon.

Abbreviations

Chapter Short-code	DGC	Devon Great Consols Mine Site (Chapter 3)
	AZO	Azores Volcanic Transplant site (Chapter 4)
	LRP	*L. rubellus* Microbial Phylogeography (Chapter 5)
	VXL	The Voxel Worm: Spatial microbial localisation (Chapter 6)
16S		16S Ribosomal RNA
ANOVA		Analysis of Variance
BH-FDR		Benjamini-Hochberg False Discovery Rate Correction
COI		Cytochrome Oxidase I
DGGE		Denaturing Gradient Gel Electrophoresis
FISH		Fluorescent In Situ Hybridization
HTS		High Throughput Sequencing
ICP-MS		Inductively Coupled Plasma-Mass Spectrometry
IT		Ion Torrent Sequencing Platform
KEGG		Kyoto Encyclopedia of Genes and Genomes
NGS		Next Generation Sequencing
OTU		Operational Taxonomic Unit
NMDS		Non-parametric Multi-Dimensional Scaling
T-RFLP		Terminal-Restriction Fragment Length Polymorphism

1
Introduction

1.1 General Introduction

Host-associated microbiota is increasingly understood to contribute to an individual's phenotype. The host's impact on its microbiota and, conversely, the impact of the microbiota on the host can be observed in species at all levels, ranging from humans to lower invertebrates (Li et al. 2008; Ley et al. 2008). This 'two way street' forms the basis of the commonly observed mutualism which can play an important role in a host organism's environmental interactions and potentially its outward phenotype. Exploring these interactions is essential in developing a comprehensive understanding of the organism in question. Furthermore, the interactions an individual has with its external microbial habitat define the local and global impact that species imparts on the worldwide microbial ecosystem.

The importance of the terrestrial soil invertebrate, summarised as earthworms (Order: Oligochaeta), is unequalled in the soil environment (Lavelle et al. 2006; Jana et al. 2010; Blouin et al. 2013). Earthworms are the single largest contributors to the soil invertebrate biomass in many ecosystems and have long been recognized for the benefits they bring to the environment. Gilbert White noted in 1777 that earthworms promote vegetation growth through perforation of the soil and drawing leaf detritus down into it (Russell & Wild 1988). In 1837, Charles Darwin also reported on these effects, followed 35 years later by his manuscript "The Formation of Vegetable Mould through the Action of Earthworms" (Darwin 1896) which has been described as the first publication on the biological nature of soil (Feller et al. 2003). Essential functions performed by earthworms include soil aeration, bioturbation and organic matter fragmentation which are vital for promoting soil structural development and nutrient cycling (Darwin 1896; Heemsbergen et al. 2004; Blouin et al. 2013).

The earthworm plays host to a number of discrete bacterial communities acquired either from the surrounding soil, as with the gut fauna (Drake & Horn 2007), or through vertical parental transmission as observed for the nephridial symbiont *Verminephrobacter* (Pinel et al. 2008). Research stemming from early earthworm microorganism studies demonstrated that the earthworm gut affects a discrete anaerobic environment and plays host to ingested communities containing unique suites of nitrogen fixing bacteria with the capacity to transform the chemistry of the soil as it traverses the gut (Drake & Horn 2007). Expanding the knowledge on the bacterial contribution to the environmentally-essential earthworm taxon is the central concept explored in this thesis.

1.2 Eukaryote–Bacteria Interactions

The interactions between eukaryotic hosts and bacterial communities form an integral part of the observed phenotype and has been the attention of increased research in the previous decade (Bright & Bulgheresi 2010; Moya et al. 2008; Hansen & Moran 2014). Host-community interactions are likely universal, having been observed at all levels of biotic life including Humans (The Human Microbiome Consortium 2012), Mammals (Ley et al. 2008), Plants (Selosse et al. 2004) and Invertebrates (Chaston & Goodrich-Blair 2010). The causative effects of eukaryotic species on the free-living microbial community is also widespread, including influencing the natural fluctuation of soils (Lavelle et al. 2006). The importance of understanding these systems are twofold. Firstly, defining the interactions of the host with external and transient communities can aid in understanding eukaryotic impact on the microbial landscape and the influence on microbially-mediated environmental processes. Alternatively, the obligate and commensal symbiotic species which form part of the *in situ* organism contribute to the observable phenotype. The interactions and contributory effects of bacteria to their host are observed in higher organisms (Humans and other mammals) through to species more applicable to laboratory study (e.g. invertebrates).

Microbiomic research of human-associated bacteria is likely the most in depth studied species to date, with an extensive number of studies profiling organ-specific communities including intestinal (Bäckhed et al. 2005; Walter & Ley 2011), skin (Grice et al. 2009), oral (Bik et al. 2010) and vaginal (Dumonceaux et al. 2009). These studies have been widely performed or integrated into the Human Microbiome Project which summarises the wide range of diverse communities present within the single species (The Human Microbiome Consortium 2012) and is supplemented by a number of studies determining the derivation of these adult communities through investigation of infant microbiota (Palmer et al. 2007; Turroni et al. 2012). Foundational studies defining the basal bacterial populations has allowed further research in determining mutualistic functionality of human-associated species (Bäckhed et al. 2005) and the influence of bacterial symbiosis on metabolic pathways (Li et al. 2008). Conversely, dysbiosis has been linked to pathologies including Inflammatory Bowel Disease (Rigottier-Gois 2013) and obesity (Delzenne et al. 2011). The human model represents a high-water mark for understanding the contribution of bacterial communities to the observed eukaryotic phenotype and the effects that can

potentially disrupt functionality, and is a realistic target when studying other species interactions such as the soil invertebrates.

Termites occupy a similar soil niche to earthworms and demonstrate a functional phenotype through the action of symbiotic bacteria. Termite species which employ wood-feeding perform an important contribution to organic matter degradation and subsequent carbon cycling that has a significant global impact (Heemsbergen et al. 2004). This central role is achieved through the action of bacterial symbionts performing cellulose and xlyan hydrolysing activity in the hindgut community (likely spirochete and fibrobacter species (Warnecke et al. 2007)). The host produces intestinal 'lampbrush' shaped fronds which increase the surface area for communities to propagate upon, raising the bacterial load and thus increasing the functional capacity (Hackstein & Stumm 1994).

Osedax polychaete worms regularly frequent the carcasses of dead whales (whalefall) where they feed upon the bones. The collagenolytic gene pathways required for this process are absent from the eukaryotic host genome but have been identified in two Epsilonproteobacteria endosymbiotic species that associate directly with those species engaged in the feeding process (Goffredi et al. 2007). *Osedax* acquire these endosymbionts via horizontal transfer from subsequent feeding events when the host tissue invades the whalebones, demonstrating important roles that can be facilitated by species which are not vertically transmitted (the most common form of symbiosis (Verna et al. 2010)).

Symbiotic bacterial species are required in the development of a mature immune system in many insects. Potentiation in *Drosophila* relies upon the presence of specific *Wolbachia* species, absence of which results in greater susceptibility to viral infection (Teixeira et al. 2008). *Glossina* (tsetse) flies suffer similarly, where aposymbiotic individuals are highly susceptible to *Escherichia coli* infection (Weiss et al. 2012). Intrauterine development in the absence of their *Wigglesworthia* symbiont causes the host's genome to be deregulated resulting in under expression of immune-response genes and an absence of immune cells, yet reintroduction of cell extracts from the symbiont to the maternal parent can restore the immune system for future offspring.

Pyrrhocorid species (*Dysdercus fasciatus* and *Pyrrhocoris apterus* aka Cotton stainers and Red Firebugs respectively) have increased mortality in the absence of parentally-

applied cocoon-surface symbionts (Coelom et al. 2012). The applied species (notably Actinobacteria species *C. glomerans* and *Gordonibacter sp.*) establish in the M3 midgut region upon hatching and perform essential nutritional functions, namely facilitating the exploitation of a novel food source (sunflower seeds) by the host.

Symbiosis localisation within the host varies, including specific organ-associated species. *Verminephrobacter* is observed solely in the nephridial excretory organ in earthworm species (Pandazis 1931; Schramm et al. 2003)(discussed further below) whereas gut-bound structures found exclusively in the hindgut within species such as termites promote biofilm-like congregations specific to the area (Hackstein & Stumm 1994). *Osedax* symbionts localise specifically to the 'root' section of the body which is embedded within the bone upon which it feeds (Verna et al. 2010).

A microbial community which could reduce environmental stress on the host would be highly beneficial, and host-microbial symbiosis could be seen as either an endpoint (i.e. an important component of the host) or as a stepping-stone in invertebrate evolution through buffering the individual from contaminant stress and enabling the host population to encroach on environments otherwise inhospitable. The aquatic insect species Chironomids host bacterial species that have been identified to function as detoxifiers of lead and hexavalent chromium (Senderovich & Halpern 2013). Upon sterilisation their survival in heavy metal inoculated waters decreased significantly, with this effect being largely reversed when the symbionts are reintroduced. Conceptually, any action which an organism could employ that would result in accumulation of a higher proportion of detoxifying bacteria would provide a practical advantage in extreme environments. This functionality, unavailable to the eukaryotic organism, would create a strong selective pressure for the host to accommodate microbes which reduce the potential toxicity of environmental stressors.

1.3 Physiochemical function of earthworms

In 1 square metre of a favourable soil environment roughly 1 litre of soil is contained within the earthworm population's gut, where 4-10% of total soil is traversed annually (Drake & Horn 2007). Extrapolating this indicates that over 10 years ~50% of soil will have passed through an earthworm and ~90% within 40 years. Within the United Kingdom an estimated 89.5 million litres of soil resides in the earthworm gut at any one time (1 L M^{-2} of favourable UK soil (Barr et al. 1978)) and therefore the

egested material clearly represents the major constituent of soil. Globally the impact that earthworms have on the soil environment is vast and integral to the microbial structure and physiochemical properties that terrestrial plants and animals rely upon. Earthworm tunnelling increases the aeration of soil and water poration, and they cycle organic matter via ingestion of leaf detritus, lining their burrows with nutrient-rich mucous (Edwards 2004). The gut environment differs greatly from the surrounding soil, for instance subjection of the contents to anoxia and pH neutralisation (Drake & Horn 2007). The abundance of N_20 increases in the gut lumen of earthworms, and peaks in the midgut region (5.6 µM *Lumbricus rubellus*, 17.6 µM *Aporrectodea caliginosa*), however the gut maintains complete anoxia (Horn et al. 2003). The levels of organic carbon are higher in the gut contents than the surrounding soil due to secretions of intestinal mucus producing a 'priming' effect (Brown et al. 2000). Transit time of soil from ingestion to egestion is rapid, reported to range from 6-8 hours for *L. rubellus* (Daniel & Anderson 1992) and 2-16 hours for other earthworm species (Brown et al. 2000). This rapid transmission raises question to the amount of change which could be effected upon the soil structure and community during transit.

Large populations of earthworms inhabit active agricultural fields and the interaction between worms and applied herbicides is a constant and dynamic process. This has provided the opportunity to study the effect that earthworms have on soil microbiota under such anthropogenic stressors. Through their burrowing activities it is proposed that *A. caliginosa* stimulate a higher abundance of microbial MCPA (2-Methyl-4-chlorophenoxyacetic acid, a phenoxyalkanoic herbicide) degraders that are typically endogenous to soil (Liu et al. 2011). Transit of soil through the *Eisenia fetida* gut lumen modifies the molecular mass of the substrate humic acids, resulting in enhanced microbial-inhibitory effects by humics than in soils devoid of earthworms (Tikhonov et al. 2011).

Earthworm species vary in their ability to alter/modify metal availability and toxicity in the environment (Sizmur & Hodson 2009; Nahmani et al. 2007). These effects are often observed to influence the plants and animals which also inhabit their immediate environment (Liu et al. 2005; Jana et al. 2010; Sizmur et al. 2011) however these interactions are often too complex to be labelled as beneficial or detrimental. The use of earthworms for remediation has been frequently discussed (Blouin et al. 2013) with examples including reducing the bioavailability of copper

and cadmium in sewage sludge (Liu et al. 2005), however earthworms have also been shown to increase metal availability to plants and other invertebrates (Sizmur & Hodson 2009). The bioremediation of oil spills has been posited because of the increase in microbial degradation pathways enhanced by earthworm presence for petroleum products (Schaefer et al. 2005).

Earthworms implement core ecological processes however it is often difficult to separate the host functionality from roles performed by the microbiota within and surrounding them. The nitrification potential of soil has long been understood to be influenced by earthworms (Day 1950) and more recently earthworms have been proposed as the origin of ~60% of global N_20 release (Drake & Horn 2007). Furthermore, their presence increases soil nitrate content and plant uptake through increased bacterial activity (Jana et al. 2010). However their efficacy as a reservoir for anoxic denitrifiers highlights the difficulty in separating the host from the community (Horn et al. 2003).

1.4 Earthworm-Microbe interactions

1.4.1 The Soil Microbiome

Soil is potentially the most diverse microbial community on Earth (Curtis et al. 2002; Hugenholtz 2002) with high levels of heterogeneity observed in small spatial areas (Frey 2006; Griffiths et al. 2011). Soil microbiota forms a central role in global carbon cycling through incorporation of degraded organic matter into the soil by chemoautotrophic bacteria (Gougoulias et al. 2014), and free living heterotrophic bacteria fixing CO_2 (Santrůcková et al. 2005). Atmospheric nitrogen fixation is solely performed by prokaryotic organisms, containing 10x higher nitrogen than plants globally (Frey 2006). pH is the dominating factor which impacts soil nutrient composition, structure, function and the micro- and macro-biota which inhabits and depends upon it (Kemmitt et al. 2006).

Carbon is released from soils predominantly as CO_2 in oxic environments and CH_4 when anoxic (e.g. flooded), and typically achieved through the action of heterotrophic species or methanogenic bacteria releasing carbon from degrading organic matter. Less than 4% of soil organic carbon is estimated to be found in the biomass of the living microbial population (Anderson & Joergensen 1997; Sparling 1992) but the microbiomic community is central in driving carbon cycling (Geider et al. 2001; Gougoulias et al. 2014).

1.4.2 Earthworm interactions with the soil microbial community

The impact of Earthworm species on the soil microbiome is varied and often contradictory in the literature, arising from both the wide range of soil conditions and earthworm species. In hardwood forest surface soils, invasive earthworm species (various) increased the total microbial biomass of the surrounding habitat (Li et al. 2002). However, earthworms (*Lumbricus terrestris* and *A. caliginosa)* reduced the number of detectable bacteria from the soil and leaf litter when ingested, particularly in the crop/gizzard organs, yet the abundance of *Cytophaga-Flavobacterium* (Bacteroidetes) increases in the hindgut and cast material (Schönholzer et al. 2002; Nechitaylo et al. 2010).

Landfill sites are a major reservoir of anthropogenic methane production (the largest source in the UK) and earthworm presence was found to manipulate the natural microbial community resulting in increased methane oxidation (Héry et al. 2008). The earthworm species *Eudrilus eugeniae* demonstrated high levels of methane production, proposed to be generated by associated Methanosarcinaceae and Methanobacteriaceae, despite the varied amount of organic material upon which the host fed (Depkat-Jakob et al. 2012). Other species (including Lumbricids) did not show this effect, potentially due to differing redox potentials within the alimentary canal.

Alternatively, earthworms have long inspired the view of generalist antimicrobial functionality (Lassalle et al. 1988). Fluid isolations from whole earthworms show antibacterial and antifungal properties that affect the surrounding soil which transitions the gut (Vasanthi & Singh 2013; Bhorgin & Uma 2014). Specifically, the coelomic fluid of *E. fetida* demonstrates significant antibacterial effects (Lassalle et al. 1988; Pan et al. 2003). A major anthropic use of earthworms is in vermicomposting (maintaining a population of earthworms (often *E. fetida*) to convert organic waste into fertilizer) due to increases in polysaccharide and other nutrient content, resulting in stimulated plant growth (Sharma et al. 2005). Vermicomposts also demonstrate antifungal properties mediated by *E. fetida* stimulation of bacterial populations exhibiting chitinase activity (Yasir et al. 2009) in addition to those generalist antibacterial effects described above.

1.4.3 Microbiota within the Earthworm

Earthworm species almost ubiquitously host a single *Acidovorax*-like extracellular symbiotic species in the ampulla of the osmoregulatory nephridia which has been described as *Verminephrobacter* (Pinel et al. 2008). The nephridial symbiont is considered a standard constituent of earthworm anatomy as it can be found in almost every species of earthworm (Davidson et al. 2013). It is a vertically transmitted symbiont which has diversified with the host over significant evolutionary time (estimates suggest that *Verminephrobacter* has been associated with Lumbricidae since the earthworm family split 62-132 million years ago, and displays the characteristic accelerated mutation rate found in other symbionts (Lund et al. (2009)). *Verminephrobacter* has had many functional roles postulated based upon its historic association (Davidson & Stahl 2008). Primarily a role in nitrogen and protein recovery was hypothesised due to anatomical location and nephridial functionality (Pandazis 1931; Schramm et al. 2003), however this has been questioned based upon the absence of extracellular proteases from the *Verminephrobacter eisinea* genome and analysis of aposymbiotic-reared individuals (Lund et al. 2010). Absence of the symbiont is detrimental to host fitness and fecundity in low nutrient conditions for the earthworm species *Aporrectodea tuberculata*, highlighting the functional role that it must provide to the host (Lund et al. 2010).

Earthworms create a microhabitat within their gut to which transient soil microbes are subjected to (Figure 1.1). Major characteristics are the early secretion of calcium carbonate from the calciferous glands in the anterior section of the earthworm that rapidly reduces the acidity to a more neutral level, and the imposition of anoxia throughout the gut lumen which provides a unique reservoir from the external soil (Drake et al. 2006). Furthermore, the secretion of carbon and nitrogen metabolites into the hindgut lumen deliver differing energy sources from the soil (Wüst et al. 2011).

Previous studies of earthworm-associated microbiota at the inception of this thesis were based largely upon culture and molecular techniques and there was yet to be analysis based upon the now widely utilised High Throughput Sequencing (HTS) technologies. Early studies demonstrated that during transit through the earthworm gut, the number of microorganisms increased by up to 1000 times (Edwards & Fletcher 1988) and caused the concentration of soluble organic carbon to increase (Daniel & Anderson 1992). Earthworm microbiota has been predicted to solely

represent the ingested soil and plant microbiota (Morgan H. 1988; Brown & Doube 2004) however scanning electron microscopy identified hindgut adhered bacterial species in *L. terrestris* (Jolly et al. 1993). It has been demonstrated that Firmicute and Psuedomonas species increase in abundance along the gut tract of *L. rubellus* (Furlong et al. 2002) however more recently, Terminal Restriction Fragment Length Polymorphism (T-RFLP) analysis of *L. terrestris* has shown highly similar microbial profiles in each compartment (transient gut contents, soil and casts), indicative of a soil-derived microbiome (Egert et al. 2004).

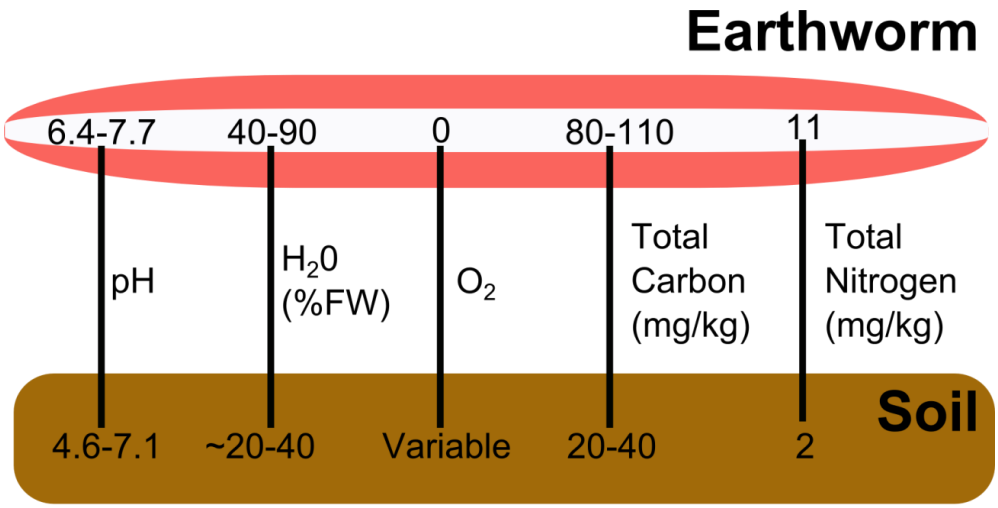


Figure 1.1 Differing conditions between the earthworm gut microhabitat and the surrounding soil. FW: Fresh Weight. Adapted from Drake and Horn (2007)

The majority of microbial activity associated with the earthworm is likely contributed by the transient community being 'activated' by the unique environment it encounters during transit, where the gut acts as a 'fermenter' (Wüst et al. 2011). This encouraged Clostridia and Enterobacteriaceae communities via metabolism of mucus- and plant-derived saccharides, resulting in nitrogenous gas production. Distinct taxonomic groups have been identified at higher levels in *L. terrestris* and *A. caliginosa* casts, notably Bacteriodetes species (Nechitaylo et al. 2010) where its role in organic matter digestion is posited. The gut wall of several earthworm species have more recently been described as cultivating Proteobacteria, Firmicutes and an Actinobacterium species (Thakuria et al. 2010). Symbionts in *A. caliginosa*, *L. terrestris*, and *E. fetida* have been postulated due to the detection of bacteria absent from the surrounding soils including Enterobacteriaceae, Flavobacteriaceae, Pseudomonadaceae, Sphingobacteriaceae (Bacteroidetes) and several Actinobacteria

(Byzov et al. 2009). *E. andrei* mediates a reduction in soil diversity but an increase in microbial activity through effects on the transient community (Gómez-Brandón et al. 2011). The gut of *A. caliginosa* has a complex antimicrobial effect on transient species, where application of isolated gut fluid was shown to directly eradicate particular bacterial species *in vivo* whilst others survived or proliferated (Khomiakov et al. 2007). This effect varied by body position, with different taxonomic targets being suppressed at anterior and mid sections. In the anterior sections a reduction in *Bacillus megaterium* and *Alcaligenes faecalis* numbers were observed however no microbicidal activity was found in the posterior section, instead increasing the colony forming capabilities of various *Psuedomonas* species. This correlates with observed rises in hindgut microbial diversity and load (Wust et al. 2009).

As is evident, there is significant controversy surrounding the existence of earthworm-associated bacterial species and communities. Utilising the latest High Throughput Sequencing technologies will allow elucidation upon the microbiome to unprecedented depth and develop a greater understanding of this biotic system.

1.5 Earthworms as Stress-Tolerant Soil Sentinels

As soil organisms, earthworms are in direct contact with changing soil chemistries and bio-available pollutants, and as such it is essential to understand how these soil organisms adapt both physiologically and genetically to these highly variable environments. Significant work has resulted in earthworms being regarded as model species for assessing environmental toxicity over several decades (Spurgeon et al. 2003) since the development of the 'acute earthworm toxicity test' (OECD 1984) and are of particular use because of the broad global distribution where edaphic and climatic conditions are favourable.

The wide prevalence of earthworm species also extends to highly toxic environments. The high level of interaction which an individual earthworm has with the soil environment, both epidermally and due to ingested material, results in high levels of contaminant contact. Despite this, earthworm species are often present in numbers indicative of tolerance to environments that are potentially toxic to many other species. Field populations of earthworms can be observed living in soils with metal/metalloid levels often exceeding LC50 laboratory-determined acute toxicity values by several orders of magnitude (Langdon et al. 2001; Morgan & Mariño 1999). Mechanisms through which earthworm species (and other invertebrates) are capable

of inhabiting such habitats have been widely studied but remain to be established definitively due to the diverse range of environmental stressors.

Genetically encoded pathways have been identified for various invertebrate species in relation to tolerance to metal toxicants, including copper and cadmium (Morgan et al. 2007). The ability to tolerate high levels of toxicity can be achieved through regulation of existing tolerance mechanisms i.e. the activity of Metallothionein, a metal ion binding peptide (Stürzenbaum et al. 2001). This performs a multitude of roles including the sequestering of metal into body tissues thus reducing toxicity (Coyle et al. 2002) and is a recognized technique of earthworm species (Andre et al. 2010; Nahmani et al. 2007).

1.6 Application of Next Generation Sequencing to Community Analysis

The introduction of novel 'sequencing-by-synthesis' approaches to DNA study in the past decade have radically changed the possibilities for analysing microbial communities, allowing much deeper investigation than previously thought possible. The technologies have also improved significantly since commencement of this work and have resulted in multiple platforms being utilised in this thesis, as described.

Microbiomic analysis is most commonly focused upon a variable gene region of 16S rRNA which allows taxonomic description and statistical analysis of the bacterial community (Huse et al. 2008; Hugenholtz 2002). Early sequencing approaches focused on the 454-Roche or Illumina platforms (Review: Mardis 2008). The 454 quickly became the most favourable for bacterial amplicon analysis, producing 10s of thousands of sequences per sample in comparison to the handful possible by classical Sanger sequencing, while maintaining a reasonable sequence length (200-400bp). This was advantageous for diversity and annotation as the higher amount of information per sequence allowed more accurate description in comparison to the shorter reads from the Illumina platform (35-75bp). The Illumina platform however, generated significantly more reads (<1 Billion) at a comparable cost to the fewer 454 FLX (~400 Thousand), and so platform selection was application specific.

Significant improvements to each platform have since occurred, with Illumina capable of generating ~500bp amplicons (from 2x300 paired end sequencing) at the same read count as previously. However, although the 454 FLX platforms have improved the read length to upwards of 900bp, the increased output has plateaued at ~1 Million per run, reducing the depth of sequencing. Concurrently, newer

technologies have emerged, notably the Ion Torrent platform (Review: Quail et al. 2012) which while similar to the 454, produces high read counts at shorter read length (~200bp). It has also found difficulty with base calling quality (similar to 454 data output), but is still to have error model correction as has been found to improve 454, raising cost/benefit questions per experiment (Bragg et al. 2013), particularly to the hugely accurate Illumina technologies (Quail et al. 2012).

Today these three platforms are still the most widely used for microbial amplicon based community analysis and the significant advantages over classical and molecular techniques have propelled the understanding of the microbial world to previously unfathomable depths.

1.7 Research Aims

The contemporary knowledge on the earthworm microbiota is predominantly based on classical microbiology and molecular techniques. Using the latest High Throughput Sequencing (HTS) technologies will allow a greater depth of analysis than previously achievable, allowing examination at a much higher resolution. This study is focused predominantly upon the species *Lumbricus rubellus* as a model for all earthworms. This is due to their prevalence and distribution in the United Kingdom, their predilection for contaminated environments which is an aspect here focused upon, and the quantity of literature available on this popular research species.

Investigation was undertaken to accomplish the following goals:

- Define the 'earthworm microbiome' with high resolution Next Generation Sequencing. **(Chapters 3,4,5,6)**
- Identify the advantages/disadvantages between HTS and classical techniques. **(Chapter 3)**
- Establish the differences between the surrounding soil microbiome and the internal and/or transient gut community. **(Chapters 3,4,6)**
- Examine the impact that chronic association with environmental contaminants has on the earthworm microbiome. **(Chapters 3,5)**
- Observe the acute influence of extreme environments on the earthworm-associated microbiota in rapid adaptation. **(Chapter 4)**
- Advance the understanding of the functional capacity of transient and associated earthworm microbiota and their contribution to the host phenotype. **(Chapters 2,4)**

2
Materials and Methods

2.1 Laboratory Based Methods

2.1.1 Soil Processing

2.1.1.1 Metal Analysis and Geochemical Characterisation

Soil characterisation was performed and provided by collaborators for Chapters 3, 4 and 5 as described in Kille et al. (2013). Briefly, triplicate samples were taken from the epigeic level (surface 10 cm), dried at 80ºC and 1 g sample of the processed soil was analysed or total concentrations of respective metals as described for earthworm tissues (2.1.2.2).

Soil organic matter content was measured using furnace combustion at 375°C. pH determination was via analysis by electrode from a 1:2.5 by volume soil:water mix to assess pH (Spurgeon et al. 2011; Allen 1989).

2.1.1.2 Soil Collection for DNA Sequencing

Soil samples for each replicate were collected from the epigeic surface layer (10 cm) in a 1 metre square 'W' formation and hand mixed in a sterile bag before being divided into 3 statistical replicates. This was either briefly stored at 4ºC followed by DNA extraction within 24 hours or immediately frozen at -80ºC until DNA extraction could be performed.

2.1.1.3 Soil DNA Extraction

In almost all cases, DNA extraction from soil was performed in exact concordance with the earthworm extraction method described below. In the Devon Great Consols study (Chapter 3) a soil-optimised extraction method was performed using the Soil PowerBio kit (MO BIO Laboratories, CA, USA) to manufacturer's specifications, however the method was highly alike to subsequent methods, and based upon the similar bead-beating methodology.

2.1.2 Earthworm Processing

2.1.2.1 Sample Collection

Earthworm individuals were collected and processed to inactivate the microbial community and eukaryotic host with immediate speed in all cases. At the sampling site, individuals were washed in 2 subsequent troughs of distilled water to remove external soil, followed by drying on filter paper. They were submerged in liquid nitrogen until frozen, placed into individual packages and stored at below -80ºC. Prior to DNA extraction, individuals were ground using a sterile pestle and mortar to

homogenise the samples (excluding Chapter 6 where samples were manually dissected as described).

2.1.2.2 Metal Analysis

Earthworm metal analysis was provided for Chapter 5. Briefly, tissue was homogenised per individual and 0.5 g measured via acid hydrolysis and Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) (Perkin Elmer 4300DV, Cambridge, UK) as previously described (Spurgeon et al. 2011). Quality control within the analyses was conducted using the standard reference material ISE 192 (International Soil Exchange, The Netherlands).

2.1.2.3 Earthworm DNA Extraction

Total DNA was extracted from randomly selected earthworms from each site, visually identified to be the correct species with later confirmation via COI barcode sequencing (2.1.2.4.3).

2.1.2.3.1 'Qiagen' DNA Extraction

DNA extraction was performed to manufacturer specifications using the Qiagen blood and tissue extraction kit (Qiagen Inc., Crawley, UK) with omission of an optional proteinase K digestion and replacement with bead-beating. Homogenised sample (~0.1 g) was mixed with 200 µl buffer ATL (Qiagen), ~0.5 g 0.1 mm glass beads and ~20 1.0 mm zirconia/silica beads (Biospec products Inc (Bartlesville, Oklahoma, USA)) and placed into 2 ml screw-cap tubes and homogenisation was performed using an MPBio FastPrep-24 tissue and cell homogeniser (Solon, Ohio, USA). The resultant supernatant was utilised in the downstream extraction with the Qiagen Blood and Tissue kit. DNA was quantified using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE) prior to amplification of the 16S rRNA gene for sequencing.

2.1.2.3.2 Automated DNA Extraction

DNA extraction for Chapter 5 (LRP) was performed from a ~10 mg of frozen tissue sample taken from a 1 cm section of the tail of each individual with a sterile scalpel by automated DNA extraction (Tepnel Ltd, Manchester, UK) using the Nucleoplex Plants Tissues DNA Extraction Kit (Nucleoplex, Manchester, UK).

2.1.2.4 DNA Sequencing Preparation

PCRs were performed in an aseptic UV cabinet with sterile plasticware and nuclease-free molecular-grade H_2O. PCR reagents utilised are referenced per chapter but were changed to utilise Hot Start *Taq* polymerase for high sample throughput studies.

2.1.2.4.1 Promega GoTaq PCR Mixture (Promega UK, Southampton, UK)

Buffer	10 µl
$MgCl_2$	3 µl
BSA	1 µl
dNTPs	0.5 µl
Forward primer (10 uM)	1 µl
Reverse primer (10 uM)	1 µl
Taq polymerase (5 U)	0.25 µl
HPLC grade H_2O	32.25 µl
Template	1 µl
Total	50 µl

2.1.2.4.2 PCRBio Taq Hot Start PCR Mixture (PCR Biosystems Ltd, London, UK)

Buffer	10 µl
Forward primer (10 uM)	1 µl
Reverse primer (10 uM)	1 µl
HS Taq polymerase (5 U/µl)	0.25 µl
HPLC grade H_2O	36.75 µl
Template	1 µl
Total	50 µl

2.1.2.4.3 Cytochrome Oxidase I (COI) genotyping

Amplification and sequencing of the Cytochrome Oxidase I 'barcode' gene was performed for species and lineage confirmation using the Promega GoTaq reagents (2.1.2.4.1).

Forward Primer (LCO-1490): 5'-GGTCAACAAATCATAAAGATATTGG-3'
Reverse Primer (HCO-2198): 5'-TAAACTTCAGGGTGACCAAAAAATCA-3

PCR procedure:

94ºC	5 minutes	
94ºC	30 seconds	35 Cycles
52ºC	30 seconds	
72ºC	60 seconds	
72ºC	10 minutes	

2.1.2.4.4 Denaturing Gradient Gel Electrophoresis

Sample preparation was performed using Promega PCR mix (2.1.2.4.1) under the following conditions:

Forward Primer 357F (GC clamp): 5′-CGCCCGCCGCGCCCCGCGCCCGGCCCGCCG CCCCCGCCCCACTCCTACGGGAGGCAGCAG-3′

Reverse Primer 907R: 5'-CCGTCAATTCMTTTGAGTTT-3'

PCR procedure:

95ºC	2 minutes	
94ºC	30 seconds	30 Cycles
52ºC	30 seconds	
72ºC	90 seconds +1 per cycle	
72ºC	5 minutes	

DGGE was performed as described by Webster et al. 2006, using a DCode™ Universal Mutation Detection System (Bio-Rad Laboratories, Hercules, CA) and 1mm-thick (16 x 16cm glass plates) 8% (w/v) polyacrylamide gels (Acrylogel 2.6 solution, acrylamide: N,N0-methylenebisacrylamide (37:1) BDH Laboratory Supplies, Poole, UK) with a gradient of formaldehyde-urea denaturant between 30 and 60%. Gels were poured with a 50 mL volume Gradient Mixer (Fisher Scientific, Loughborough, UK) and prepared with 1x TAE buffer (pH 8; 40mM Tris base, 20mM acetic acid, 1mM EDTA).

Electrophoresis was at 200V for 5 h (with an initial 10 min at 80V) at 60ºC in 1x TAE buffer. Polyacrylamide gels were stained with SYBRGold nucleic acid gel stain (Molecular Probes) for 30 min and viewed under UV. Images were taken on GeneGenious Bio Imaging Systems (Syngene, Cambridge, UK).

2.1.3 Next Generation Sequencing

Three systems of DNA sequencing were used to generate the data in this report. The downstream effect on the data is discussed and contrasted in Chapter 7. In all cases, triplicate PCRs were performed for each sample before pooling to minimise PCR bias. The amplicon regions sequenced are summarised in Figure 2.1.

2.1.3.1 454 Pyrosequencing

454 Titanium, FLX+ and FLX++ sequencing (Roche, Branford, CT, USA) was completed by Research and Testing laboratories (Lubbrock, USA) and sample amplification was performed by them to the following specification to obtain approximately 10,000 reads per sample. Primers were incorporated with 12bp barcode and 454 sequencing adaptors to system specification (Appendix 1a, 1b).

Promega PCR mix (2.1.2.4.1) was used with the following specifications:

Forward Primer (357f): 5'-CCTACGGGAGGCAGCAG-3'
Reverse Primer (907r): 5'-CCGTCAATTCMTTTGAGTTT-3'

PCR Procedure:

95ºC	5 minutes	
95ºC	30 seconds	35 Cycles
54ºC	40 seconds	
72ºC	60 seconds	
72ºC	10 minutes	

2.1.3.2 Ion Torrent Sequencing

Ion Torrent sequencing was performed on the Personal Genome Machine (PGM, Life Technologies, UK) using chip specification 316 (Centre for Ecology and Hydrology, Wallingford, UK). Primers were adapted with linker and barcode sequences as per manufacturer specification (Appendix 1c). Triplicate sample pools were quantified using Quant-iT PicoGreen dsDNA Assay (Invitrogen, Darmstadt, Germany) with a Tecan plate reader (Tecan UK Ltd, Reading, UK) and equimolar pooled to produce each library.

PCRbio PCR mix (2.1.2.4.2) was used with the following specifications:

Forward Primer (357F): 5'-CCTACGGGAGGCAGCAG-3'
Reverse Primer (518R): 5'-ATTACCGCGGCTGCTGG-3'

PCR Procedure

95ºC	5 minutes	
95ºC	30 seconds	35 Cycles
54ºC	40 seconds	
72ºC	60 seconds	
72ºC	10 minutes	

2.1.3.3 Illumina MiSeq Sequencing

Sequencing was performed on the MiSeq platform (Illumina United Kingdom, Essex, UK) to the protocol described by Kozich et al. (2013) at the Centre for Ecology and Hydrology (Wallingford, UK). Primers were adapted with linker and barcode sequences as per manufacturer specification (Appendix 1d). Triplicate sample pools were quantified using Quant-iT PicoGreen dsDNA Assay (Invitrogen, Darmstadt, Germany) with a Tecan plate reader (Tecan UK Ltd, Reading, UK) and equimolar pooled to produce each library.

PCRbio PCR mix (2.1.2.4.2) was used with the following specifications:

Forward Primer (357f / V3): 5'-CCTACGGGAGGCAGCAG-3'
Reverse Primer (907r / V5): 5'-CCGTCAATTCMTTTRAGT-3'

PCR Procedure

95ºC	5 minutes	
95ºC	30 seconds	30 Cycles
55ºC	40 seconds	
72ºC	60 seconds	
72ºC	10 minutes	

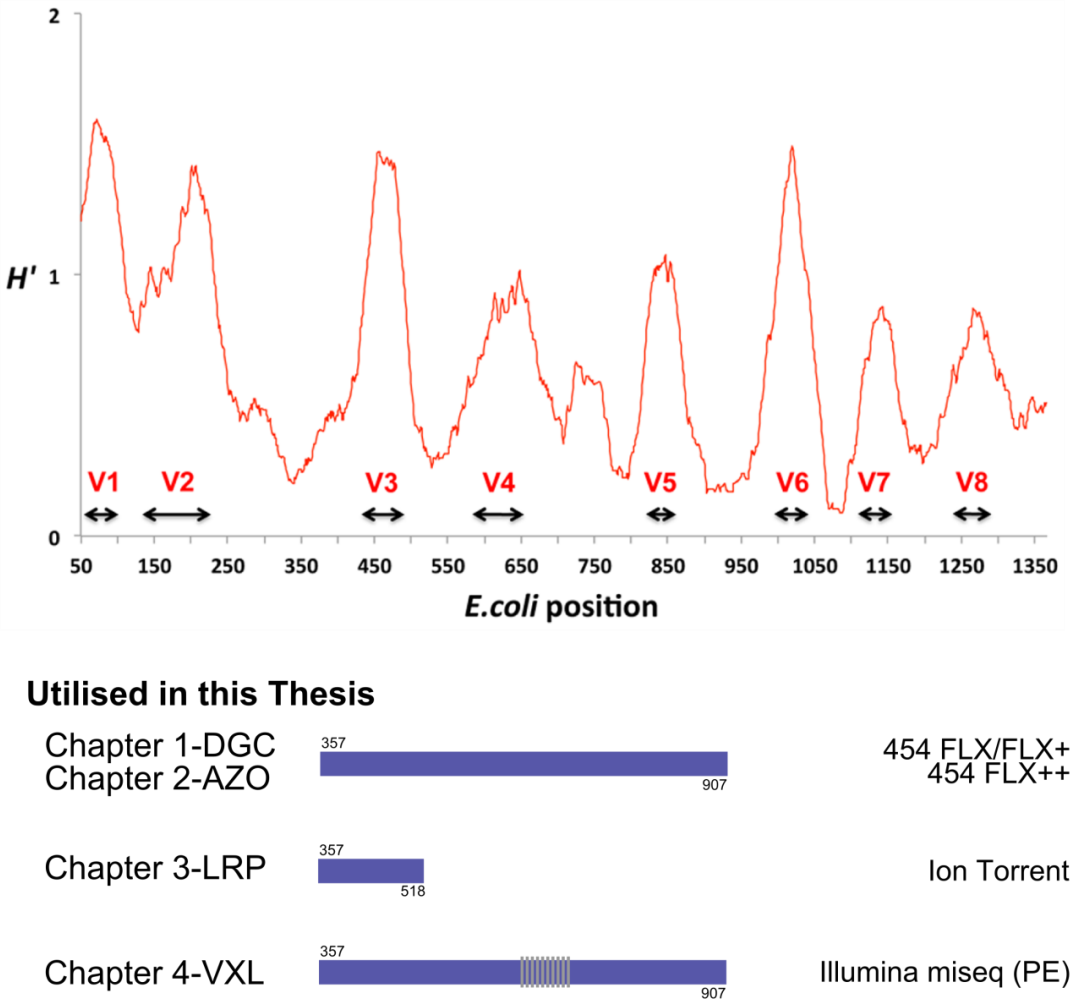


Figure 2.1 - Amplicons and sequencing technologies which were used in this thesis. Upper graph details variability of the 16S rRNA SSU gene at base positions (*E. coli* reference) with the commonplace variable name indicator using the Shannon H' measure of variability. Below are primers utilised in this thesis and the sequencing instrument utilised. Dashed lines on Chapter 4 amplicon indicate overlap position of paired-end reads.

2.2 Informatic Analysis

Differential treatment was applied to each dataset on a case-by-case basis as described in each respective chapter, using the following conventions and software packages. Sequence data was received in fastq format or the sff pre-cursor from 454 sequencing. All processing and analysis was performed on the Bio-Linux 8 operating system (Field et al. 2006) on a local compute cluster.

2.2.1 Sequence Quality Assessment

Initial superficial assessment was performed using FastQC (Andrews 2014) to rapidly determine sequencing success and to identify if any biases are apparent. Where required, samples were demultiplexed using Qiime (Caporaso et al. 2010) to reclassify samples according to their artificially added barcodes.

2.2.1.1 Denoising

The most appropriate method of correcting inherent sequencing error in 454 and Ion Torrent data is disputed but can be achieved with a range of tools summarised in Figure 2.2. In Chapter 3, Acacia was utilised but in Chapters 4, 5 and 6 USEARCH/UPARSE was applied to remove erroneous data (discussed further in respective Chapters).

Software Package	Note	Reference
AmpliconNoise (Prior: PyroNoise)	Most accurate and computationally intensive, based on 454 binary flowgrams.	(Quince et al. 2011; Quince et al. 2009)
Denoiser (Qiime)	More rapid and lower processing requirement but less accurate.	(Reeder & Knight 2010)
Acacia	Similar capabilities to 'Denoiser' but with fewer indel errors. Incorporates the 'Quince' model from AmpliconNoise.	(Bragg et al. 2012)
USEARCH/UPARSE	Not specifically a denoising program but circumvents the requirement through algorithm implementation.	(Edgar 2013)

Figure 2.3 - Software packages utilised in 'denoising' sequence data.

2.2.1.2 Trimming and Filtering

Reduction of sequence quality at the 3' terminus is commonly required in 454 and Ion Torrent sequence data. With 454 data, removing bases from the first instance of low quality base calls (<25 Phred score), homopolymers >6 and discarding reads >3 s.d. from mean retained high quality sequences for downstream analysis. Ion Torrent produced a smaller size range therefore all reads were quality filtered and trimmed to 150 bp (shorter reads discarded) due to poor mean quality (<25 phred score) being detected beyond that point.

Due to the use of broad spectrum primers, in some datasets a proportion of 18S rRNA contamination was observed originating from the earthworm host and the commonly observed eukaryotic parasite *Monocystis agillis*. These sequences were informatically removed prior to any downstream assessment.

2.2.2 OTU generation and annotation

Generation of Operational Taxonomic Units (OTUs) as appropriate substitutes for defined species was performed by UCLUST/USEARCH (Edgar 2010; Edgar 2013). Input parameters vary in relation to differing sequence length, quality and technology, and are described per-chapter. Although 0.97 sequence identity of the 16S rRNA sequenced region was the default parameter for OTU generation, statistical interpretation at 0.88 and 0.94 was performed where appropriate. For samples to be compared equally each independent sample was subsampled to the minimum sequencing count of the respective dataset.

OTUs were annotated according to the latest (05/2013) Greengenes database (DeSantis et al. 2006) using BLAST (Altschul et al. 1990) and the Qiime software pipeline. Sequence alignment for phylogenetic based diversity analysis used FastTree (Price et al. 2009).

2.2.3 Data Analysis and Interpretation

Analysis varies by chapter and necessity; however core analysis was performed as follows.

Calculation of intra-sample (alpha) and inter-sample (beta) diversity was performed using core Qiime functionality (Caporaso et al. 2010). Statistical analysis was performed using R (R Core Team & R Development Core Team 2013) including the Vegan package (Oksanen et al. 2011). To describe and compare community structure

Shannon diversity, Chao1 richness and observed species metrics' were calculated with QIIME.

To visually examine the relationship between samples, Non Metric Multidimensional Scaling (NMDS) from unifrac distances (Lozupone & Knight 2005) was performed. Network analysis was performed with QIIME and analysed with Cytoscape (Shannon et al. 2003). Sequence alignment was performed with ClustalW (Larkin et al. 2007) and phylogenetic tree generation was computed as per chapter using MEGA (Tamura et al. 2011).

Metagenomic Inference was performed using PICRUSt (Langille et al. 2013) using uclust OTU generation (Edgar 2010) and taxonomic assignment with Greengenes database (05/2013) (DeSantis et al. 2006).

2.2.4 Data Visualisation

Figures were mostly generated using R (R Core Team & R Development Core Team 2013) and particularly utilising the ggplot2 package (Wickham 2009). Fixed axis phylum plots were produced in Excel (Microsoft, USA). Where appropriate, biom OTU tables were assessed using STAMP (Parks et al. 2014). The network in Chapter 3 was generated in Cytoscape (Shannon et al. 2003). Post-production of all figures was performed using Inkscape (www.inkscape.org).

3

The earthworm microbiome and the effect of anthropogenic arsenic contamination

See also:

Pass, D.A., Morgan, A.J., Read, D.S., Field, D., Weightman, A.J. and Kille, P. 2014. The effect of anthropogenic arsenic contamination on the earthworm microbiome. *Environmental microbiology*.
DOI: 10.1111/1462-2920.12712

3.1 Chapter Introduction

This chapter describes the effect of external contamination on the microbiome of the *Lumbricus rubellus* earthworm species. Primarily, given the significant environmental differences between the internal host and the external soil, establishing the differences between the microbiome associated with the earthworm and the surrounding soil was pertinent. The host-imparted conditions encountered during gut transit represent a higher moisture, more neutral pH and fully anoxic environment in contrast to the surrounding soil habitat (Drake & Horn 2007). Furthermore, the secretion of organic metabolites from the host into the hind gut can 'prime' particular microbes causing community shifts (Wüst et al. 2011). The rapid transit of ingested material and previous low-resolution studies suggest an inconsequential effect of microbial passage however the observable effects that earthworms have on the microbial diversity and structure of soil prompted the further investigation here described.

L. rubellus is an endogeic species ubiquitous across the UK. The organisms are essential in the maintenance of healthy soils due to their role in aeration, organic matter degradation and bioturbation, and their role in prompting soil structure development and nutrient cycling is core to agricultural practices (Drake et al. 2006; Blouin et al. 2013). They are commonly referred to as a 'sentinel' species due to their resilience to environmental stressors commonly found to be toxic to other species (Stürzenbaum et al. 2009; Donnelly et al. 2014). Their presence in and surrounding the disused arsenic and copper mine site, Devon Great Consols (Tavistock, UK), is such an example and allows an *in situ* examination of the effect of high contamination on the host microbiome.

Arsenic is a potent metalloid element commonly present in sulphurous minerals but can also be found in many forms in the soil i.e. oxides, arsenites and arsenates, that are influenced by other elemental abundance (phosphorus, aluminium, iron) and with pH and organic matter content (Langdon et al. 2003).

Earthworms are often labelled 'extremophiles' due to regularly occupying habitats with extreme geochemical gradients and high anthropogenic contamination (Morgan et al. 2007). *L. rubellus* demonstrates an acquired ability to survive in high-arsenic environments. Control populations of *L. rubellus* have been shown to have a significantly reduced survival rate and rapidly lose body condition when exposed to

sodium arsenate compared to mine site resident populations (Langdon et al. 2001). Additionally, they avoid arsenic treated soils when provided with an uncontaminated alternative whereas individuals sourced from arsenic contaminated sites are less discriminating and inhabit the contaminated soil more readily. The adaptation is preserved in subsequent F1 and F2 generations of laboratory-bred offspring thus demonstrating a genetic basis for arsenic resistance (Langdon et al. 2009). Beyond the genetic investigation of the host, the associated microbial contribution has never been assessed.

This study assesses the microbial community associated with *L. rubellus* and the surrounding soil across 7 levels of arsenic and other metal contamination from an off-site control to ~400x higher. This aims to understand the basal state of the earthworm microbiome, the variance between this and the soil community (if any) and the effect that anthropogenic contamination has on the bacterial populations.

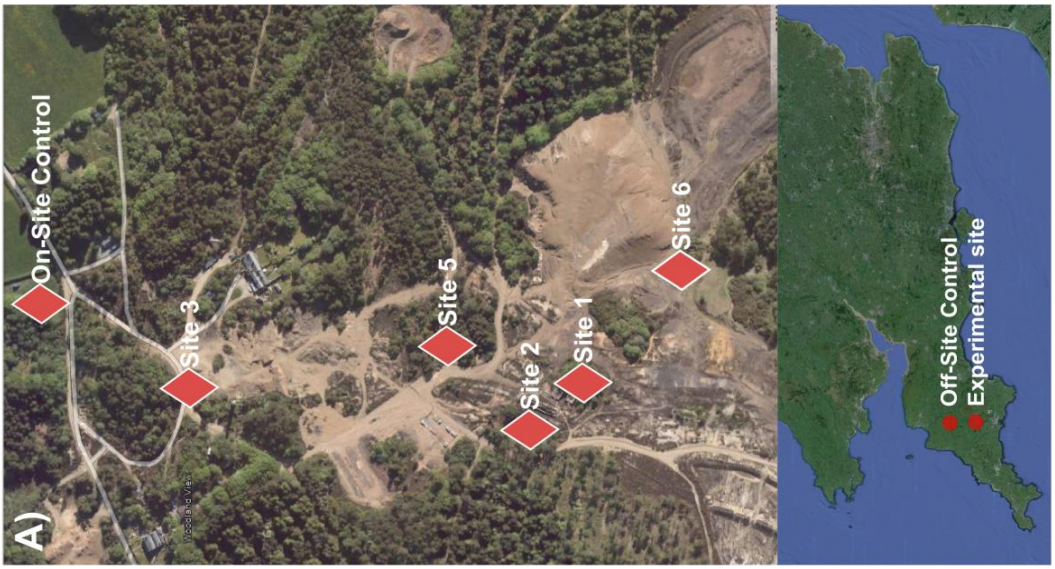
3.2 Methodological Approach

3.2.1 Site Description and Experimental Design

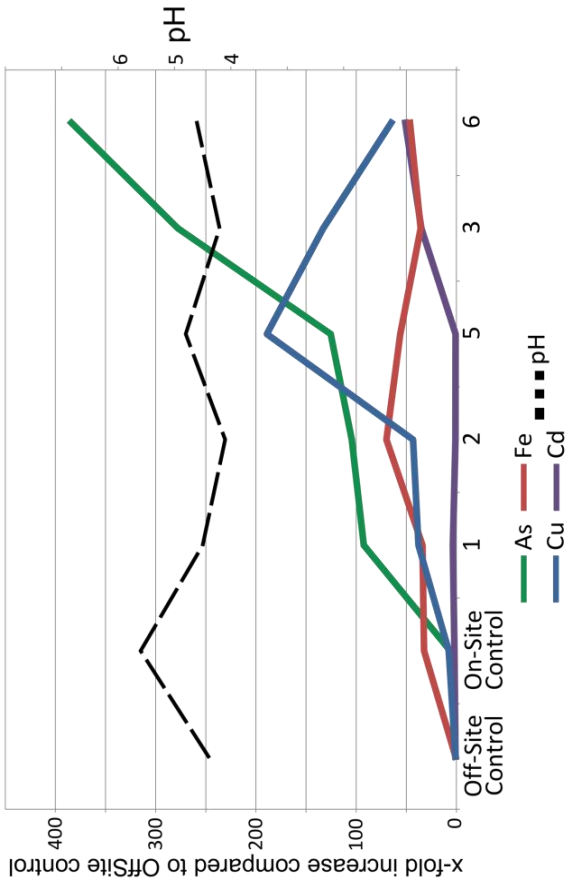
Lumbricus rubellus and soil samples were obtained from the disused Devon Great Consols mine site in the Tamar Valley, Devon, South-West UK (Mine centre: Latitude: 50.538456, Longitude: 355.777252). The site has historically mined copper then later arsenic and an extreme arsenic gradient is still observed at discrete site locations, as has been previously documented (Kille et al. 2013). Soil characterisation was performed with metal content analysed via aqua regia digestion (Chapter 2.1.1.1) (Figure 3.1c). pH varied within small boundaries although was independent of the arsenic gradient. Five sites were identified within the mine in addition to two 'clean' reference sites. These were at a site adjacent to the contaminated area, which displays the areas relatively increased arsenic level (On-Site Control) and a site 20 km distant from DGC which was outside the geological area of arsenic rich soils present in the Tamar Valley (Off-Site Control, Latitude: 50.688863 Longitude: 355.75955).

Earthworms were visually identified as *L. rubellus* with later confirmation via COI barcode sequencing (Chapter 2.1.2.4.3) and immediately processed (Chapter 2.1.3.1). Soil samples were collected from the epigeic surface layer, chilled and DNA extracted within 24 hours (Chapter 2.1.1.2). DNA extraction for earthworm samples was performed to the Qiagen Blood and Tissue protocol (Chapter 2 1.2.3.1) and soil samples extracted to the MoBio Soil PowerBio protocol (Chapter 2.1.1.3).

Samples were confirmed for extraction success and amplification viability via DGGE (Chapter 2.1.2.4.4 and Appendix 2). Amplification of the V3-4 'barcode' region of 16S rRNA was performed using Promega GoTaq (Chapter 2.1.2.4.1) with unique barcodes included in the primer design for post-sequencing sample identification, followed by submission of 56 samples for 454 Titanium/454+ sequencing (Chapter 2.1.3.1).



B)



C)

Site	Al	As	Ba	Cd	Co	Cr	Cu	Fe	Mn	Ni	Pb	Zn	pH	LOI
Off-Site Control	7,840	50	37	0	4	10	14	1,420	427	4	21	69	4	15
On site Control	21,500	310	46	0	4	32	107	45,800	585	28	68	140	6	17
1	9,430	4,620	47	1	20	11	529	48,000	430	16	61	134	5	4
2	6,060	5,220	70	0	53	7	606	99,200	802	28	191	164	4	50
5	17,300	6,270	46	0	26	18	2,647	79,600	630	22	225	277	5	30
3	4,940	13,888	70	7	21	8	1,860	50,615	686	9	99	120	4	66
6	13,600	19,200	71	10	4	20	910	65,900	262	10	148	63	5	17

Figure 3.1 - **Topological Site map and Soil Chemistry. (A) Aerial image showing the location of the 6 sample sites and Off-site control (GPS locations in main text). (B) Chart relating proportional increase of major soil metals to Off-site control, alongside pH. (C) Full quantification of measured environmental variables via aqua regia digestion (metals) Loss on ignition (g) and pH.**

3.2.2 Bioinformatic Methodology

The informatics processing steps performed are described in detail with source references in Chapter 2 and sample descriptions in Appendix 1a. Approximately 1.2 million reads were produced from 1 454 FLX sequencing run and supplemented with 1 FLX+ run. After quality filtering 550,179 were utilised for analysis. Each amplicon sample was subsampled to 2,811 reads which was minimum level while maintaining 3 replicates. A brief overview of the steps taken follows:

Stage	Chapter	Software	Description
Input Data	2.1.3.1	454 Sequencing	~1,200,000 sequence reads (530320 FLX+ reads expanded with 681,891 FLX Titanium reads)
Quality control	2.2.1.2	Qiime	Mean >25 phred quality. Within 3 sd. from mean length. Truncated to 650bp
Denoising	2.2.1.1	Acacia	Quince model correction
Quality control	2.2.1.2	Qiime	Length filtering: 350bp<X<600bp. Removal of homopolymers >6 Filtering of mismatched primers
Contamination Filtering	2.2.1.2	Bespoke script	Remove host (~22k) and parasite (~6.9k) contamination sequences
Subsampling (normalisation)	2.2.2	Qiime	Randomly subsample to minimum sample sequence count while maintaining 3 replicates (2,811)
OTU generation	2.2.2	UCLUST	OTUs were generated at 0.97, 0.94 and 0.88
Taxonomic Annotation	2.2.2	BLAST, Greengenes database	Blast annotation at <0.001 E value
Phylogenetic Alignment	2.2.2	FastTree	For downstream use in diversity analysis
Alpha Diversity Analysis	2.2.3	Qiime	Shannon, Chao1, Observed Species
Beta Diversity Analysis	2.2.3	Qiime/Unifrac	Jackknifed UPGMA analysis
Ordinance Analysis	2.2.3	R	Non-parametric Multidimensional Scaling & Principal Co-ordinate analysis
Network Analysis	2.2.4	Qiime, Cytoscape	Limited to highly abundant OTUs (>7% in ≥1 sample)

3.3 Results

3.3.1 The Basal Earthworm Microbiome

The observed taxonomic profiles and community structure represent the combination of transient soil- and inherently host-associated microbiota i.e. the known nephridial symbiont, *Verminephrobacter*. All earthworm samples included total gut contents (ingested soil) at time of harvesting, therefore any variation when performing comparisons with soil relates to direct influence of the host and represents the true microbial population present at the time of sampling.

The microbial composition (at the phylum level) of all *L. rubellus* analysed in this study, including on and off site controls together with the 5 sites originating from the arsenic mine site, were analysed and compared with the soil microbial composition (Figure 3.2). For the earthworms, Proteobacteria was the most abundant phylum in the majority of individuals (28/32, 52.3% total average). The next most abundant phyla were Actinobacteria (28.0%), Bacterioidetes (5.9%), and Acidobacteria (3.2%).

In earthworms Alphaproteobacteria was the predominant class in most samples, primarily comprising Rhizobiales (57%) and Rhodospirillales (29%) which likely originated from soil and was subsequently selected for by the anoxic gut environment (Depkat-Jakob et al. 2013).

Betaproteobacteria abundance was largely attributable to a single OTU of the known symbiont genus; *Verminephrobacter*, which comprised up to 93% of this microbial class in some individual earthworms. The presence of this taxon was highly sensitive to high arsenic contamination, resulting in near or total absence in all individuals from sites 1, 2, and 6, and 3/5 individuals from site 3 (high arsenic sites). *Verminephrobacter* presence in both control sites and site 5 individuals were responsible for ~77% of Betaproteobacteria and ~22% total microbiota represented.

The remaining earthworm Betaproteobacteria was largely soil-derived with 17 of 18 Betaproteobacteria genera being identified in both earthworm and soil communities. A proportion (16%) remains unclassifiable beyond Comamonadaceae (Family; 7%), (of which *Verminephrobacter* is member), Burkholderiales (Order; 6%) or Betaproteobacteria (Class; 3%). Unclassified Comamonadaceae displayed significantly increased presence in the host compared with soil, as was also observed in the

identified symbiont, and may indicate the presence of a *Verminephrobacter*-like species sufficiently distinct from known sequences as to form a distinct OTU.

Deltaproteobacteria comprised 2.8% average relative abundance of the earthworm community compared to 4.2% presence in soil. Gammaproteobacteria was present in approximately equal abundance between earthworm and soil communities (6.5% and 7.3% respectively), however at the class-level an increased Enterobacteriales and reduced presence of Chromatiales was observed in the earthworm community (excluding the off-site control) when compared to that recorded in the soils.

The presence of Actinobacteria (28.0%) was consistent amongst all earthworm individuals, displaying an increased abundance compared to soil communities (8.7%). The relative abundance of major contributing classes was raised in host samples versus soils; Actinomycetales (13.6% vs. 4.2%), Acidimicrobiales (5.9% vs. 1.7%) Solirubrobacterales (5.5% vs. 1.16%). Low levels of the phyla Bacterioidetes (5.9%) and Acidobacteria (3.2%) were present in host earthworm communities. This demonstrates a major decrease of soil Acidobacteria (34.6%) where it is the second most abundant phylum. Chloroflexi appeared at a higher rate in the microbiota of individuals from low contaminant sites (1.9% Off- and On-site controls compared to 0.8% contaminant sites), although this did not correspond with the soil communities, where Chloroflexi was identified in both high and low arsenic-enriched soils (Total: 1.6%).

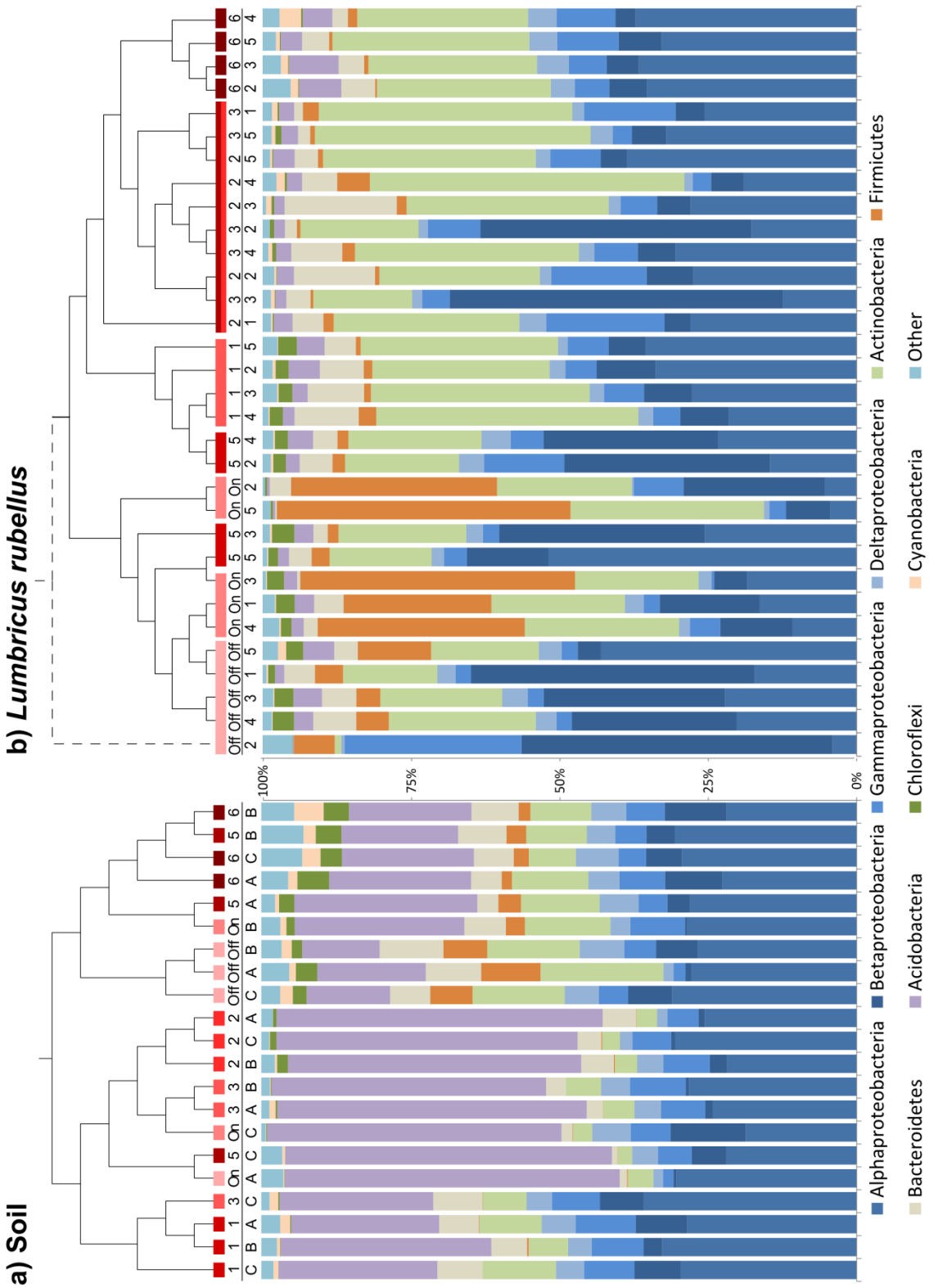


Figure 3.2 - **Phylum-level diversity chart for Soil and *L. rubellus* samples arranged by UPGMA phylogenetic sample similarity. Vertical columns indicate relative proportion of microbial phyla per sample. Proteobacteria has been displayed at class level as the largest Phyla. Columns labelled: Site/Replicate and coloured according to arsenic contaminant level by indicative boxes [High arsenic: dark] -> [Low arsenic: Pale]. Phylogenetic analysis indicates individuals sourced from the same site cluster closely by microbiome profile. More detailed taxonomic analysis in main text body.**

3.3.2 Host versus Habitat

In total 26,618 OTUs were generated at 97% homology linkage (10,895 excluding singletons) after normalisation (expected with this technique due to high variability in the soil environment (Griffiths et al. 2011)). Figure 3.3 shows OTU generation and diversity measures at 97%, 94% and 88%.

		97%			94%			88%		
	Total input sequences	OTUs (singletons)	Shannon (sd)	Chao (sd)	OTUs (singletons)	Shannon (sd)	Chao (sd)	OTUs (singletons)	Shannon (sd)	Chao (sd)
Soil	59,031	17,310 (11,991)	9.26 (±0.73)	4,545.36 (±1,443.0)	10,530 (6,455)	8.41 (±0.83)	2,789.99 (±982.38)	1,938 (702)	6.19 (±0.86)	675.82 (±231.4)
L. rubellus	89,952	12,201 (6,013)	7.66 (±1.34)	1,994.91 (±511.36)	5,861 (2,365)	6.93 (±1.25)	1,131.86 (±354.48)	1,194 (346)	5.32 (±0.86)	332.97 (±112.88)

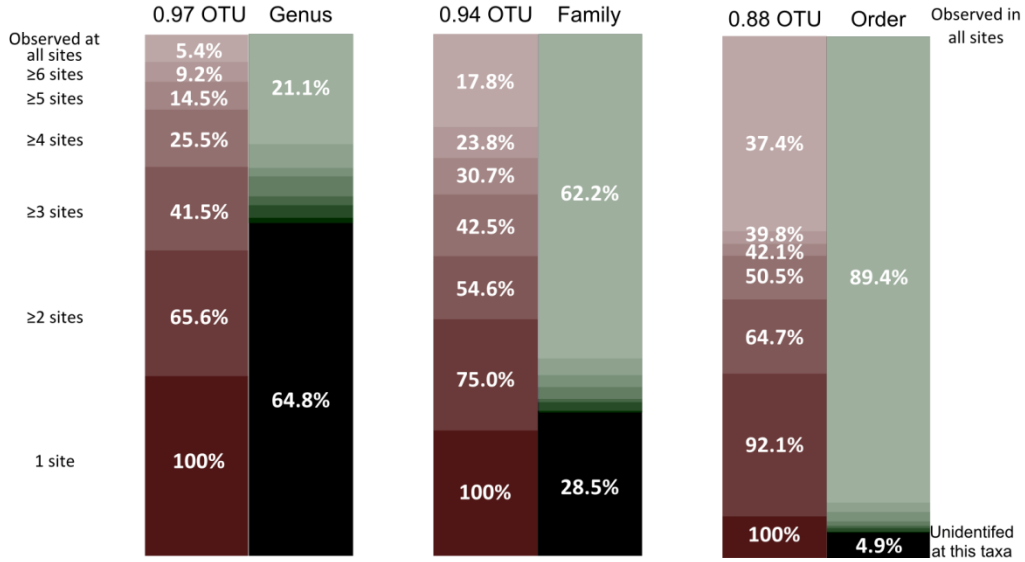


Figure 3.3 - **Diversity, richness and co-occurrence at various levels for *L. rubellus* and soil microbiome. The upper table shows estimates for diversity (Shannon, number of OTUs) and richness (Chao1) defined at 0.97, 0.94 and 0.88 sequence homology. All Soil vs Worm statistics are significantly different (P<0.5 t-test). Lower bars plot values representing the proportion of sequences that contribute to each taxonomic level or OTUs, occurring in at least one individual from each site. 5.4% of OTUs are found in host microbiota from every site, predominantly the *Serratia* species (discussed in text). 21.1% of genera are found at each site and 89.4% of orders, although un-assignable sequences may obscure a higher proportion.**

Principal Co-ordinate Analysis of Unifrac (Lozupone & Knight 2005) distances showed bacterial communities to differ between soil and host-resident microbiota (Figure 3.4). The largest differences were phylum level shifts where relative abundance of Acidobacteria reduced, and Actinobacteria increased from soil to *L. rubellus*, however, significant Family level abundance shifts were observed in the earthworm community (Figure 3.4b). Taxa are ordered by magnitude of difference between soil and host and indicate that large shifts can be attributed to family level changes.

Diversity and richness is summarised in Figure 3.5 and detailed in Figure 3.6. A general reduction in Shannon diversity was observed in host communities in comparison to the surrounding soil although not significant in all individuals (*t-test*, $P<0.05$, Figure 3.6). Chao1 richness was significantly lowered in all but one site (t-test, $P<0.05$) and Observed Species was significantly reduced in 5 of the 7 sites. To assess the *L. rubellus* microbiomic differences from non-contaminated control sites, separate analysis of these samples was performed. Sample pooling generated 4 data points with high sequence depth (OnSiteControl-Worm, OnSiteControl-Soil, OffSiteControl-Worm, OffSiteControl-Soil. Subsampled to 20,626 sequence reads per site). 16,725 OTUs were generated at the 97% homology. Diversity and richness estimates at this deeper level of sequencing maintained the same relationships as with the main dataset (Figure 3.8) but also highlights that a large amount of diversity is yet to be captured.

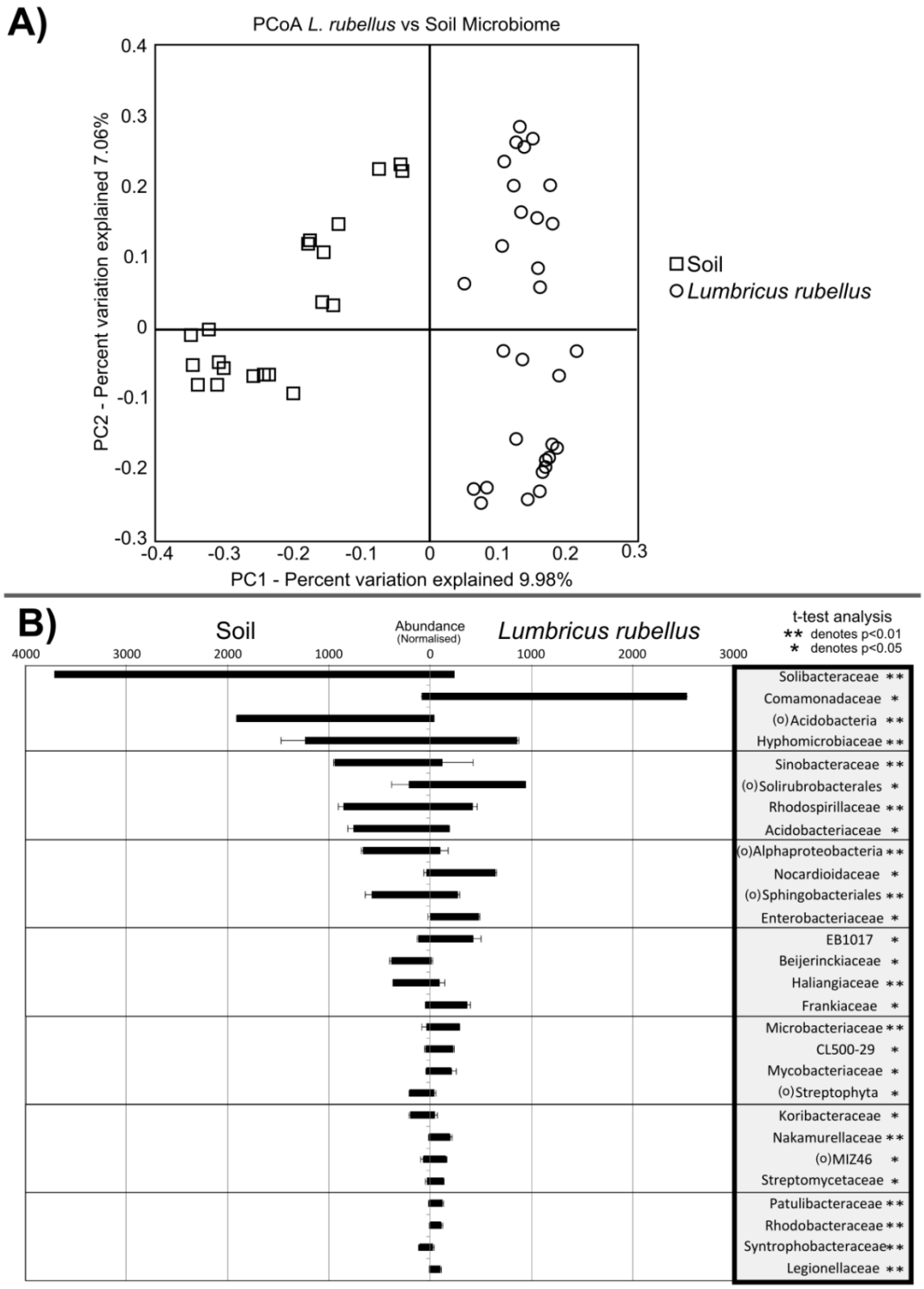


Figure 3.4 - **Contrasting the *L. rubellus* and soil microbiomes. (A) PCoA of unifrac distances with distinct separation on the primary axis between *L. rubellus* (circles) and soils (squares). (B) Bacterial families with >3.5% host or soil reads and significant (t-test, P<0.05) difference between host and soil. If Family level annotation was not possible Order was denoted by (o). Additive presence for all sites ordered by magnitude and plotted with standard deviation error bars.**

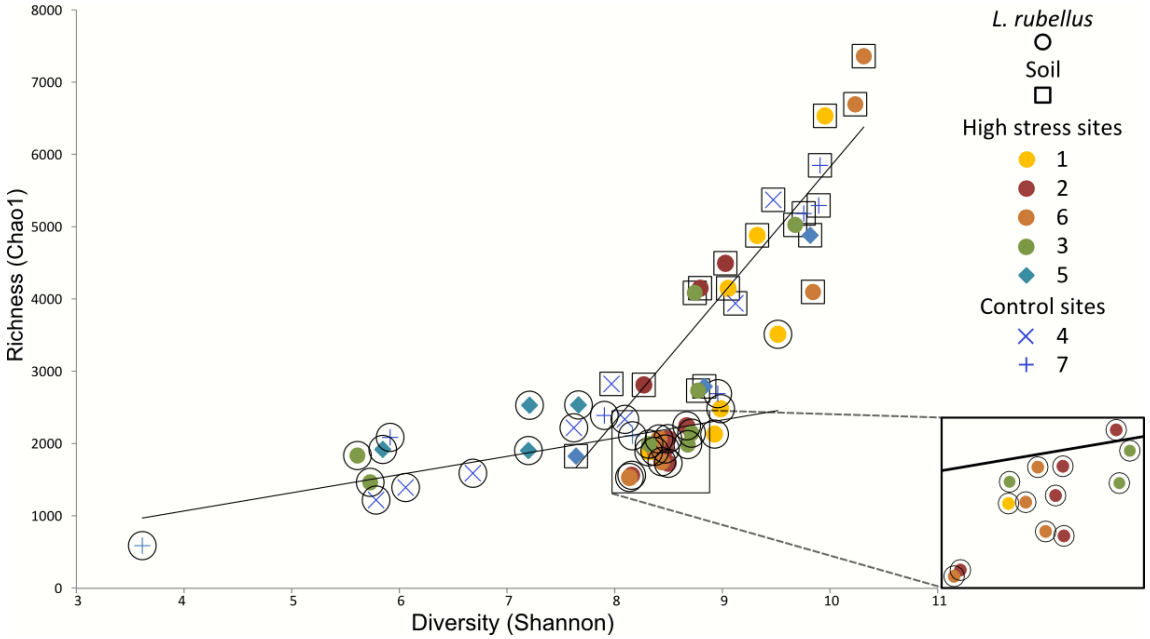


Figure 3.5 - **The effect of anthropogenic stress on community structure. (A) Overview of Diversity and Richness (Shannon and Chao1 respectively) for all soil (Squares) and *Lumbricus rubellus* (circles) microbiomes as coloured by site origin. Lower right box displays magnified area for clarity.**

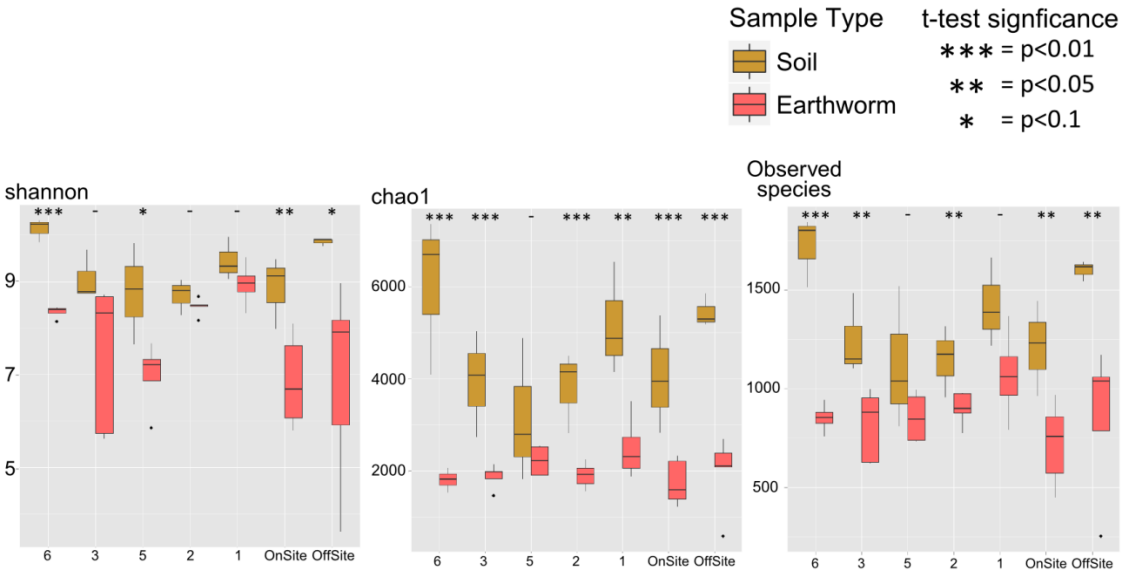


Figure 3.6 - **Diversity and richness estimates of *L. rubellus* and soil. Shannon, Chao1 and Observed species metrics of each site, showing reduction in the earthworm host in all cases. T-test significance indicated by (*).**

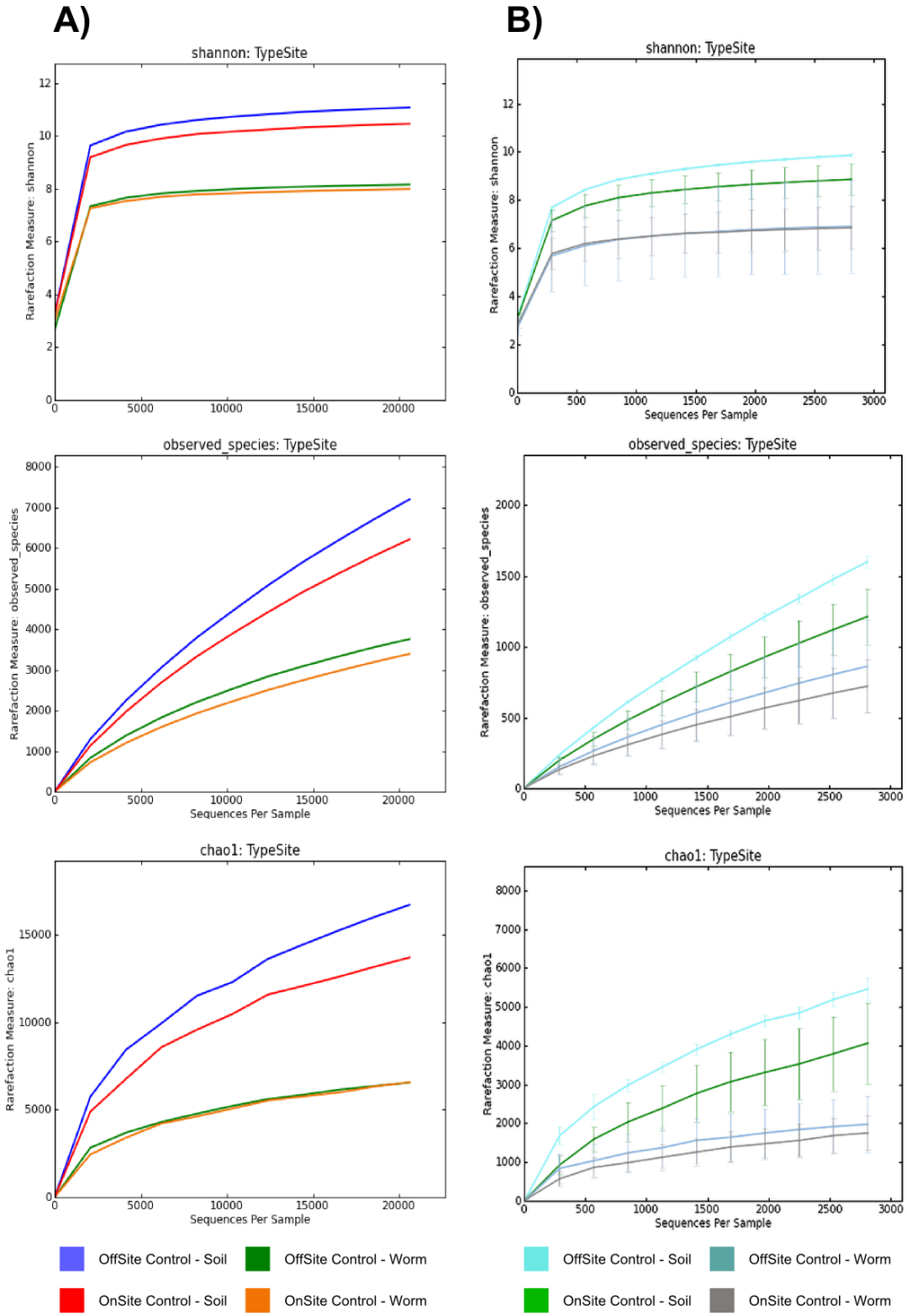


Figure 3.7 - **Diversity and richness estimates of pooled control *L. rubellus* and soil microbiomes. Control samples were pooled by site to establish whether further depth would show a plateau in diversity and richness estimates (A). This was not observed and trends seen in the subsampling approach (B) were consistent, indicating analysis was acceptable at the achieved sequencing level, however it is clear that a substantially higher amount of diversity is yet to be captured.**

3.3.3 Core Community

A consistent community structure was observed at the phylum level, as described above. 9,122 OTUs (at 97% homology) were found solely in the earthworm host microbiome but were absent from the soil. Due to the large variation in site conditions, a significant amount of diversity was observed across the dataset.

Earthworms shared 21% of genera between individuals at all sites (Figure 3.3). These were predominantly genera from Proteobacteria (61%) and Actinobacteria (28%). Greater conservation is likely, however 64.8% could not be accurately identified at this taxonomic level. Earthworms from both contaminated and control soils shared 13 genera which could be annotated from the reference database, and were not observed in soils. Seven 'core' OTUs were detected at all sites in at least 1 individual, and these OTUs contributed to 5.4% of all earthworm-derived reads (Figure 3.3). Of these core OTUs, 6 were identified as Actinobacteria (Class) representing 28% of the abundance, predominantly Nocardioides and Patulibacteraceae. A single OTU representing the Gammaproteobacteria genus *Serratia*, a genus which contains a known symbiont in aphids (Sabri et al. 2011), represented 72% of the core OTUs abundance and was found at distinct abundance at all sites excluding the on-site control (1.4% of total host-associated reads) although not every individual earthworm profile.

3.3.4 The effect of anthropogenic contamination on the microbial community

There was an implied, but non-significant trend observed in host community diversity between *L. rubellus* from control and contaminated sites (Figure 3.8). No significant trend was observed in correlation to arsenic availability or pH in either soil or earthworm microbiota with tested diversity and richness estimates (Shannon, Chao1, Observed OTUs; Figure 3.6). Low resolution through subsampling normalisation may obscure minor trends.

Non-parametric Multidimensional Scaling (NMDS) analysis of unifrac distance profiles (Lozupone & Knight 2005) of all individual worm microbiomes demonstrated a consistent microbial population being present in earthworms from the same site (Figure 3.8) and also highlights the major environmental variables correlating with the host-microbiome, primarily the strong correlation with pH in the control sites. In the presence of the other measured environmental stressors, pH becomes less significant and the arsenic-iron complex is observed as the dominant determinant of

microbiome composition. Cadmium appears to contribute strongly to the observed spatial patterning although stochastic presence/absence (5 sites <0.7 mg kg^{-1} Cd | 2 sites >7mg kg^{-1} Cd) may over-represent the impact.

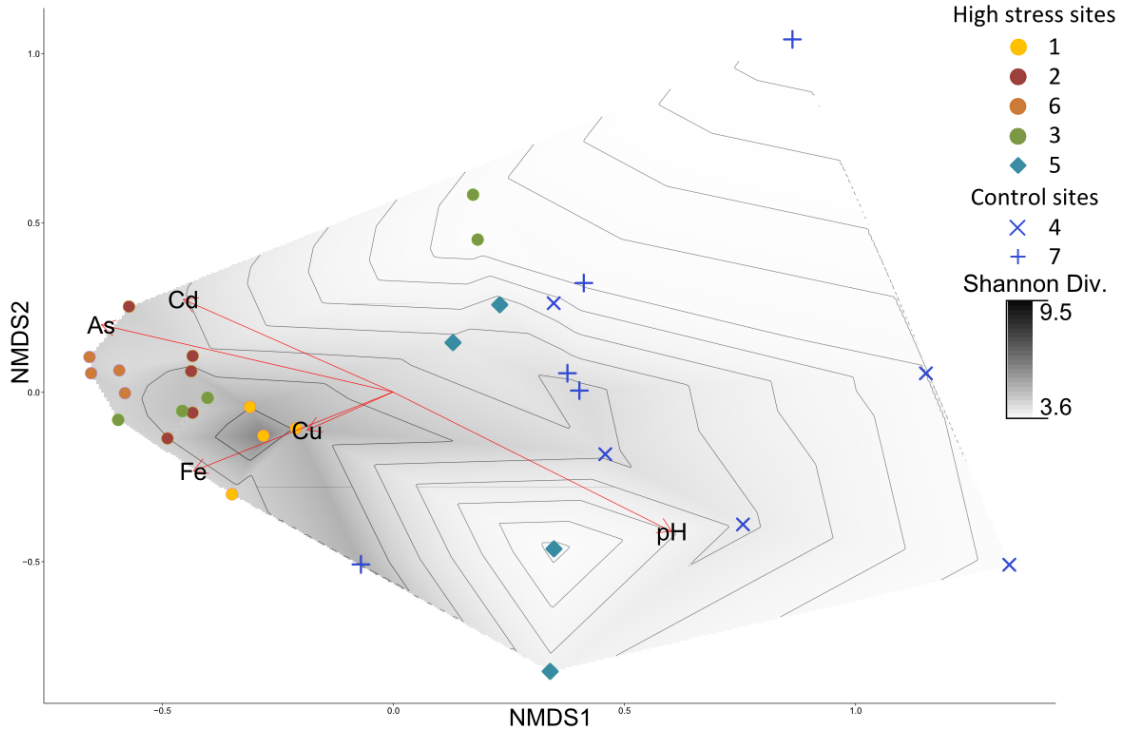


Figure 3.8 - **Non-parametric Multi-Dimensional Scaling (NMDS) plot representing divergence of *L. rubellus* microbiota profile and site similarity in conjunction with environmental factors. Site-specific grouping is observed, as is the effect of increasing stress on the microbiome community structure. pH is shown as the major contributor to community structure variation in individuals from control soils replicating known soil effects. Arsenic abundance appears to cause a combinatorial effect with iron.**

OTUs which drive the observed variance are identified in Figure 3.9. Network generation based upon the 47 most abundant earthworm-identified OTUs (>7% abundance) separate *L. rubellus* individuals into control and contaminated groups, with Site 5 spanning the 2 clusters (ANOVA $P<0.05$ = association to contamination factor, $P>0.05$ = equal association between sites, (BH-FDR correction)). Site 5 samples were omitted from association calculations due to individuals from this site being outliers but still included in network plotting. Of the abundant OTUs, 11 of 48 associate with the contaminated sites whereas 8 associate only with control sites and are largely absent from contaminated site locations. 29 OTUs were not significantly associated with either cluster implying co-occurrence in both control and contaminated site samples.

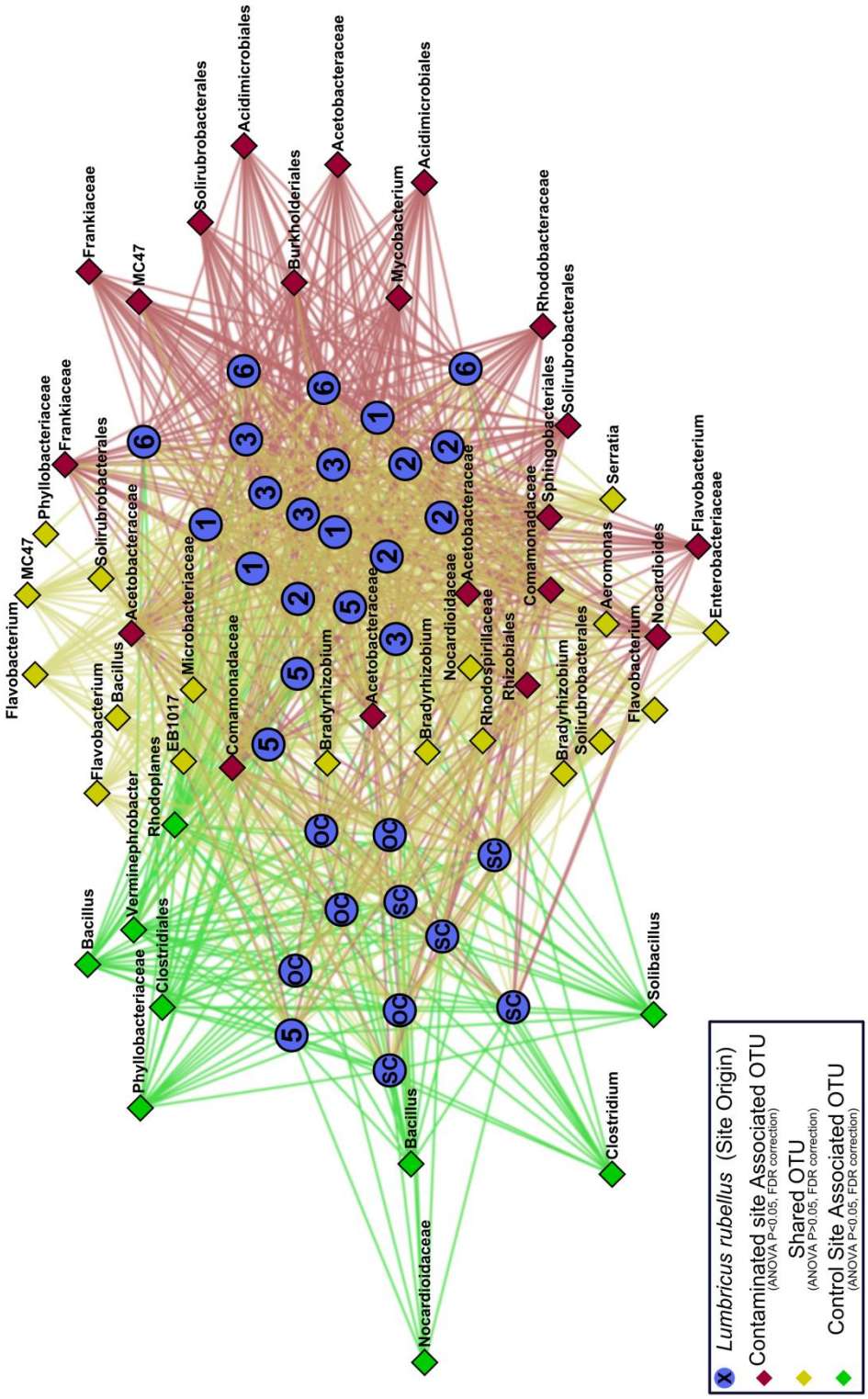


Figure 3.9 - **Network analysis of earthworm samples with associated abundant OTUs. Significantly present OTUs (>7% abundance, diamonds) associated with *L. rubellus*, (blue circles). Coloured by association to site origin conditions when ANOVA testing associates OTU with condition (P<0.05 = association, BH-FDR correction). All samples were incorporated in generation of network, however Site 5 outlier individuals were omitted from association calculations.**

3.4 Discussion

The results here describe the earthworm microbiome as distinct from the surrounding soil microbial community. The *L. rubellus* microbiome is dominated by Proteobacteria (~50%) and Actinobacteria (~30%). Bacteroidetes (~6%), Acidobacteria (~3%), Firmicutes, Chloroflexi and Cyanobacteria also appear regularly at lower abundance levels. Approximately 1/3 of Genera/OTUs (29.4% and 34.3% respectively) appear as earthworm-specific (not observed in the soil profiles), but only 7 OTUs are repeatedly observed in individuals sourced from across the seven sites. Sequencing depth is a limiting factor; however, these results support the concept that the community shift occurs in response to increases in the abundance of quiescent soil species via stimulatory effects in the gut environment, coupled with the environmental filtering of certain soil- and plant-associated species either by inter-specific competition or by unfavourable conditions (Figure 3.10). These observations contrast with earlier literature describing a high degree of similarity in the diversity of microbial communities within the earthworm gut and surrounding bulk soil (Egert et al. 2004), but concur with a later study describing the same major taxonomic groups at different proportions (Nechitaylo et al. 2010).

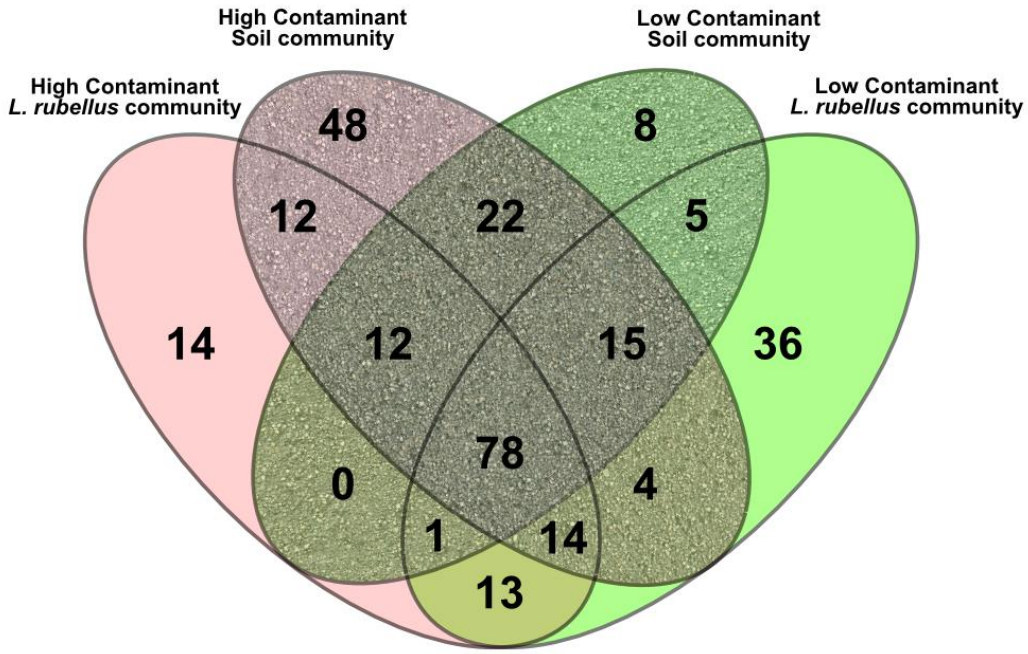


Figure 3.10 - **Venn diagram summarising shared OTUs between soil and earthworm samples at High (Sites 1, 2, 3 and 6) and Low (Off and On site controls) contaminant levels. A high number of OTUs were observed in all sample types, correlating with the soil-derived microbiome hypothesis, however a small number of *L. rubellus* OTUs were observed, implying presence of host-associated species. Only non-singleton sequences included.**

The earthworm-associated microbiome displays a significantly reduced level of diversity and richness in comparison to the surrounding soil, an observation in agreement with Gómez-Brandón *et al*. (2011). This reduction is likely due to both the prominence of the *Verminephrobacter* symbiont, and proliferation of minor soil species in the favourable conditions of the host gut environment (neutral pH, mucosal saccharides, organic acids (Wust et al. 2009)) in conjunction with decreasing numbers of transient species. A diversity closer to soil was observed in host earthworms inhabiting contaminated soils where the symbiont was eliminated. This suggests that egested material is more similar to soil diversity despite taxonomic shifts and that the reduced measures observed are due in part to host-bound species.

Significant reductions are observed in the oligotrophic and acidophilic Acidobacteria families (including Solibacteraceae and Koribacteraceae) when passing from soil to host, which likely reflects both the impact of circumneutral gut pH and increases in carbon sources derived from gut secretions (Drake & Horn 2007). Conversely, increases in Actinobacterial families typically described in soil communities suggest a stimulating effect of the host environment and may contribute to the acknowledged activity of earthworm species in nutrient cycling. For example, the increased earthworm abundance of Streptomycetaceae can contribute to cellulose degradation through enzymatic activity (Thakuria et al. 2010), Mycobacteriaceae utilise soil humic acids and participate in nitrogen cycling (Ventura et al. 2007). Additionally, the total absence (at this sequencing depth) of Enterobacteriacea from soils, and the significant abundance in host communities, strongly suggests a microbial community curated by earthworms and indicates the potential presence of functionally beneficial symbiotic communities.

Anthropogenic soil contamination, particularly in the form of arsenic and iron, caused significant shifts in the composition of the earthworm microbiome. However several species of Actinobacteria and one species of Gammaprotobacteria were identified as being present in individuals from all sites (albeit not consistently in all individuals at this sequencing depth). The prominence of *Serratia* (Gammaproteobacteria) has not been previously noted in earthworms, although it may be a constituent of the Enterobacteriaceae community previously described (Wüst et al. 2011). In free living communities, *Serratia* is known to digest a wide

range of carbon sources through production of various hydrolases (Farmer et al. 1985), yet *Serratia symbiotica* is an intracellular symbiotic species in aphids which has lost many of these attributes during chronic host-association and vertical transmission (Sabri et al. 2011). If the *Serratia* here observed is indeed a symbiotic species then a chronic, vertically transmitted, association may account for such divergence. Further analysis will be needed to establish the nature of the *Serratia*-earthworm association and to determine the functional role of this highly prevalent species within its host.

The observed ubiquity of the symbiotic *Verminephrobacter* species in *L. rubellus* inhabiting non-contaminated control soils was predicted (Davidson et al. 2013); however, it appears highly sensitive to environmental arsenic contamination demonstrating total absence in high contaminant sites. As a long-known symbiont of *L. rubellus* nephridia (Pandazis 1931), the absence of *Verminephrobacter* has been shown to reduce earthworm fitness in nutritionally impoverished environments (Lund et al. 2010). The symbiont has been shown to be actively recruited by the earthworm whilst in the cocoon (Davidson & Stahl 2008) but the abundant presence of *L. rubellus* at the contaminated sites (Langdon *et al.*, 2001) suggests that absence of the symbiont does not cause overt detriment to the host population and revives the question of its function.

The effect of elevated arsenic and iron on the host microbiota produces a conserved earthworm-associated community structure which is distinct from that extant in the surrounding soil. Furthermore, earthworm microbiome profiles were more similar between sites than between individual earthworms and their site-specific soil. The combinatorial effect of iron with arsenic may relate to Fe-As complexes affecting arsenic speciation and promoting the oxidation of arsenic to As(V) species (Bednar et al. 2005). Leaching of arsenic from soils by the action of microbiota is increased in the presence of a carbon source (Turpeinen et al. 1999) and may contribute to the effect of earthworm species on arsenic mobility (Sizmur et al. 2011). Microbiome profiles originating from Site 5 earthworms consistently appeared unaffected by the high arsenic levels according to NMDS and Principal Co-ordinate Analysis. This correlates with the high copper concentration recorded in the soil at this site, which may reduce arsenic availability through formation of copper-arsenide compounds (Cao et al. 2003).

There were 18 abundant OTUs identified with a statistically significant increased abundance in *L. rubellus* from arsenic contaminated sites. These include unknown species of Burkholderiales, Acidimicrobiales, several Acetobacteria OTUs and the Actinomycetales *Frankia* and *Mycobactaria*. Additionally, two Comamonadaceae OTUs (closely related to the sensitive *Verminephrobacter* symbiont) were associated with the contaminated microbiomes and may represent a divergent, tolerant lineage. In the terrestrial isopod *Porcellio scaber*, environmental mercury contamination causes a shift in gut community and an increased abundance of Hg-resistance bacterial genes, potentially contributing to the isopod's resistant phenotype (Lapanje et al. 2010). Similarly, species identified in this study could be of interest in future investigations into the basis of local adaptations of earthworm field populations to chronic arsenic exposure, and also in understanding the increased mobility of soil arsenic in the presence of earthworms (Sizmur et al. 2011).

Twenty highly abundant OTUs were found not to significantly associate with either contaminated or control site earthworms. These core OTUs consisted of several flavobacterium species, including Actinobacteria, Rhizobiales and a *Serratia* species, and form the most likely candidates for defining a core functional community. However, distinguishing active species from those inactive in transit are beyond the possibilities of this study and requires further study.

There were 9 contaminant-sensitive OTUs identified, including *Bacillus*, *Clostridia*, *Rhizobiales,* and the *Verminephrobacter* symbiont. All of these were significantly associated with unpolluted reference sites. Given their high abundance in the *L. rubellus* microbiome from control sites, their absence could result in major changes in the functional output of the microbial population and may potentially disrupt fundamental host processes (e.g. fecundity via *Verminephrobacter* symbiont). Additionally, in light of the essential environmental roles that *L. rubellus* performs (Edwards 2004; Bernard et al. 2012; Nahmani et al. 2007), alteration of the stable microbial community structure could have large impacts upon global processes such as greenhouse gas production (Lubbers et al. 2013; Ihssen et al. 2003).

Given the high microbial community variability at the genus/species level, few species form major constituents or contribute towards a 'core community' as observed in some other invertebrates, for instance termites (Warnecke et al. 2007). This means that any broad functional roles arising from the microbiome (e.g.

denitrification (Drake et al. 2006; Ihssen et al. 2003)) would have to be enacted by communities acting in concert, rather than by a single dominant species. However, it is reasonable to expect that disparate ingested communities can differentially proliferate to a functionally convergent, active, microbial population to exploit the stable conditions maintained by the host environment. The host-induced propagation of Enterobacteriales (facultative aerobes) validates one proposed origin of nitrogenous gasses (Wüst et al. 2011) and supports the notion that some roles are derived from the action of a wider microbial community rather than an individual species.

3.5 Conclusion

Earthworms are globally distributed and perform essential roles in organic matter fragmentation, carbon and nitrogen cycle regulation and the modulation of soil microbial composition (Lavelle et al. 2006; Li et al. 2002; Brown et al. 2000). The present study posits that the earthworm species *L. rubellus* accommodates, *in situ*, a significantly divergent microbiome community compared with that found in the surrounding bulk soil that it inhabits including largely increased Actinobacteria and reduced Acidobacteria. Furthermore, the identification of significantly associated taxa including Flavobacterium and Enterobacteriales provide targets to further understand earthworm-associated communities. Understanding the interplay between transient/resident microbial communities and their ecosystem-engineering geophagic hosts is key to explaining the environmental effects earthworms have, as well as improving our knowledge of the benefits of mutualism for soil invertebrates. The demonstrated impact of anthropogenic contaminants on the microbial community including eradication of the obligate *Verminephrobacter* symbiont of an ecologically-important species such as *L. rubellus* raises concerns for both host health and causal effects on the global environment.

4

Examining the microbiome of an extremophilic earthworm living on volcanic geothermal soils

4.1 Chapter Introduction

The propensity for bacterial species and communities to colonise extreme environments is one of the microbial domain's most noted characteristics, as are the novel pathways often employed to counteract environmental conditions often fatal to eukaryotic species. Volcanic geothermal vents represent some of the most extreme non-aquatic environments globally and novel bacteria are frequently identified (Norris et al. 2002; Johnson et al. 2003; Imperio et al. 2008). The Azorean island São Miguel demonstrates substantial volcanic degassing in the active calderas which, at the site investigated in this study, exhibit CO_2 levels approaching 50%, with further notable composition including hydrogen sulphide (H_2S), sulphur dioxide (SO_2), hydrogen chloride (HCl), and lesser amounts of hydrogen fluoride (HF) and radon (Rn) (Silva et al. 2007; Rinaldi et al. 2012). Soil temperatures in several of these key locations can reach upwards of 50ºC and the soil is permeated by high levels of heavy metals originating from the mantle.

Co-occurrence can often be observed between microbial species/communities and a eukaryotic host, and it is frequently revealed to include symbiotic and/or mutualistic functions e.g. collagen degradation in boneworms (Goffredi et al. 2007), digestion pathways in wood-feeding termites for cellulose and xylan (Warnecke et al. 2007), and immune system development of *Drosophila* (Teixeira et al. 2008). As such, bacterial populations capable of buffering environmental stress or providing access to novel food sources (Hansen & Moran 2014) would be highly beneficial to soil invertebrates such as earthworms, in the high-strain environment of a geothermal vent.

Little information is presently available on the soil bacterial communities of the São Miguel Azorean island, either in the volcanic degassing areas or sites of lower stress. The majority of studies to date have focused upon the marine hydrothermal vents encircling the islands (Sahm et al. 2013), or the biofilms of subterranean lava caves (Hathaway et al. 2014). However, the fumerole soil community of the geothermal sites on the adjacent Azorean island Terceira was studied (Alexandrino et al. 2011), but no taxonomic information was reported. Therefore the novel soil environment of São Miguel provides an ideal location for expanding the fundamental knowledge of bacterial communities associated with volcanically-active terrestrial calderas and especially those associated with the metazoan mesoflora that live within these extreme environments.

The megascolecid *Amynthas gracilis* is an extremophilic epigeic earthworm thought to be native to the eastern Palaearctic and known to increase microbial biomass of the soil (Burtelow et al. 1998). It is noted for its invasive nature due to inherent environmental plasticity, rapid growth and fecundity (García & Fragoso 2002), and can be found living in soils surrounding the extreme conditions of geothermal vents. It has been shown to respond physiologically to the extreme environment described in this chapter through thinning of the epidermal cell layer and differential transcriptional expression (Novo et al. *In Preparation*), although the populations that inhabit the area are not genetically different to those inhabiting a less stressed reference site, indicating plasticity rather than adaptation.

This chapter describes the transplant study which aimed to describe the microbial community associated with *A. gracillis* earthworms due to their ability to inhabit the extreme active volcanic site. This study investigated the bacterial communities found within the earthworm inhabiting both an Active Volcanic (AV) site and a Reference Field (RF) site, both in the context of the surrounding soil. This was achieved through a 30 day reciprocal *in situ* mesocosm cross-transplantation to observe which communities originated from the soil after transplantation of the host and which maintained their presence after 30 days in foreign soil. Identification of changes to the population both taxonomically, and the differential frequency of predicted gene pathways associated with these taxa, were key to understanding the impact of the tripartite stressors associated with the active volcanic site and the way in which the earthworm host assimilates into a novel microbial environment.

4.2 Methodological Approach

4.2.1 Site Description and Experimental Design

Two exposure sites on the largest Azorean island São Miguel were selected; one representative of the Active Volcanic (AV) habitat (Furnas, 37.773, -25.304) and a second Reference Field (RF) site which did not display volcanic characteristics (Macela, 37.764, -25.525). Due to the large number of individuals required for the reciprocal transplantation experiment, earthworms were harvested from source populations resident to nearby locations with similar characteristics: AV_{worm} (37.770, -25.304) and RF_{worm} (37.764, -25.534). Environmental characteristics and metal abundance levels were performed previously and are reproduced in Figure 4.1 with permission from the authors (Novo et al. *In Preparation*).

Adult *Amynthas gracilis* were visually identified upon capture prior to transplant with later confirmation via COI barcode sequencing at the completion of the exposure phase of the experiment (Chapter 2.1.2.4.3). Six perforated mesh bags were filled with 15 L soil excavated from their respective site (12 in total) and buried *in situ* along with 10 adult earthworms per bag. Replicate AV_{worm} were added to 3 bags per site and RF_{worm} to the remaining 3, as visualised in Figure 4.2. The mesocosm bags were opened after 30 days at which time triplicate soil samples were collected (Chapter 2.1.1.2) from each bag whilst harvesting the earthworms (Chapter 2 1.2.1). Additionally, 3 native individuals and a representative soil sample were taken from each site and immediately processed representing native reference samples ($NatAV_{worm}$ and $NatRF_{worm}$) unaffected by the mesocosm environment and showing minor variation from the site of origin.

Soil and earthworm DNA extractions were performed using the Qiagen method (Chapter 2.1.2.3.1). PCR of the amplicon region 357F-907R was performed as with unique barcodes included in the primer design for post-sequencing sample identification followed by submission of 56 samples for 454 FLX++ sequencing described in Chapter 2 (2.1.3.1).

A)	Mesocosm RF_{soil}			$NatRF_{soil}$			Mesocosm AV_{soil}			$NatAV_{soil}$		
	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min
CO_2 (vol.%) 50 cm	0.8	2.8	0.1	0.7	2.3	0.2	48.6	96.5	14.0	6.9	8.8	2.4
CO_2 flow (g m^{-2} d^{-1})	19.5	27.5	12.3	15.4	32.4	6.2	181.0	533.4	58.3	19.8	25.0	12.5
O_2 (vol.%) 50 cm	18.5	19.9	13.1	18.8	20.0	15.6	10.4	17.1	1.4	9.9	18.3	1.2
Temperature (ºC) 30 cm	17.0	17.6	16.2	17.0	18.9	15.9	48.6	96.5	14.0	18.3	19.3	17.8

B)	Mesocosm RF_{soil}		$NatRF_{soil}$		Mesocosm AV_{soil}		$NatAV_{soil}$	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
pH	5.84	0.32	5.92	0.13	5.58	0.08	5.73	0.08
Loss on ignition (Organic Matter) (g per 100g oven dried soil)	20.06	11.39	29.30	9.88	9.08	0.87	13.59	1.58
Soil moisture content (H2Og per 100g soil)	5.70	2.05	6.67	0.06	3.50	0.23	2.82	0.43
Water Holding Capacity (%)	103.14	22.10	115.93	27.28	80.78	3.82	97.02	4.65
Clay (< 2 micron)	1.33	0.58	1.50	0.58	3.33	0.58	1.67	1.15
Silt (2 - 63 micron)	60.00	6.08	58.25	7.27	64.33	3.51	38.67	19.50
Sand (63 micron - 2 mm)	41.67	6.81	43.50	6.95	35.67	3.51	62.00	19.52

C)	Transplant RF_{soil}		$NatRF_{soil}$		Transplant AV_{soil}		$NatAV_{soil}$	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Al	29856.81	4783.86	32172.76	4055.47	21695.94	800.09	12507.00	307.79
Ba	73.32	26.99	133.31	20.45	73.60	3.62	32.18	9.60
Ca	6605.19	6186.71	8807.07	1545.20	2746.45	123.65	2982.32	676.61
Cr	25.41	15.90	25.48	0.99	16.80	1.99	6.11	4.04
Cu	19.12	11.35	27.64	2.81	79.47	6.12	11.27	8.10
Fe	29492.22	5528.81	27864.15	2616.12	20899.31	401.29	13344.68	1540.64
Ga	192.60	55.49	174.17	22.50	63.78	4.01	57.47	2.32
K	1547.25	695.81	1578.70	234.47	2216.08	194.58	3173.71	352.62
Li	6.51	1.33	6.98	1.11	6.59	0.16	5.65	0.02
Mg	8111.04	8312.83	6481.89	1140.01	1619.22	100.50	1945.52	356.02
Mn	807.39	219.43	996.94	133.22	586.45	2.17	623.71	54.64
Na	1712.09	1603.80	1743.44	284.42	868.97	39.42	1438.18	135.42
Ni	35.27	39.17	24.87	5.79	5.44	1.33	4.46	0.06
Pb	14.90	6.31	10.91	1.69	145.05	14.43	17.63	13.71
Sr	64.79	67.27	85.43	10.39	29.79	1.11	22.40	3.36
Ti	3824.97	1191.80	3829.65	405.35	1220.06	103.84	1405.64	35.74
Zn	131.18	50.27	153.35	22.65	263.53	13.50	111.36	66.15

Figure 4.1 - Environmental characteristics of source site (Nat) and transplant mesocosm sites. Gas flow and temperature data provided by site volcanologists (A), with additional physiochemical characteristics (B) and soil metal abundance (C).

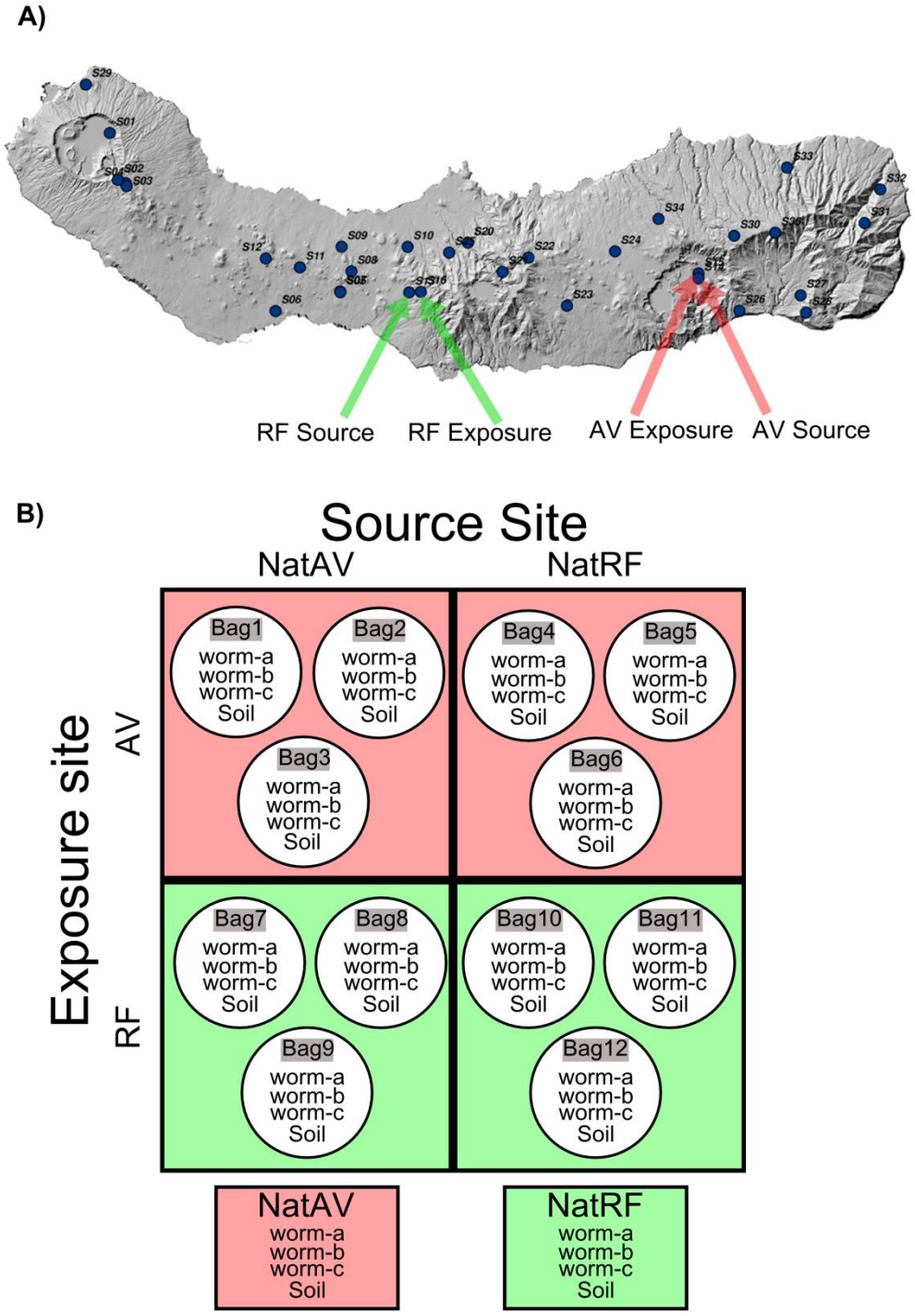


Figure 4.2 - Graphical representation of the AV/RF transplantation experiment. (A) The geographic position of the source of the experimental earthworms and the exposure transplant site. (B) Further explanation of the transplant cross, demonstrating the number of individuals per mesocosm and native (non-transplanted) samples.

4.2.2 Bioinformatic Methodology

The informatics processing steps performed are described in detail with source references in Chapter 2.2 and sample descriptions in Appendix 1b. Approximately 580,000 reads were produced from 454 FLX++ and after quality filtering 504,026 were utilised for analysis. Each amplicon sample was subsampled to 4,500 reads. A brief overview of the steps taken follows:

Stage	Chapter	Software	Description
Input Data	2.1.3.1	454 FLX++ Sequencing	579,847 sequence reads
Quality Control	2.2.1.2	UPARSE	Primer mismatches removed, <400bp discarded, Sequences trimmed to 450bp, Singletons removed, 504,026 sequences utilised
Denoising	2.2.1.1	N/A	Incorporated in UPARSE pipeline
Contamination Filtering	2.2.3	Bespoke script	Remove host & eukaryotic sequences
OTU generation	2.2.2	UPARSE	5,108 OTUs generated at 0.97
Subsampling (normalisation)	2.2.2	Qiime	Randomly subsample to minimum sample sequence count (4,500)
Taxonomic Annotation	2.2.2	BLAST, Greengenes database	Blast annotation at <0.001 E value
Phylogenetic Alignment	2.2.2	FastTree	For downstream use in diversity analysis
Alpha Diversity Analysis	2.2.3	Qiime	Shannon, Chao1, Observed Species
Beta Diversity Analysis	2.2.3	Qiime/Unifrac	Jackknifed UPGMA analysis
Ordinance Analysis	2.2.3	R	Non-parametric Multidimensional Scaling & Principal Co-ordinate analysis
Functional Inference	2.2.4	PICRUSt	Predict gene pathway abundance from amplicon data

4.3 Results

4.3.1 Microbiome Community Characteristics

In total 5,108 OTUs were generated at 97% homology (excluding singletons). Distribution between the soil microbial community and that observed within the earthworm host is described in Figure 4.3. Overall, earthworm samples displayed a higher range of OTUs than soils but diversity was higher in individual soil samples (Figure 4.4). This is likely as more earthworm samples were analysed than soils due to the focus of the study (42 versus 14 respectively) therefore more data is available for the earthworm microbiomes, incorporating more OTU species.

	Soil	Earthworm	Shared
Total OTUs	3,374	4,485	2,751
OTUs present in greater than half of samples	391	380	500
OTUs present in greater than half of samples but absent from alternative sample type	0	11	

Figure 4.3 - OTU distribution between soil and earthworm microbiomes. A high proportion of OTUs were shared between earthworms and soils, but a number of host-specific OTUs were observed.

Diversity and Richness was estimated using the Shannon and Chao1 metrics, supplemented with expected Observed Species calculation. Figure 4.4 shows the calculated values and the percentage difference between soil and earthworms from each site. Both soil and earthworm microbiomes had lower diversity and richness in the AV_{soil} exposure site (Figure 4.4A).

Native soils and earthworms demonstrated higher diversity and richness metrics than any samples under transplant conditions. The diversity of soil communities from native sites (non-transplanted) was not significantly different from the earthworm microbiota, but demonstrated higher richness. This may indicate a bias in experimental design where the transplantation mesocosms artificially reduced diversity.

There was lower difference between the metrics of individuals transplanted to AV_{soil} communities than the RF counterparts. However, only the transplanted RF_{worm}/AV_{soil} earthworms demonstrate a greater diversity and richness than their soil habitat. The inverse transplantation (AV_{worm}/RF_{soil}) had markedly lower metrics than the RF_{worm}/RF_{soil} earthworms also indicating a lack of full assimilation to the new soil.

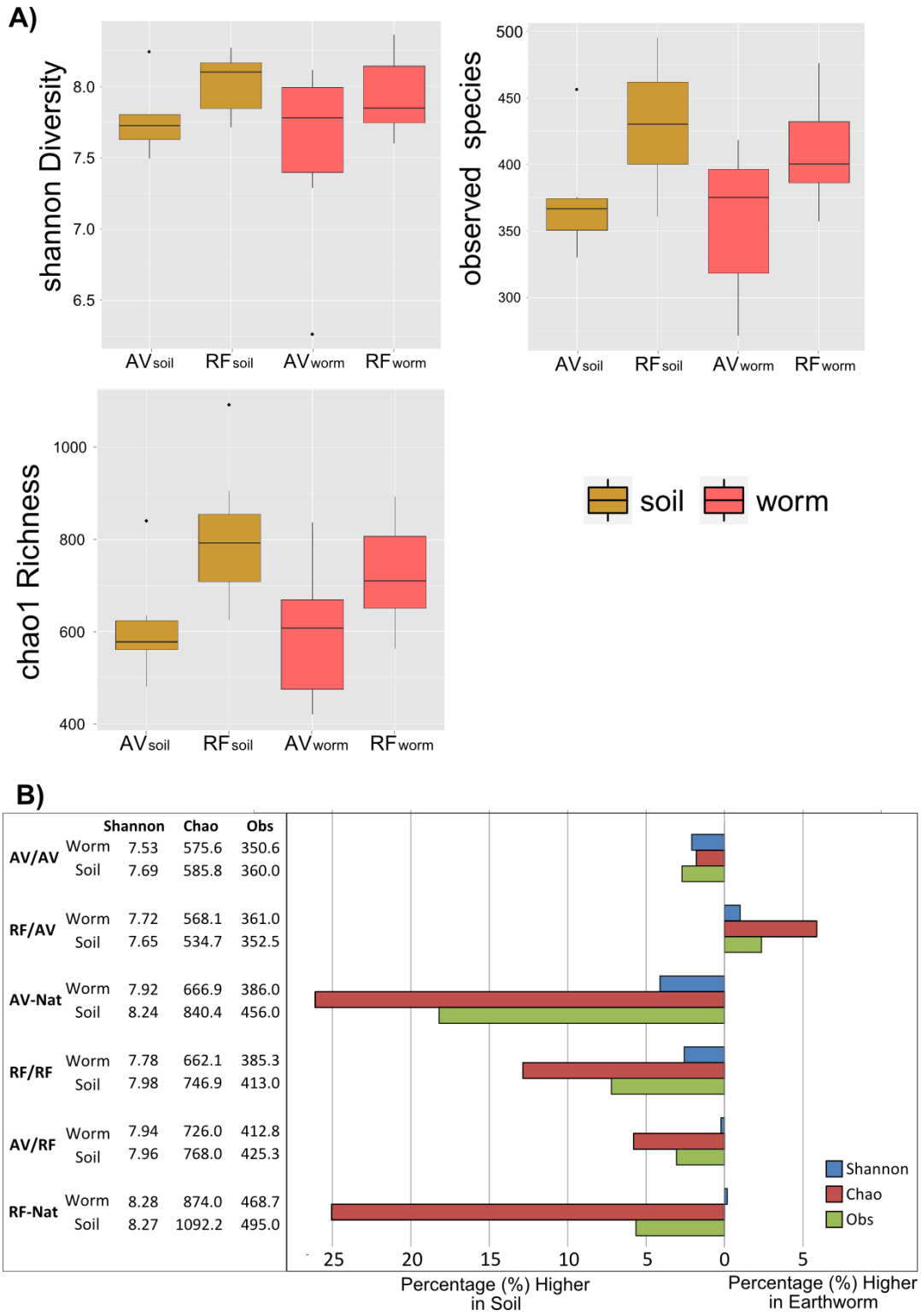


Figure 4.4 - Differences in bacterial diversity and richness between soil and earthworm communities. (A) All indices were higher in individuals transplanted to RF_{soil} which was consistent with transplant and native samples (displayed). (B) Soil-worm comparisons demonstrate higher soil diversity in all cases except $NatRF_{worm}$ in AV_{soil} mesocosms.

The soil and earthworm community was markedly similar at the Phylum level where 80% (±1%) was accounted for by Proteobacteria, Actinobacteria and Acidobacteria in both datasets (Figure 4.5). The foremost community change was the increased Actinobacteria abundance (24.8% earthworm versus 11.3% soil) and the reduced Acidobacteria presence (8.1% earthworm into 26.0% soil) which supports previous observations between earthworms and their surrounding soil (Chapter 3; Pass et al 2014). Also notable is the change in abundance of the Ternicute phyla which, although a minor proportion of the earthworm microbiome (average 0.69%), was almost absent from soil profiles (0.028%)

The microbial community of the earthworm was largely similar across all transplanted and native sites at the Phylum level (Figure 4.6), comprising predominantly Proteobacteria (48.3%, (Alphaproteobacteria: 24.4%)), Actinobacteria (24.8%), Acidobacteria (8.1%), Firmicutes (6.9%), Bacteroidetes (5.0%) and Chloroflexi (4.2%). The remaining 2.5% consisted of Tenericutes, Nitrospirae, Gemmatimonadetes, Cyanobacteria, Verrucomicrobia and Planctomycetes.

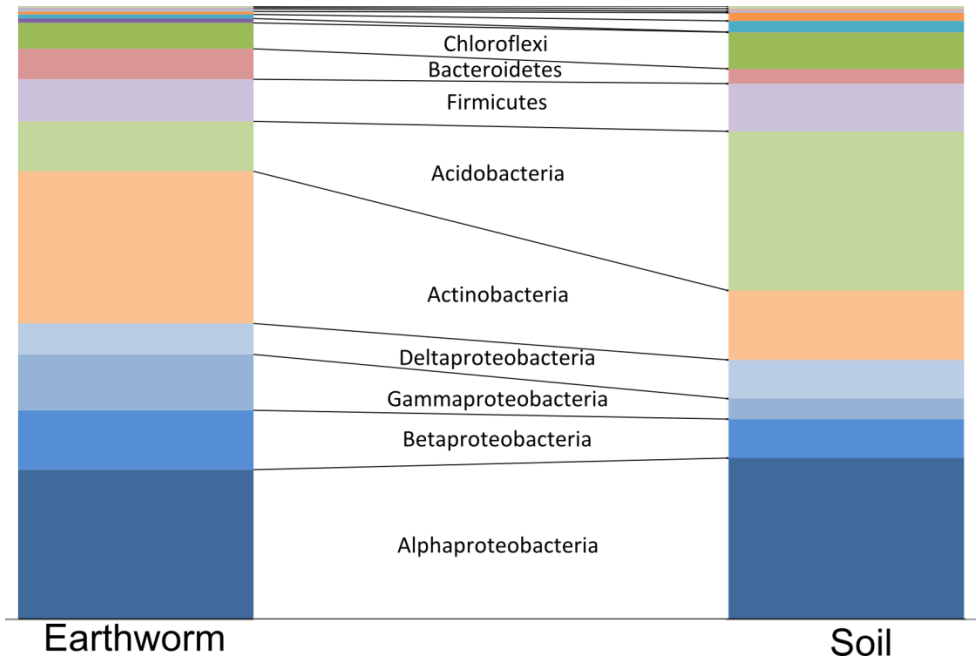


Figure 4.5 - Phylum level differences in earthworm and soil microbiomes. Significant reduction in the earthworm microbiome of Acidobacteria is approximately equal with increased Actinobacteria. Minor abundant phyla are omitted from labelling. Colouring is consistent with Figure 4.5.

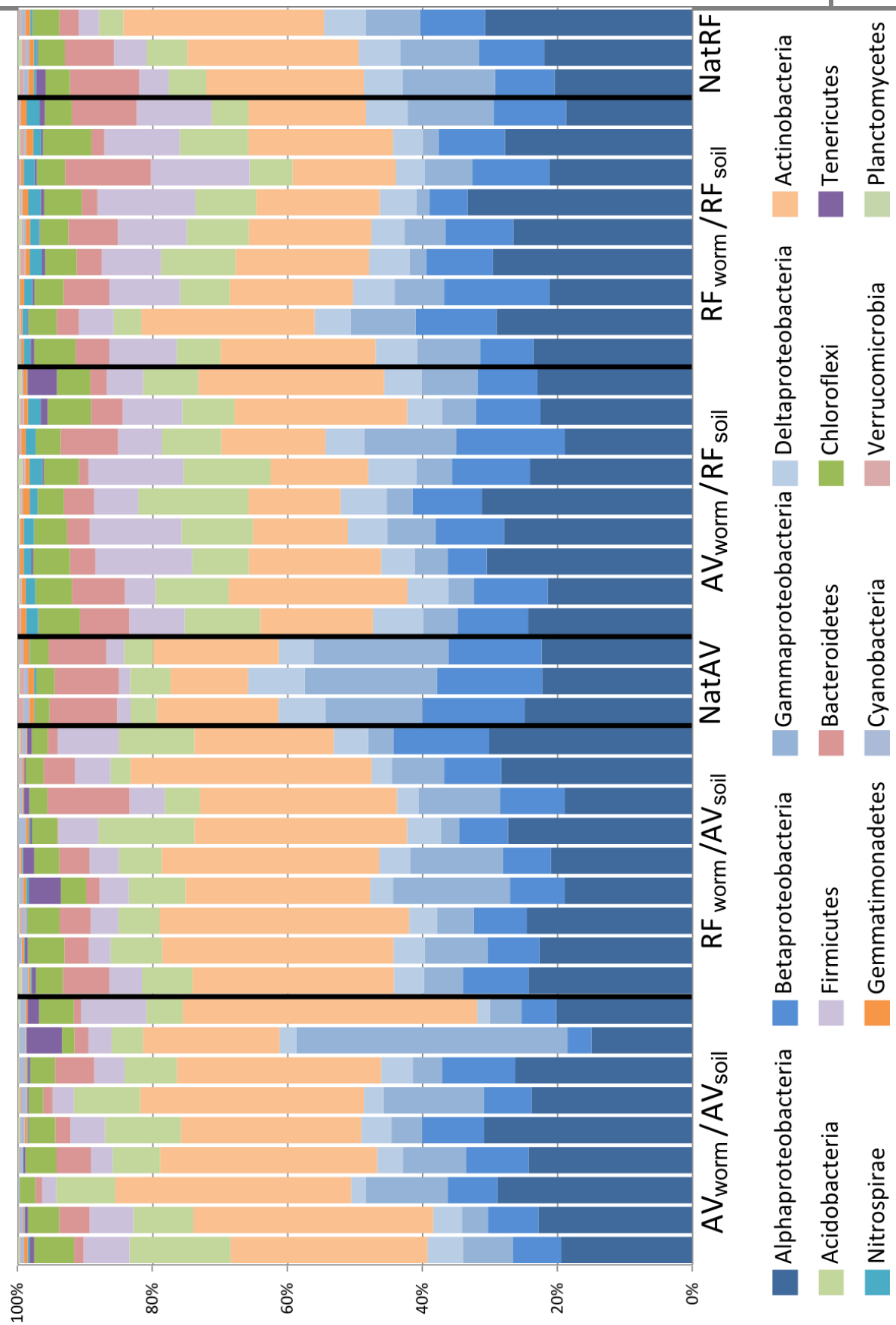


Figure 4.6 - Phylum level chart of the *A. gracilis* microbiome. Community structure is largely consistent at the phylum level across native and transplanted earthworms.

4.3.2 Association of microbial communities to environmental conditions

The earthworm microbial community profile was separated significantly between the transplantation sites, where transplanted individuals more closely represented the mesocosm site than their source habitat, as demonstrated through NMDS analysis (Figure 4.7). The primary axis displays the separation driven by exposure to the physical conditions of the transplantation site, whereas the secondary axis identifies (A) the separation of soil and the earthworms, and (B) the grouping of native earthworms from both sites ($NatAV_{worms}$ and $NatRF_{worms}$) which may indicate a bias imposed by the transplant conditions.

A number of OTUs associated significantly with particular transplant conditions (Figure 4.8). Although all taxa presented were significant at $P<0.05$ (g-test), only the first OTU (Exiguobacterales) was significant when False Discovery Rate correction (BH-FDR) was applied. In all but one instance of those charted, OTU abundance was directly related to the mesocosm exposure site (i.e. the species was acquired during the transplant) and was typically less abundant in the cross-transplanted individuals than those transplanted to the same-condition, perhaps indicating partial colonisation. Also of interest is the inverted abundance of distinct Aeromonadales species (OTUs 3, 62 & 23) which demonstrates a site-associated species that is genetically distinct.

At the conclusion of the transplant period, earthworm bacterial communities overwhelmingly resembled the soil of exposure sites regardless of their origin. Differences between the microbiomes of alternate origin sites were observable in a minor number of OTUs but were non-significant when BH-FDR multiple test correction was applied. Due to depth of sequencing, those candidate OTUs for origin-site specificity were typically low in abundance which made statistical comparison difficult. In all instances, potential source site-specific OTUs had abundance which was lower in AV_{worm} than RF_{worm}.

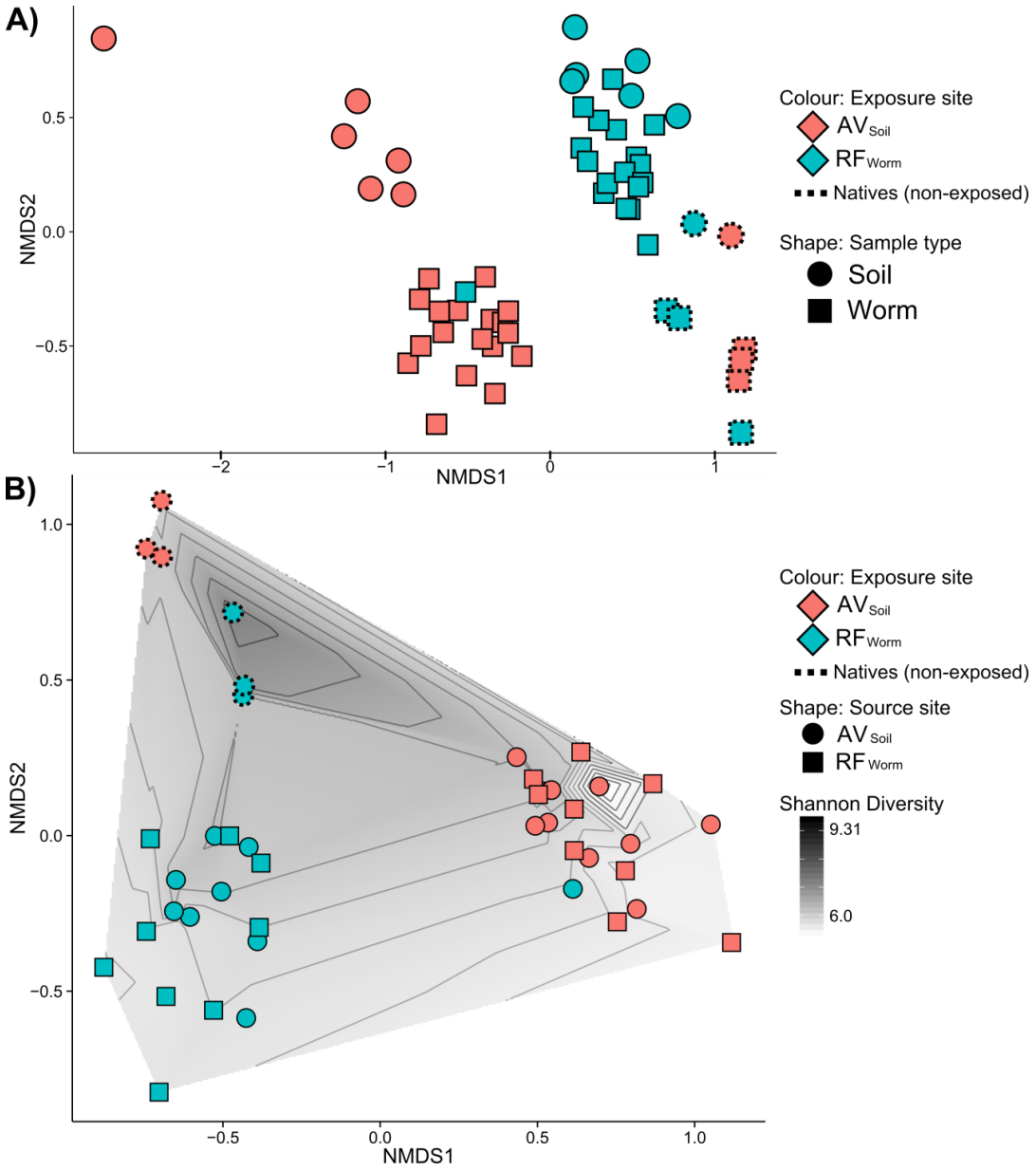


Figure 4.7 - Community structure NMDS analysis of (A) soils and earthworms and (B) all earthworms (Unifrac distance matrices). Colour denotes site during the transplantation study (native earthworms identified with broken line border). In (A) shape denotes sample type; in (B) shape denotes source site. There was significant separation between transplant sites on the primary axis but no differentiation between the source sites of the experimental earthworms. Notable is the separation of native earthworms which demonstrate high variability from the transplanted individuals and the exposure soil.

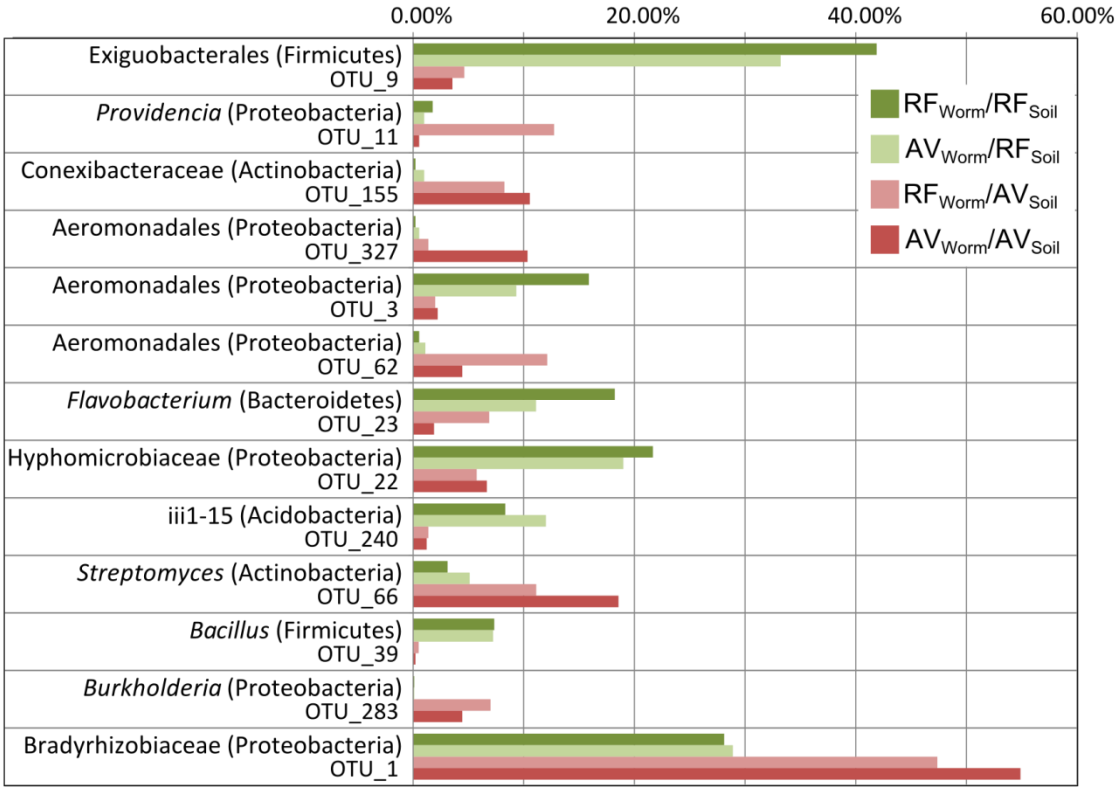


Figure 4.8 - Major OTUs demonstrating significant difference between transplant conditions. Whilst demonstrative of site variation, only Exigobacterales was significant with application of BH-FDR correction (g-test P<0.05)

4.3.3 Functional inference of host-associated microbial groups

Distinct species and community changes were observed between the different transplant conditions which suggest that the associated microbiota have a differing functional capacity. This was investigated through use of PICRUSt, which enables functional inference from the microbial community present. For this analysis, the full metagenome of identifiable species was predicted based upon reference genomes, and the abundance of genes/pathways used to determine an increased/reduced capability of the microbiome to perform various functions.

Five pathways demonstrated a significant difference in predicted abundance between the earthworm host community and the soil environment (ANOVA, BH-FDR correction, P<0.05) with an effect size greater than 0.7 (eta-squared) (Figure 4.9). A significant reduction was found in the host community's predicted ability to synthesise lipopolysaccharides and metabolise glycerophospholipids. Additionally, genes for flagellum production were reduced, indicating a higher abundance of non-motile bacteria within the earthworm host, and reduction of oxidative phosphorylation (not shown), likely due to the internal anoxic environment inhibiting

this pathway's exploitation by the microbiome. The most significantly increased pathway in the earthworm community from both sites was that for ethylbenzene degradation (Figure 4.9c). Although a minor contributing factor, it is among a wide range of other organic compound degrading functions which vary by site (Figure 4.10). The predicted genomic abundance of degradation pathways in the microbiome is higher across the range of *A. gracilis*, significantly more so than the surrounding soil microbiota.

There were no identifiable pathways that showed significant association with the site of origin for transplanted earthworm microbiota, however there were significant effects demonstrated by the incubation time at the transplant site upon the communities functional capacity. The impact of the volcanic conditions on the microbial community can be observed in the biosynthesis of fatty acids (increased) and thiamine metabolism (decreased) in the population of AV_{worm} individuals (Figure 4.11), which may have an impact on the *A. gracilis* phenotype.

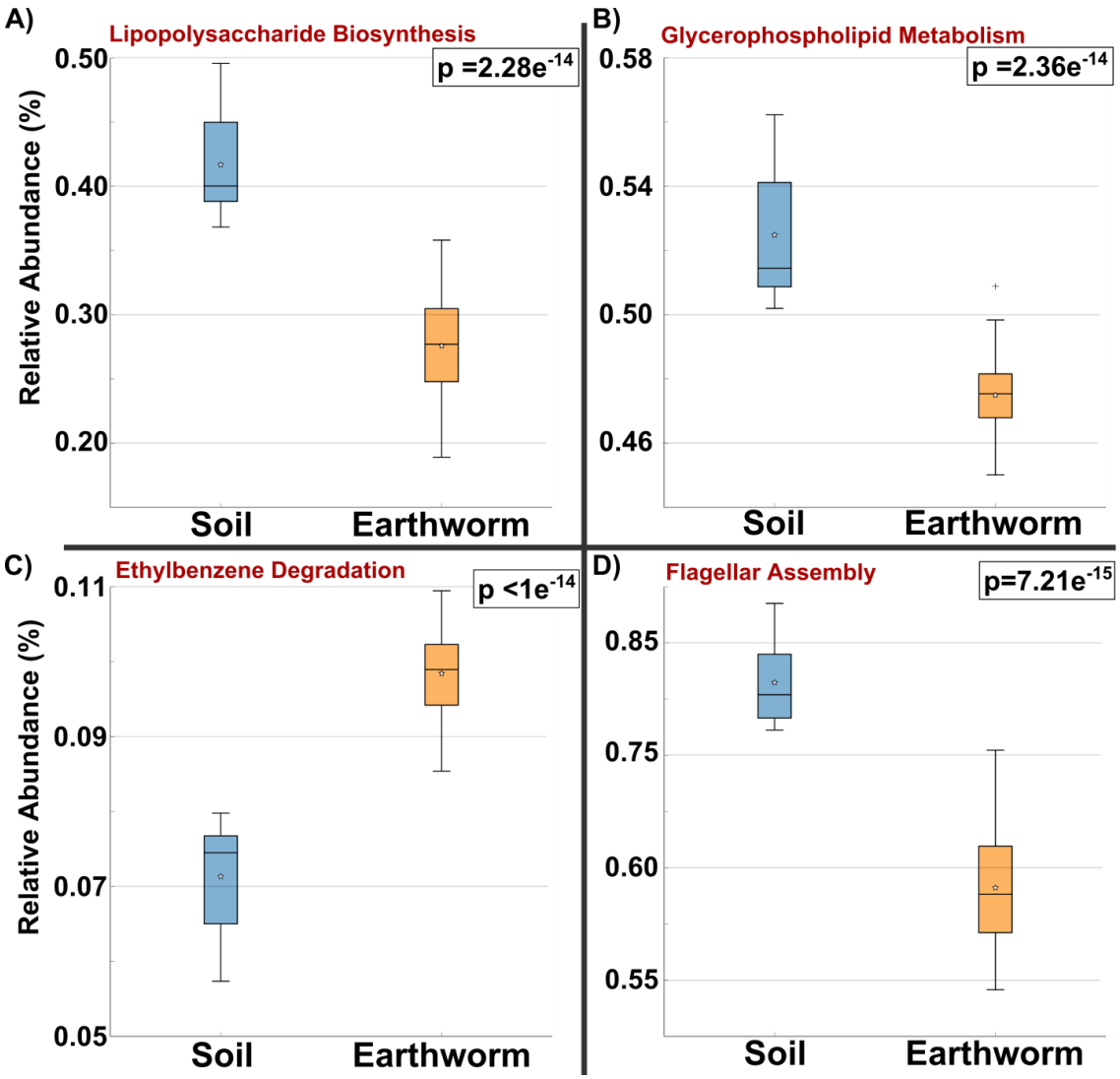


Figure 4.9 - Selected KEGG pathways proposing differing genomic capacity of earthworm-associated microbiota from their habitat soil community. Predicted reduction in relative abundance was observed in (A) Lipopolysaccharide biosynthesis, (B) Glycerophospholipid and (D) Flagellar Assembly. Ethylbenzene Degradation (C) was the most significant earthworm-increase but at relatively minor abundance. Pathway significance displayed in box (ANOVA, BH-FDR correction).

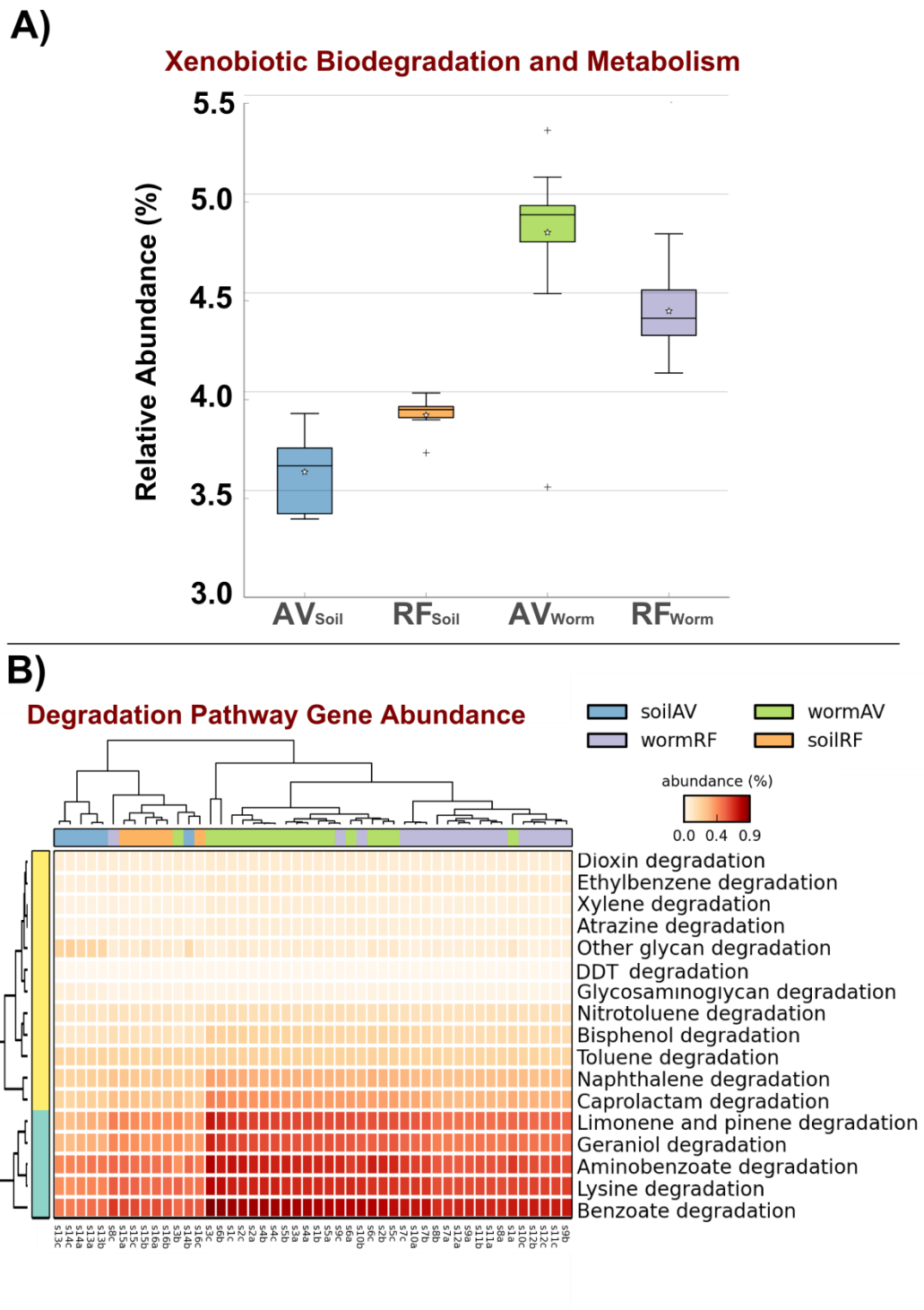


Figure 4.10 - Predicted Xenobiotic degradation KEGG pathways observed throughout microbiomic samples. Samples grouped by transplant condition i.e. includes samples from both origin sites. (A) Higher abundance in the earthworm is increased further in the high stress (AV_{soil}) environment. (B) Additional detail demonstrating most prevalent substrate targets up to 1% relative abundance.

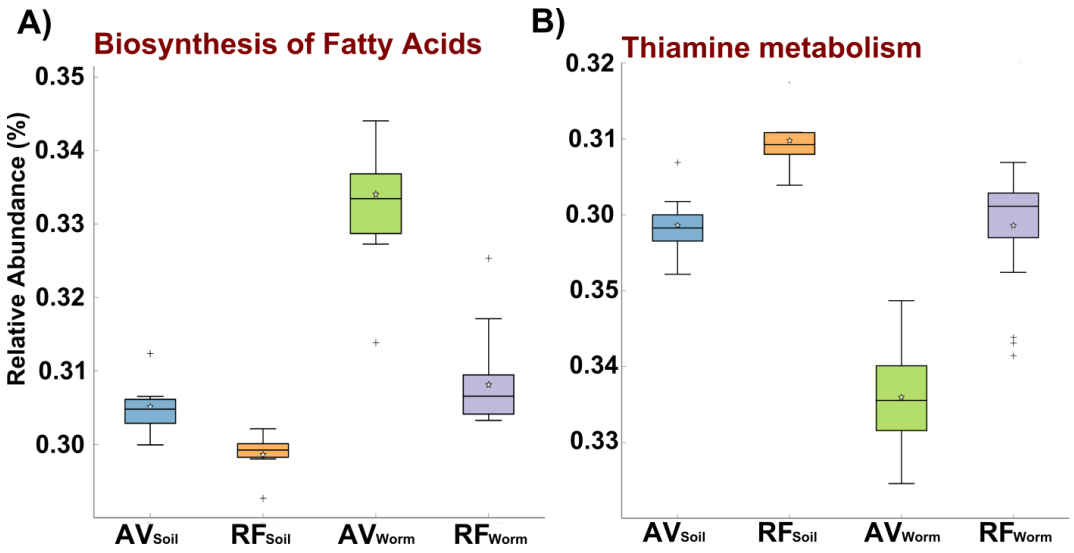


Figure 4.11 - Predicted abundance of AV_{worm} associated KEGG Pathways which were (A) significantly increased (Fatty Acids) or (B) significantly decreased (Thiamine) in the microbial community (ANOVA, P<0.05 (BH-FDR correction)). Samples are grouped by transplant condition i.e. includes samples from both origin sites. Although only a small abundance change is observed, the consistency within site and significance suggests real impact on the community structure.

4.4 Discussion

Through a 30 day cross-transplant study, individuals from both volcanic soils (AV_{Soil}) and a Reference Field site which does not display geothermal vent characteristics (RF_{Soil}) were analysed. It was observed that the microbiome of the earthworm *Amynthas gracilis* was distinctly altered from its control state when residing in an AV_{soil} environment. The bacterial community of the earthworm host was distinct from the surrounding soil environment and had significant changes both taxonomically and in diversity and richness as has been observed previously in *Lumbricus rubellus* (Chapter 3, Pass et al. 2014). The origin of the earthworms had minor effect on the community of the host after transplantation which demonstrated the major effect of the soil microbiomic habitat which individual earthworms resided on during the experiment.

RF_{worm} individuals had a higher bacterial diversity and richness than those exposed to the AV_{soil} conditions which mirrors the microbial communities observed in the soil at the two sites. Major differentiating characteristics of the two experimental transplant sites were CO_2, temperature, flow percentage and metal/metalloid abundance however notably, pH varied between experimental sites by less than 0.3 units. The bottlenecking effect of the earthworm on ingested soil previously described (Drake & Horn 2007; Nechitaylo et al. 2010; Pass et al. 2014) was non-significant in this dataset. The earthworm gut presents a positive environment for transient bacterial species through circumneutral pH and metabolite secretion into the gut lumen (Horn et al. 2003). Individuals inhabiting RF_{soil} display the expected reduced diversity and richness when compared to soil, however in the high environmental stress of the AV_{Soil} the indices are greater in the host when RF_{worm} individuals are transplanted. The intestinal earthworm habitat has the potential to be more sheltered from the high stress characteristics of the volcanic soil allowing species to proliferate, contributing towards greater diversity and richness than within soil where non-native earthworms are transplanted. RF_{Worms} transplanted to AV_{soil} displayed indices higher than the surrounding soil, potentially indicating that the earthworm gut presents a different habitat to the native individuals, supporting a wider bacterial diversity.

The amount of soil contained within an earthworm individual at time of harvest accounted for a significant proportion of the community and differentiating between transient and host-associated communities was difficult at this experimental resolution, both with sequencing depth and accounting for natural soil diversity. In

depth analysis of gut bacteria in fresh and depurated earthworm is further discussed in Chapter 6.

The microbiome of *A. gracilis* individuals were highly similar by community NMDS based upon the mesocosm site on which they resided during the transplant, regardless of their origin site. This supports a modified soil derived microbiome being the basis for the *in situ* bacterial community observed with the earthworm host. At the Phylum level, all earthworms had a markedly similar community profile consisting mainly of the three major Proteobacteria Classes (Alpha, Beta, Gamma), Actinobacteria, Acidiobacteria, Firmicutes, Bacteroidetes and Chloroflexi. At OTU-level there were a number of taxa exhibiting significantly different abundances, most notably from the Proteobacteria clade. The identified Aeromonodales species are an example of taxa which appear specific to their origin site where three genetically distinct OTUs demonstrated uniquely abundant presence at AV_{worm}/AV_{Soil}, RF_{worm}/AV_{Soil} and both transplants to RF_{Soil} respectively. However most taxa are separated only by the transplant site and results indicate that *A. gracilis* does not host an innate microbiome at abundances detectable in the presence of the soil bolus.

The earthworm curates a microbiome which has the potential to provide various greater advantages to the host than the soil community which is a trait observed in other soil invertebrates (Warnecke et al. 2007; Goffredi et al. 2007; Moya et al. 2008). Through metagenomic inference there was significant predicted variation between soil- and host-communities in the abundance of several KEGG pathway associated genes. Host-communities suggested a lower abundance of genes involved in production of flagellum structures, implying the higher relative abundance of non-motile species within the earthworm. This implies a greater degree of structure-bound species yet does not relate to the *Verminephrobacter* symbiont which is known to require flagella for nephridial colonisation (Dulla et al. 2012). An increased proportion of bacterial species with predicted degradation pathway genes (most significantly variant being ethylbenzene degradation) were found in earthworms in comparison to the soil, although notably a number of degradation pathways were associated much more significantly with the volcanic soils. It is important to recognise however that the inferred metagenome process has a number of assumptions and likely over-represents functionality based upon a subset of the KEGG pathway being defined. Whilst highlighted pathways may be conflicting, it is indicative of a homologous system and provides additional evidence when considering the effects of potential taxonomic change upon the functional capacity. It is likely

that the higher toxicity of this environment is more survivable by bacterial species with the genomic capacity of stress reduction, resulting in their increased proportion. However, in the context of the earthworm host, there is benefit to be achieved by encouraging a microbial community which can actively reduce the toxicity of the soil which the individual ingests. Pathways for anoxic degradation of aromatic hydrocarbons such as those described above would be of benefit to the earthworm when inhabiting the stressful volcanic soil environment (Fuchs et al. 2011). Conceivable is a mutualism similar to the microbiota known to enable a range of insects to feed on plants with toxic metabolites through degradation of lethal compounds (Hansen & Moran 2014), however significant study would be required to demonstrate the activity of the identified degradation pathways, the utility of them for the host and furthermore the action of the host in encouraging such communities.

The abundance of genes enabling the microbial community to synthesise fatty acids was moderately (yet significantly) greater in AV_{soil} transplanted earthworms. The *A. gracilis* individuals here studied have previously been shown to increase fatty acid metabolising pathways in the host transcriptome of AV_{soil} earthworms (Novo et al. n.d.), however linking these aspects is beyond the possibilities of this data. Such an interaction between the microbiome and the host, where the bacterial community is producing a metabolite directly utilised by the earthworm, would be a novel discovery in earthworms and implication that the host is active in cultivating a beneficial microbiome is of interest for future study. Thiamine (vitamin B1) metabolism was one other significantly altered pathway present in a notable abundance. Here although only minorly modified, the ability for the host community to perform this function was reduced solely in the AV_{worm} individuals, even in relation to the surrounding soil. With an intrinsic metabolite such as this, it is difficult to discern the impact which this could have with the data present, but the absence of these pathways from the host community again indicates a host-environment specific, functionally-relevant microbiome.

No significant variations in the functional pathways were identified based upon the origin of the host microbial communities, further indicating that the earthworm-associated microbiome was almost entirely directed by the environment they were incubated within. However, the host's ability to alter the microbiome through its unique internal environment can benefit the animal regardless, through encouraging the proliferation of valuable taxonomic groups. The greater anoxia in the volcanic environment, in connection with higher toxicity and temperature could alter

metabolic pathways performed by the host and thusly the metabolites which earthworms are known to secrete into their mid- and hindgut, priming bacterial growth (Wüst et al. 2011; Drake & Horn 2007). A multi-omic approach to identify changes in metabolite secretions by the earthworm would assist in understanding any positive effect that the host could have on beneficial gut bacterial species.

Of note are the native earthworms, whose microbiomic characteristics demonstrated distinct variance from those used in the transplant. Acute stress of the physical transplantation would have incurred only a transient effect on the individuals and the two major factors were likely the differing characteristics between the earthworm source site and the transplant fields, and the unnatural characteristics of the 15 L mesh mesocosm. Visualisation of the site data through NMDS analysis highlights that the geochemical variation between the AV_{soil} origin/exposure sites is much higher than the RF_{soil} (Figure 4.12 (source data: Figure 4.1)). Particularly the CO_2 flow and volume are elevated at the exposure site than the AV source site, in addition to copper and lead abundance. While investigation of this highly stressed site was the intent, the AV exposure site (AV_{soil}) also showed that the individuals which were utilised in the study (from $NatAV_{soil}$) were not fully acclimatised to the AV exposure site prior to the experiment and could account for variations. Also, the AV source site has a soil with a more sandy composition than the other sites which may impact on the soil and host-associated microbiome composition. The reference (RF) sites are clearly much better associated, with almost identical environmental characteristics, although reporting higher abundance of several metals than the volcanic sites.

Finally, it must be noted that the 'native' individuals were not under the same level of environmental stress due to not inhabiting the transplant sites. In the RF_{soil} sites, the similar environmental characteristics between source and exposure still exhibit taxonomic and diversity variation between Native and mesocosm individuals, however the severity of variation between the $NatAV_{soil}$ source and the AV_{soil} of the transplant mesocosms likely accounts for the differing profiles of native and experimental earthworm microbiomes. Regardless, this does not impact the direct comparison of experimental sites and is of more interest when discerning the effect of the minor site differences.

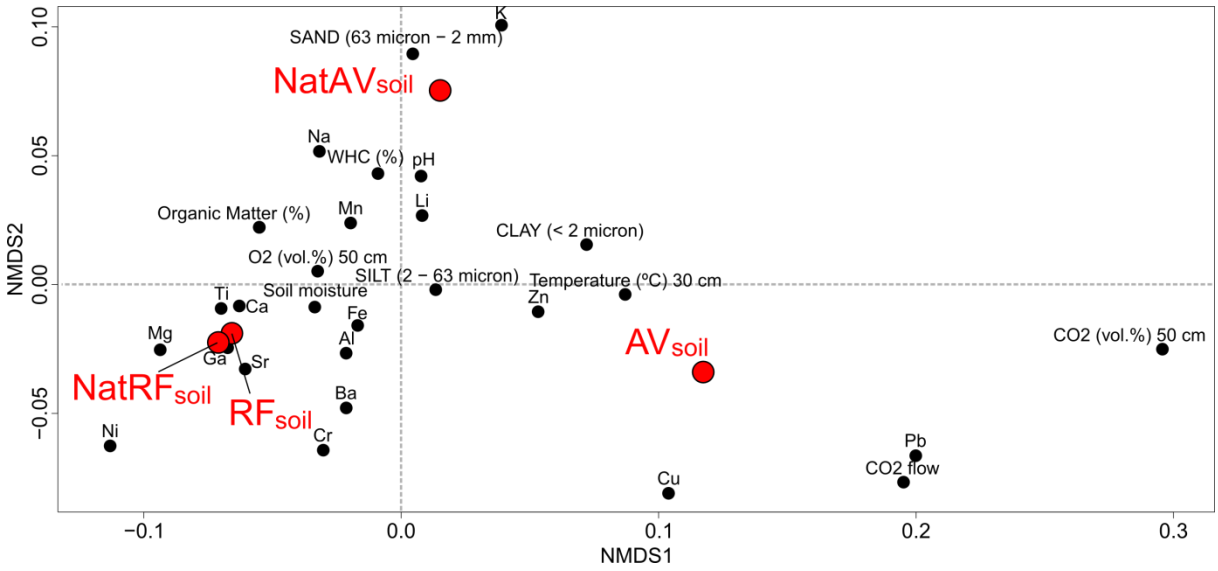


Figure 4.12 - NMDS dispersion of sites based upon environmental characteristics. The sites from which *A. gracilis* were harvested (Nat##$_{soil}$) demonstrated differential association with the mesocosms exposure sites, where the AV$_{soil}$ environment exhibited (primarily) greatly increased levels of CO_2 flow and volume, copper and lead.

4.5 Conclusion

Results here presented corroborate previous studies which indicate that earthworm species host a modified microbial community *in situ* in regards to that of the surrounding soil, and is the first example in this taxonomic order. However, the cross-transplantation technique has shown that at this sequencing resolution, at least in the context of full gut bolus, the host microbiome is derived almost entirely by the site microflora which the earthworm ingests. The influence of geothermal volcanic-vent stressors on the host has distinct impact on the internalised microbiota but notably, there are instances where the associated bacteria vary solely in the host itself, aside from the surrounding soil community. This indicates a specialisation which may have functional connotations for the host, and is of importance when attempting to identify any beneficial, mutualistic or symbiotic species between earthworms and associated bacteria.

5

The impact of diverse soil geochemistry and host phylogeography on the microbiomic hindgut community of *Lumbricus rubellus*

5.1 Chapter Introduction

Previous work has described the earthworm-associated microbial communities arising from transient soil and internalised species, however these studies have been limited to single geographic locations (Chapter 3) or a binary comparison (Chapter 4) between two distinct sites. Given the complex multifactorial nature of the soil environment and the evident soil-derived microbiome of the earthworm associated bacterial population, analysis of earthworm microbial profiles from a wide range of habitats are required to determine if a core bacterial community is present independent of the soil community, and the impact of varied biotic and abiotic conditions. In soil invertebrates the hindgut often contains a unique microbiomic community arising from specialised host structures e.g. Termites (Warnecke et al. 2007) and Beetles (Egert et al. 2003; Egert et al. 2005; Schloss et al. 2006) where bacterial populations increase the capability of organic carbon uptake by the host (Kappler & Brune 1999). Given the linear nature of the earthworm gut there is less capacity for adherence of microbes to gut structures, however hindgut-associated bacteria have previously been described which demonstrate a higher abundance than the soil microbiota and anterior gut locations (Singleton et al. 2003; Wust et al. 2009). The ecological importance of earthworms stems from the effects on the transient soil substrate and subsequently excreted cast material (Blouin et al. 2013). Determining the nature of the hindgut community would benefit both knowledge of the earthworm microbial association and the impact of this ecosystem engineer's microbiota on the environment for which it plays such an essential role.

The microbial structure of the soil which earthworms inhabit and ingest is known to be most influenced by the pH, with organic carbon, nitrogen and toxic metals as secondary influences (Griffiths et al. 2011). The earthworm gut offers a more pH-neutral environment for transient microbiota (Drake & Horn 2007) and is a likely cause of the reduced Acidobacteria observed in the gut lumen (Chapters 3 & 4; Pass et al. 2014). However notable variances in soil pH can occur within relatively small spatial distance and are a major influence on the soil microbial structure (Rousk et al. 2010) from which earthworms largely derive their associated microbiota (Ihssen et al. 2003; Pass et al. 2014). Earthworms are capable of withstanding soils contaminated with heavy metals, often at levels toxic to other species including humans (Morgan et al. 2007). In doing so they are able to alter the bioavailability of the soil metals (Blouin et al. 2013) which is further augmented by their ability to bioaccumulate the metal into their body tissues (Nahmani et al. 2007).

The phylogeographic distribution of *L. rubellus* is complex across the UK and Europe given that the taxonomically named species consists of at least four cryptic species when assessed molecularly (Sechi et al. submitted). Within the UK two lineages are established (A and B) and are distinct with 14% divergence of the Cytochrome Oxidase II gene (Donnelly & Harper 2013). Phenotypic difference between the lineages of *L. rubellus* are not overtly apparent but morphological differentiation can be reliably achieved through visual assessment of the anterior glandular tumescence (Donnelly et al. 2014). The first metabolic differentiations have only recently been identified through detection of varied laminine production between lineages A and B (Liebeke et al. 2014), however further work is required to determine the nature of the lineage variance. No study to date has assessed the microbial communities of *L. rubellus* with Next Generation Sequencing across a geographically scale of more than 20 km (Pass et al. 2014) nor addressed the lineages when describing the communities. The role of bacterial species in invertebrate speciation has been discussed, demonstrating the unviability of interspecies hybrids due to gut-symbiont incompatibility (Brucker & Bordenstein 2013) however the concept of adaptive codivergence is contested (Chandler & Turelli 2014). The relative local isolation of many earthworm species is likely instrumental in the genomic separation of the lineages and if the earthworm population was found to be dependent upon certain clades of the prokaryotic community then associated bacterial changes could occur within rapid time frames.

This study assesses the microbial community associated with *L. rubellus* of various lineages and sampled from across the UK from 17 clean and polluted sites, representing a wide range in environmental conditions from which to establish the core microbiome of the earthworm, and the microbiota responsive to varied soil types.

5.2 Methodological Approach

5.2.1 Site Description and Experimental Design

L. rubellus individuals were collected from 17 sites across the UK, 6 sites were known areas of high anthropogenic contamination and were directly paired with 6 from a close, ostensibly uncontaminated site. Additionally, 4 locations were sampled from Environmental Change Network (ECN) sites (Morecroft et al. 2009) which have been historically monitored and represent known clean sites. The sample origins are summarised in Figure 5.1 below.

UK Site	Short code	Site Type	Latitude	Longitude
Avonmouth	AMT	Control	51.536596	-2.6212709
		Polluted 1	51.531898	-2.6631155
		Polluted 2	51.51173	-2.6662809
Cwymyswyth	CWM	Control	52.352801	-3.7692123
		Polluted	52.361558	-3.7511454
Devon Great Consuls	DGC	Control	50.544686	-4.2234306
		Polluted	50.536848	-4.2255316
Port Talbot	PTB	Control	51.548033	-3.6787916
		Polluted	51.555036	-3.7466883
Scunthorpe	SCT	Control	53.596303	-0.5939021
		Polluted	53.630468	-0.5679044
Shipham	SHP	Control	51.331229	-2.7719859
		Polluted	51.311879	-2.7933687
Alice Holt	AHT	ECN	51.152057	-0.85670149
Drayton	DRA	ECN	52.193547	-1.7616291
Porton	PDW	ECN	51.138167	-1.7215882
Snowdon	SND	ECN	53.075822	-4.0364476

Figure 5.1 - Geographic location and descriptions of sites investigated

Adult *L. rubellus* were visually identified upon capture with later confirmation via COI barcode sequencing prior to selecting individuals for sequencing (Chapter 2.1.2.4.3). Stochastic geographic distribution of the two major UK *L. rubellus* lineages (A and B) was observed as expected i.e. some sites were monogenic and some polygenic for the two major earthworm lineages previously identified in the UK. Both lineages were included within this study to enable an assessment of potential for lineage specific microbial communities. Sections measuring ~1 cm were taken from the tail end of the earthworm, from which total DNA was extracted. DNA extraction was performed using the Tepal automated extraction method described previously (Chapter 2.1.2.3.2). PCR of the amplicon region 357F-518R was performed with unique barcodes included in the primer design for post-sequencing sample

identification and subsequently, 170 samples (10 individuals per site) plus 24 negative controls were submitted for Ion Torrent sequencing (Chapter 2.1.3.2). In total, 118 individuals were available for analysis following quality control of the sequencing data.

5.2.2 Bioinformatic Methodology

The informatics processing steps performed are described in detail with source references in Chapter 2 and sample descriptions in Appendix 1c. Approximately 11.5 million reads were produced from 6 IT sequencing runs (2 partially successful runs and 4 with full output) with 7.18 million being retained after quality filtering. Each amplicon sample was subsampled to 20,000 reads which was the most common minimum level. At least 5 individuals from each site were of sufficient quality for analysis. A brief overview of the steps taken follows:

Stage	Chapter	Software	Description
Input Data	2.1.3.2	Ion Torrent PGM	11,470,266 sequence reads (6x 316 chip)
Quality Control	2.2.1	UPARSE	Primer mismatches removed, <150bp reads discarded, Reads trimmed to 150bp, Singletons removed 7,852,007 sequences utilised
'Denoising'	2.2.1.1	UPARSE	Incorporated in UPARSE pipeline.
Contamination Filtering	2.2.1.2	Bespoke script	Remove host & eukaryotic sequences (7.5% across dataset, ~0.5 Million)
Subsampling (normalisation)	2.2.2	Qiime	Randomly subsample to minimum sample sequence count (20,000 reads)
OTU generation	2.2.2	UPARSE	OTUs were generated at 0.97
Taxonomic Annotation	2.2.2	BLAST, Greengenes database	Blast annotation at <0.001 E value
Phylogenetic Alignment	2.2.2	FastTree	Downstream diversity analysis
Alpha Diversity Analysis	2.2.3	Qiime	Shannon, Chao1, Observed Species
Beta Diversity Analysis	2.2.3	Qiime/Unifrac	Jackknifed UPGMA analysis

5.3 Results

5.3.1 Profiling the *Lumbricus rubellus* microbiome from multiple UK sites.

The bacterial microbiome of *L. rubellus* inhabiting polluted and control sites from 6 UK locations, and 4 clean sites from the ECN were analysed, resulting in 17 site-specific biologically replicated profiles. Geographic distribution of the sites is represented in Figure 5.2 with the major bacterial phyla composition. Proteobacteria were annotated individually due to collectively composing over 90% of communities in some individuals. Gammaproteobacteria were the most dominant taxon (the class presented a higher abundance than any phylum excluding Proteobacteria) but varied significantly between sites (Figure 5.3). No phyla or Proteobacteria class were in abundance greater than 7.7% at all sites. A number of key phyla (i.e. Acidobacteria, Firmicutes and Chloroflexi) were present at less than 0.5% at some sites suggesting that their presence in the *L. rubellus* microbiome is not essential for host survival and varies in response to the environmental conditions.

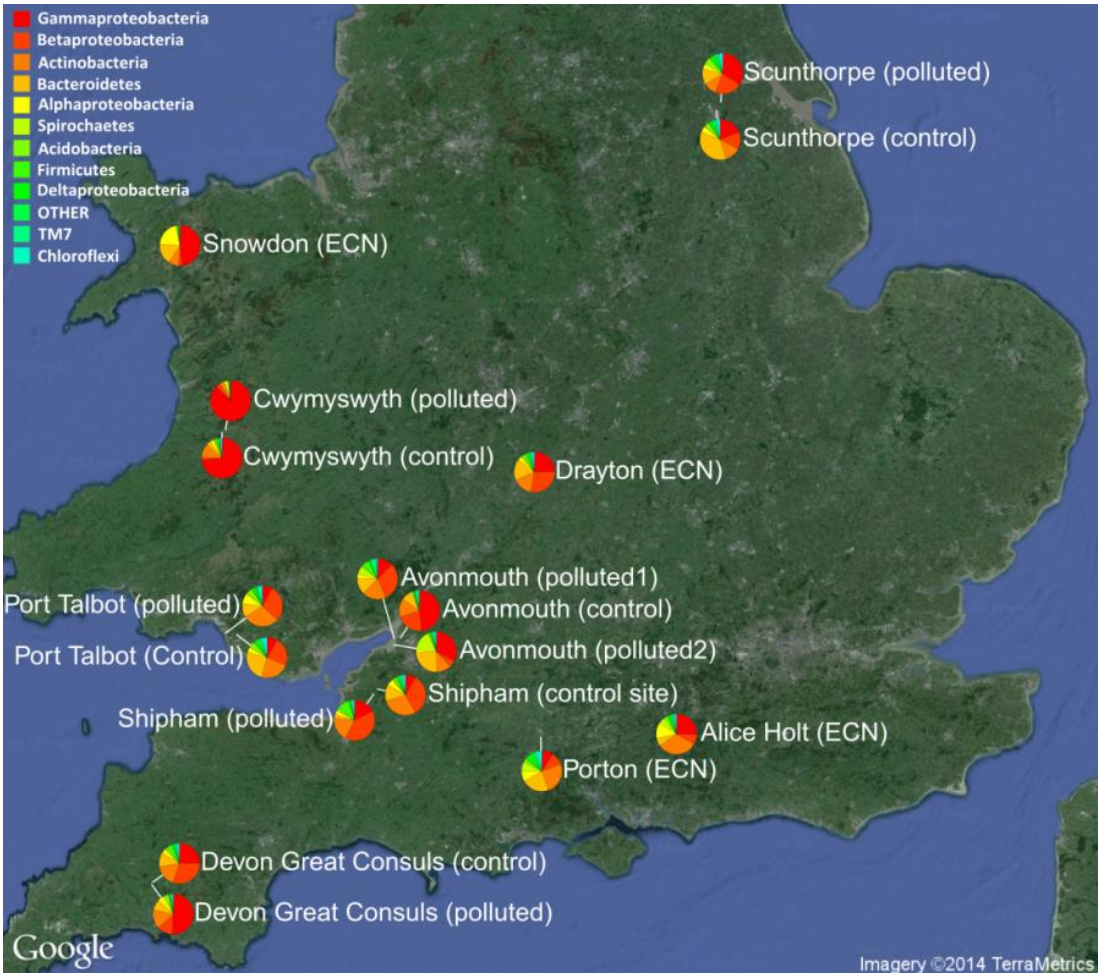


Figure 5.2 - Locations of earthworms from 17 UK sites which were harvested, with bacterial phyla overview demonstrating high variance in the major taxa. Proteobacteria annotated at class level.

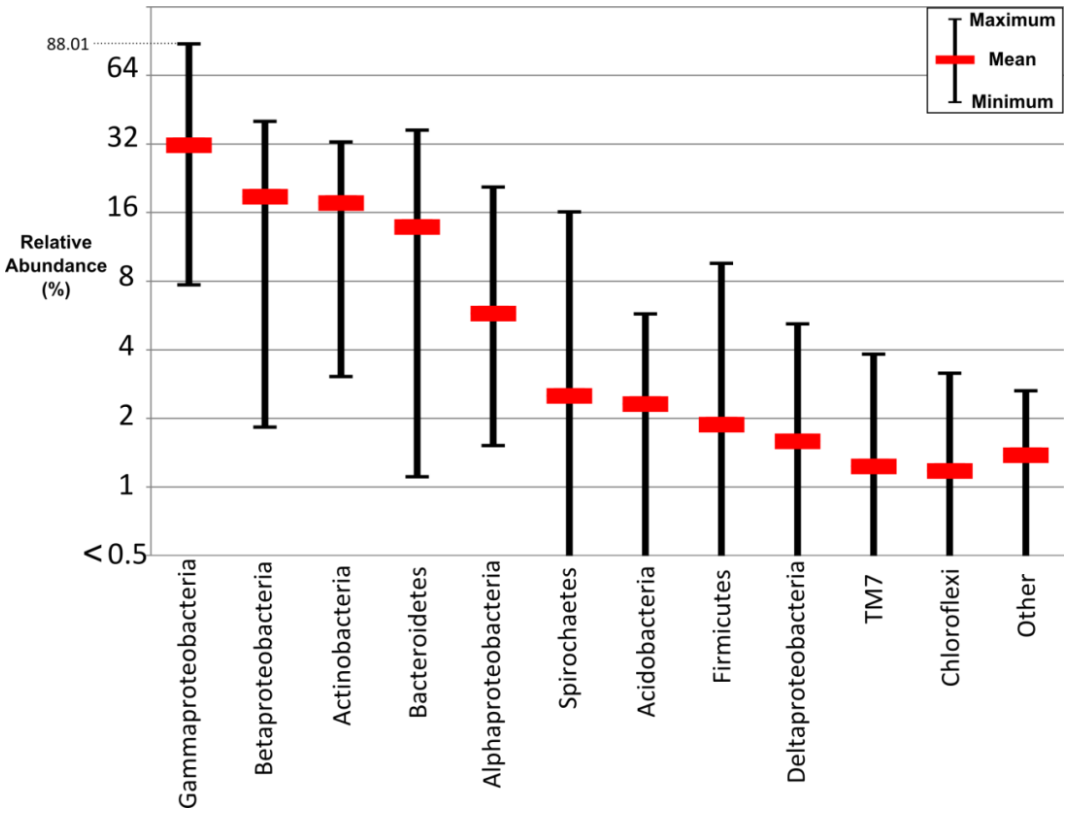


Figure 5.3 - Average, Minimum and Maximum phylum abundance across the UK dataset. Each site was summarised prior to calculation, therefore max/min heights represent average for whole sites i.e. not individual earthworm microbiome profiles. High levels of variance were observed resulting in only 5 phyla or classes present at every assessed site.

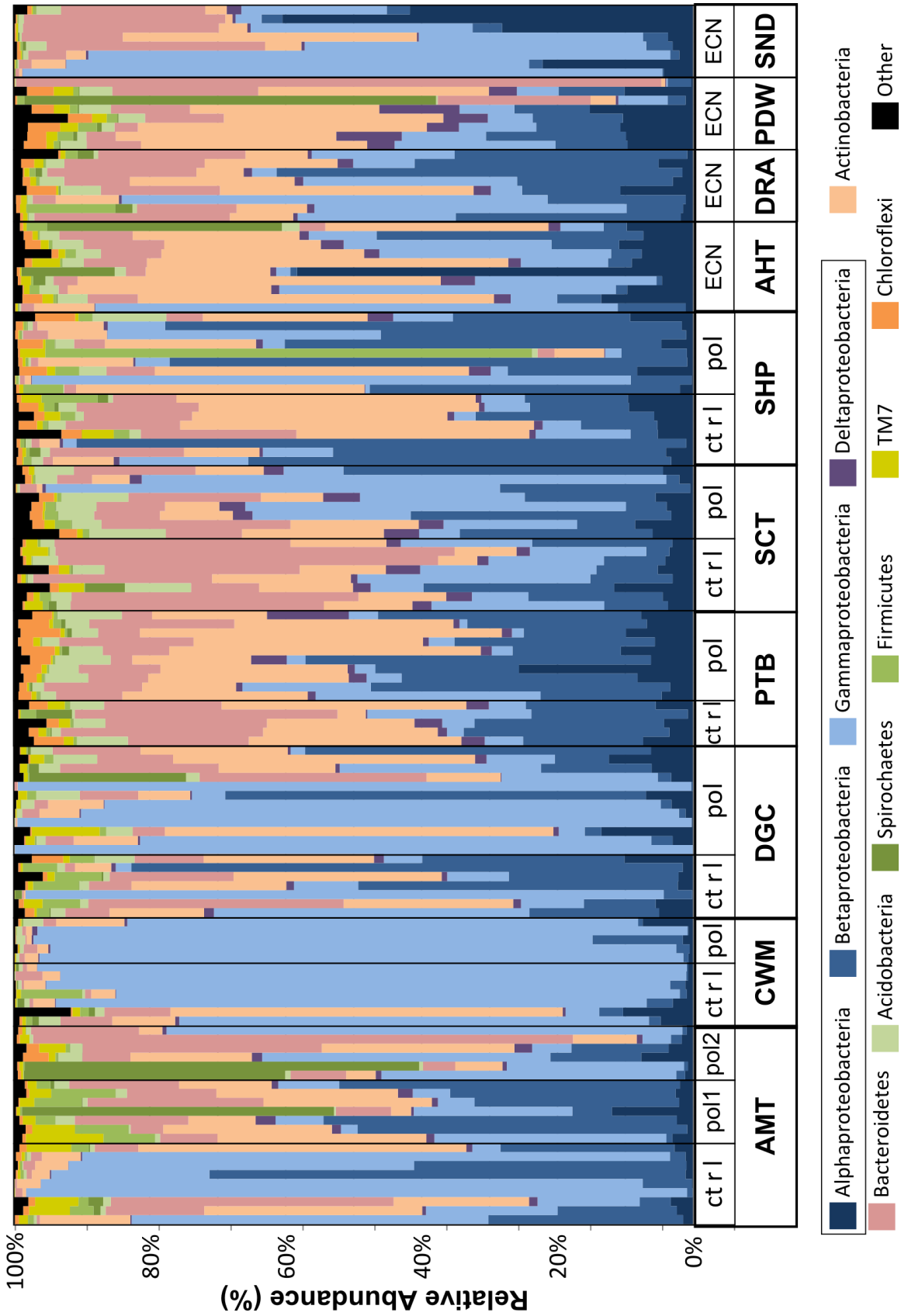


Figure 5.4 - **Relative phyla abundance of UK *L. rubellus* microbiomes. Proteobacteria classes individually represented (blue). Sites differentiated by contamination status: ctrl (control), pol (polluted). A high variance was observed across the range but was consistently dominated by Proteobacteria and Actinobacteria.**

5.3.2 The impact of environmental characteristics on the bacterial community

Diversity and Richness metrics were calculated for each individual and contrasted by polluted and control sites. There was not a consistent response to general pollution on the community structure of the *L. rubellus* microbiome (Figure 5.5). The DGC site showed significantly reduced diversity (observed species) and richness (t-test, $P<0.05$) in the contaminated site corresponds with previous analysis of this site (Chapter 3; Pass et al. 2014). Conversely the AMT site demonstrated significantly higher diversity and richness in the contaminated soil in comparison to the cleaner site (t-test, $P<0.05$).

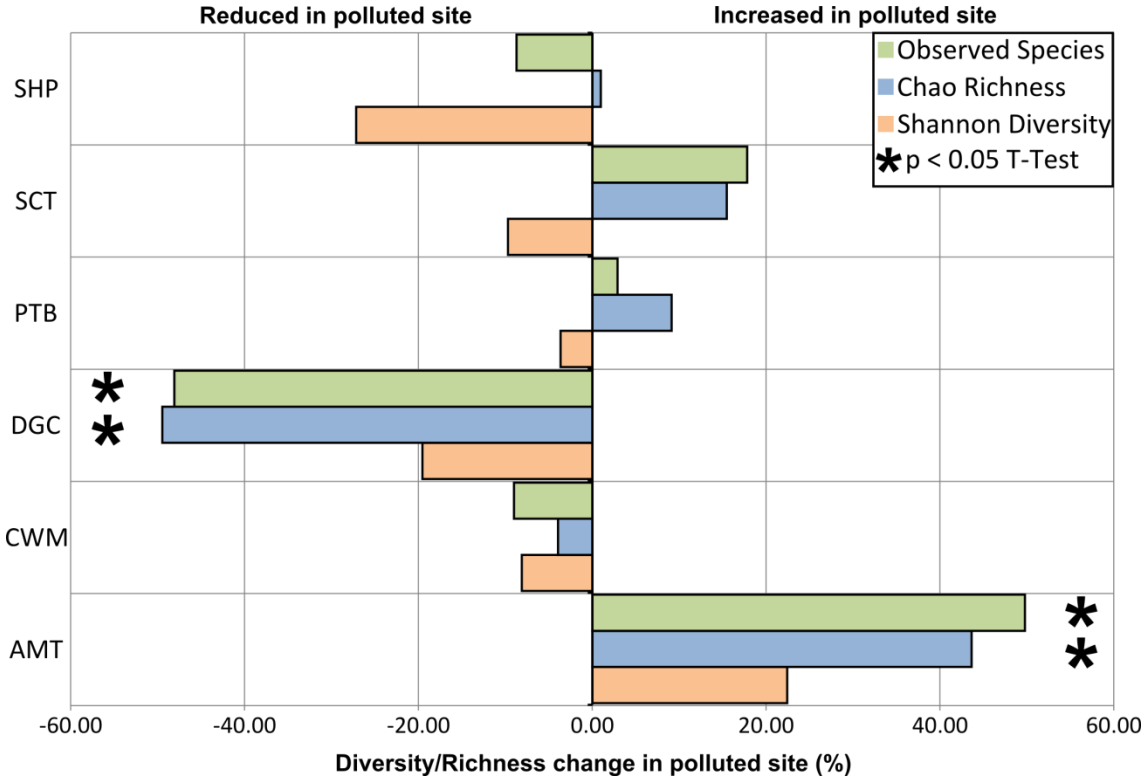


Figure 5.5 - Changes in diversity and richness between earthworm microbiomes from polluted and control sites. Differences were only significant in populations from DGC and AMT (t-test, $P<0.05$).

Assessment of the community profile of each individual was performed utilising the unifrac measure of distance and assessed with Unweighted Pair Group Method with Arithmetic Mean (UPGMA) hierarchical clustering (Figure 5.6). The association between individuals was largely capture-site specific although did not correspond to geographic location indicating that environmental characteristics were more influential than the distance between sites in most cases. Additionally, the lineage of the *L. rubellus* individual was not a determining factor on the association of

community profiles. No sites provided equal numbers of lineage A and B when harvested for earthworms and analysis of the dataset both in entirety and individual assessment of control, ECN and polluted sites was performed. There were no taxonomic families which associated with either lineage A or B for Control-site sourced, ECN-sourced or Control and ECN sourced individuals (ANOVA, $P<0.05$, BH-FDR correction, Not shown). Polluted sites were omitted from calculations due to the conflicting variety in environmental conditions.

In determining the correlation of environmental factors/contamination with the identified community structure, 4 distinct groups were identified based upon unifrac distance profiles (Figure 5.6). These 'ecotypes' demonstrated only minor association with environmental characteristics (Figures 5.7 & 5.8) although the breadth of group 2 encompassed a wide range of sites that made it impossible to significantly associate with soil or worm measurements. To ensure accurate associations, groups were correlated with tissue metal abundance rather than measurements taken from soil where available. Group 1a was characterised by higher abundance of iron, chromium and aluminium in the tissue of the earthworm host and also demonstrated the highest diversity and richness of the groups. Group 1b did not display distinguishing features but consisted primarily of individuals from ECN i.e. unpolluted sites. Group 3 were associated with high arsenic and lead as measured in the body tissue of the earthworms.

The soil moisture, organic matter and pH are key influencers of diversity and richness in the soil bacterial community (Griffiths et al. 2011) and were assessed for impact within the earthworm gut (Figure 5.8). Increasing moisture content of the soil resulted in reduced bacterial richness (Chao1, R^2: 0.3815) and to a weaker correlation diversity (Shannon, R^2: 0.1629, Observed Species, R^2:0.2925). This directly correlated with the organic matter content of the soil demonstrating a higher diversity and richness in low organic content environments. Acidic soils had a low correlation but indicated a reduction in diversity and richness, with more neutral soils demonstrating a richer community.

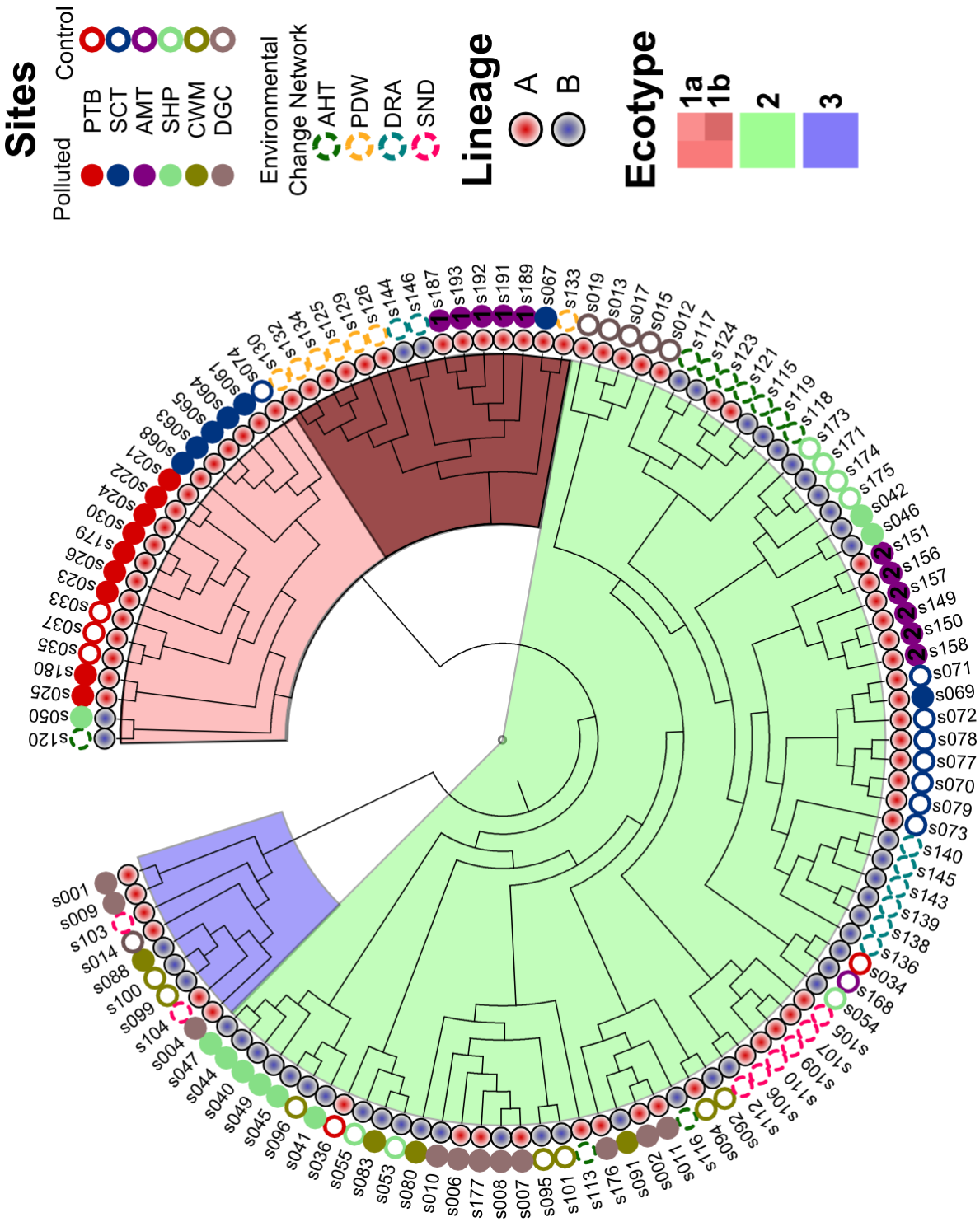


Figure 5.6 - **Weighted unifrac UPGMA tree (Jackknifed) of all microbiome profiles. Inner ring: *L. rubellus* lineage, outer ring: Site. Filled circles: polluted sites, Empty circles: control sites. ECN sites are noted with intermittent outer lines. 3 distinct separations were detected based upon community profile and denoted as ecotypes for further analysis.**

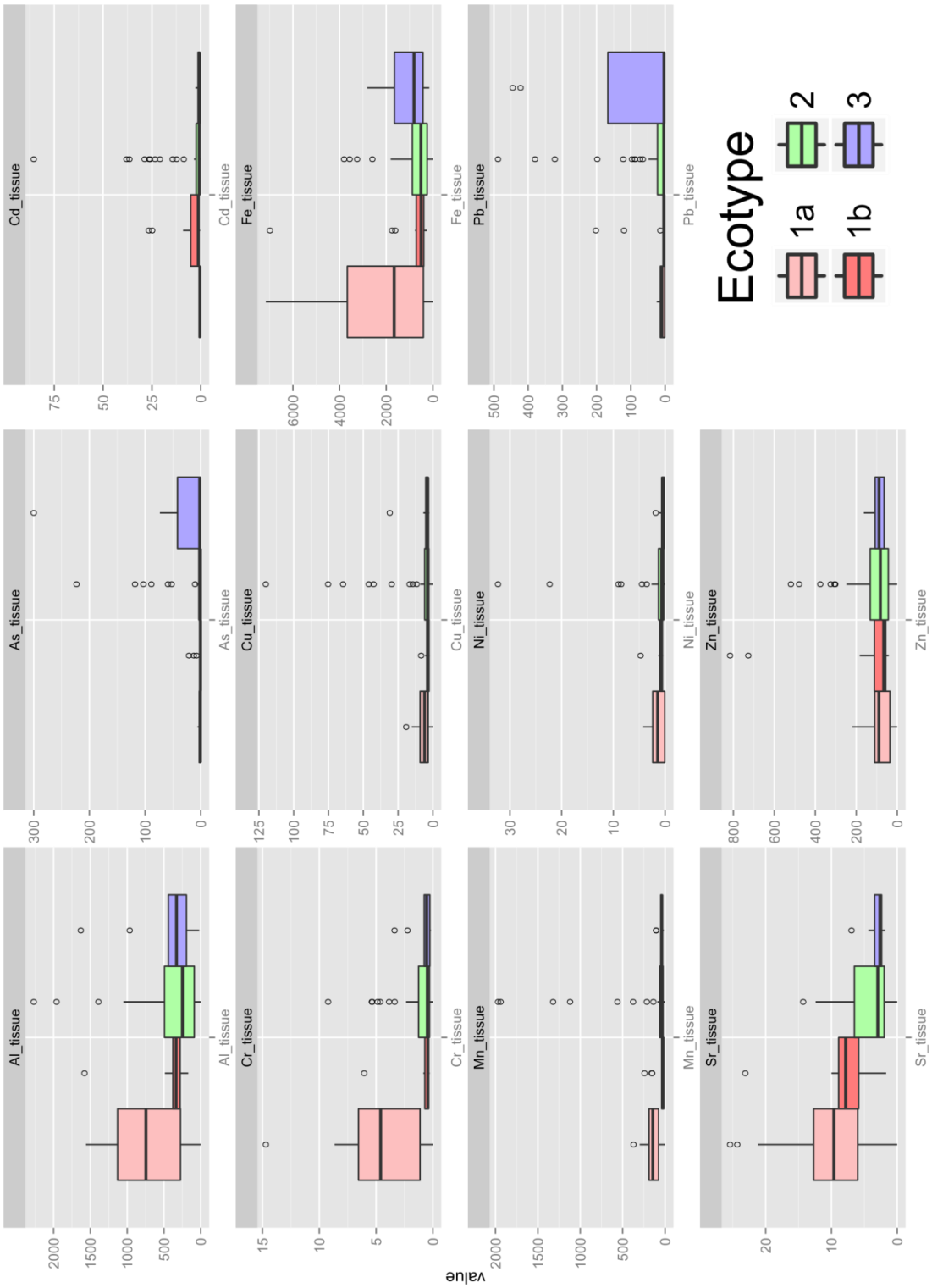


Figure 5.7 - Relationship between metal tissue abundance and community structure ecotypes. Group 1a demonstrated the greatest deviation based upon metal contamination due to the association of Al, Cr and Fe. The samples denoted 'Group 2' include a wider range than the other identified groupings and appears too diverse to associate on particular metal contamination.

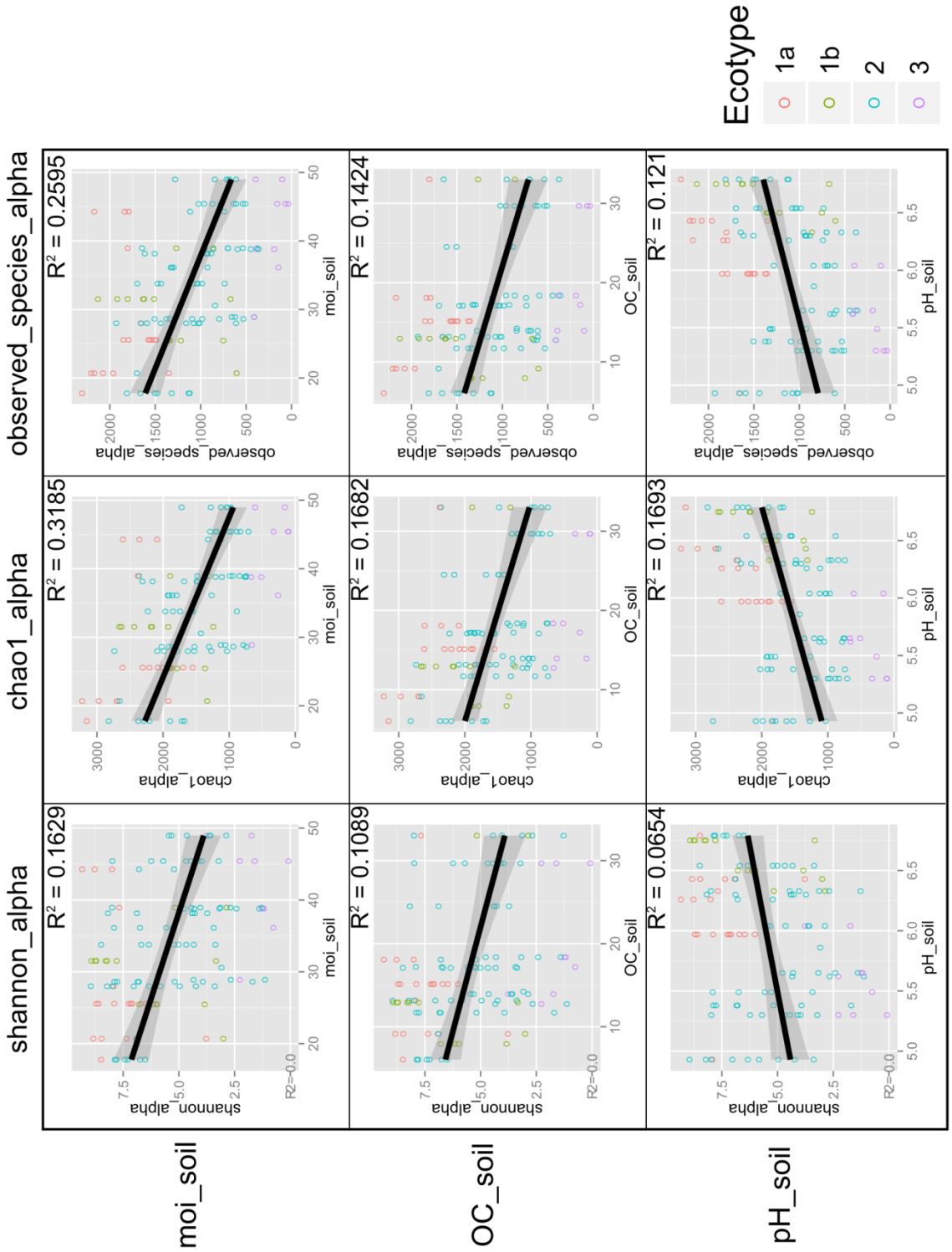


Figure 5.8 - **Soil characteristics correlating with diversity and richness metrics. Linear model and SE plotted, with R^2 in upper corner of plots. Increased moisture (moi) content of the soil had reduced bacterial richness and to a lesser degree diversity. Organic matter content (OC) of the soil demonstrated weaker negative association. Acidic soils had a low correlation but indicated a reduction in diversity and richness.**

5.3.3 Association of bacterial families to environmental characteristics

Beyond the whole-community association, the correlation of individual taxa with the surrounding metal contamination and soil geochemistry was investigated. There was significant association of a number of families when relative abundance was correlated as determined by Spearman's rank correlation (Figure 5.9). The top 40 taxonomic families (greater than 0.21 correlation) to at least 1 environmental factor were included alongside the 10 factors which had the largest influence.

There was linkage between the iron and aluminium factors possibly due to their common co-occurrence, most frequently as a negative correlation to the bacterial family. Phosphorus and calcium had inverse taxonomic correlations frequently associated with increases in bacterial abundance. They also associated with pH likely due to the influence these metals have on soil acidity.

Legionellaceae (which singularly represents the *Legionella* genus) demonstrated the strongest correlation to a single parameter through positive association with arsenic tissue levels (Spearman's Rho: 0.87). The majority of families displayed greater abundance with the more neutral pH that related to the general increase in richness and diversity (Figure 5.8). Koribacteraceae was conversely impacted by the factors, positively correlating with raised iron/aluminium and soil water content, and higher abundance in acidic soils. Propionibacteriaceae was highlighted due to the strong positive correlation with soil manganese although to a lesser degree with tissue levels.

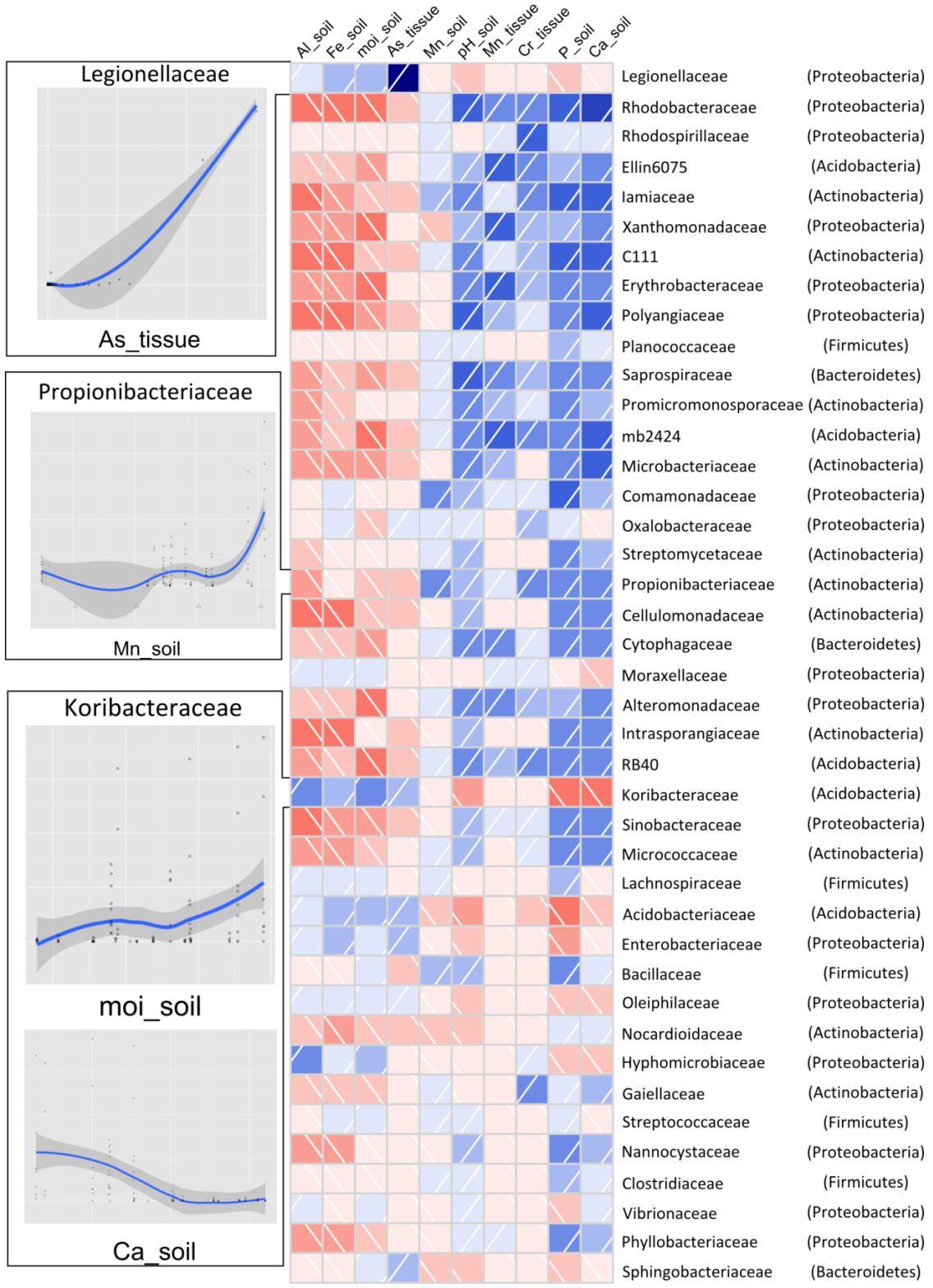


Figure 5.9 - Taxonomic Family associations with key environmental factors. Interaction boxes coloured by direction and intensity of Spearman's Rho correlation (Blue: Positive, Red: Negative). Families ordered descending by largest Rho statistic. Inserts demonstrate individual correlations for notable interactions (Loess smoothing).

5.3.4 The core microbial community of *Lumbricus rubellus* hindgut

A large variation in bacterial community structure was present across the nationwide dataset; however a number of taxonomic families were consistently identified. There were 37 taxonomic families identified in 90% of all individuals profiled, but only 3 in every individual across the sample range: Comamonadaceae (the *Verminephrobacter* symbiont), Flavobacteriaceae and Mycobacteriaceae. There was higher consistency when viewing the control and ECN sites in isolation, with 19 Families detected in 98% of *L. rubellus* microbiomes, and a greater proportion detected at >0.1% abundance richness demonstrating a more reliable core community associated with lower-stressed sites and the polluted sites displaying a more varied microbial habitat (Figure 5.10).

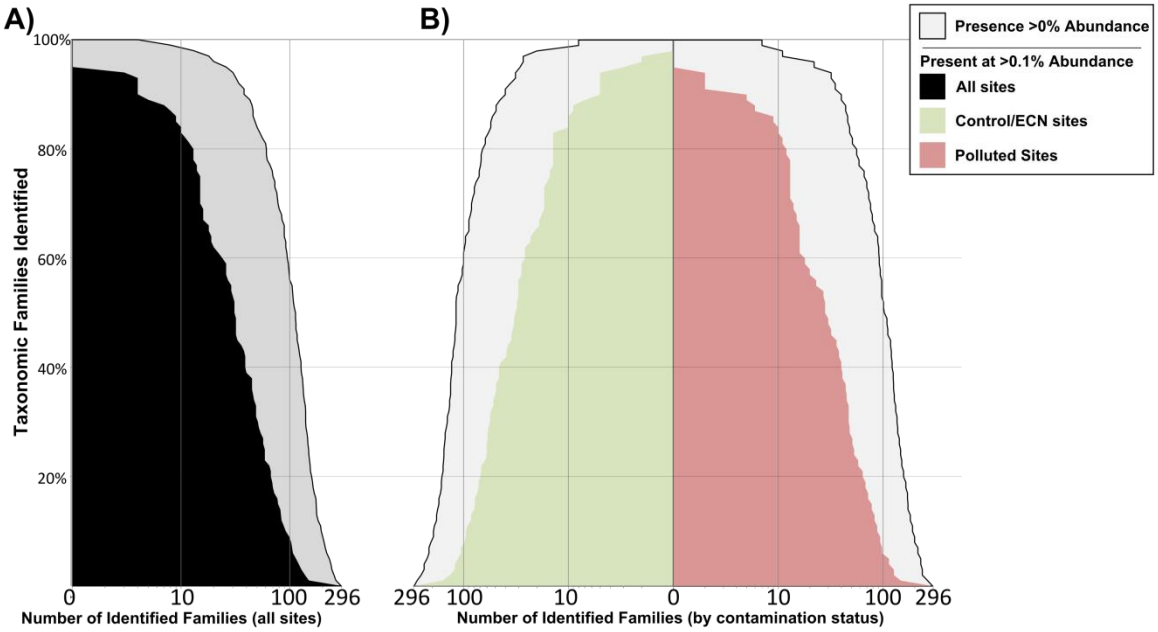


Figure 5.10 - Proportion of taxonomic families identified across the range of sites and individuals assessed. Outer shape (grey) denotes family presence at percentage of sites, inner shape (coloured) denotes only presence at >0.1% for (A) all sites and (B) control/ECN and polluted sites respectively. X-axis $\log_{10}$ scaled for clarity at low values. Only taxonomic units annotatable at family level included. Control sites demonstrated a higher number of families consistent across sites than polluted, which showed higher variability.

5.4 Discussion

This study assessed *Lumbricus rubellus* hindgut-associated microbiota from 17 geographically separated locations across the United Kingdom which encompassed a wide range of environmental stressors. Across 118 high quality deeply sequenced individuals, a core bacterial community was identified at the family level which was resistant to severe contamination. A number of taxa associated particularly with the measured increases in metal/metalloid stress on the host but conversely some bacterial families were eradicated by it. All earthworms were collected *in situ*, representing the total associated microbiota including the body symbionts, transient soil bolus and internalised microbiota allowing a realistic snapshot of the *L. rubellus*-bacterial hindgut community.

The microbiomic structure of the *L. rubellus* hindgut analysed here, demonstrated higher variation at the phylum level than has been previously observed in earthworms within a single site (Chapters 3 & 4). This indicates the potential isolationism of single-site studies and care must be taken before describing host-associated microbiota in a wider context than is under scrutiny. Observations of the earthworm microbiome at the phylum level were largely consistent with previous sequencing based studies (Chapters 3 & 4; Pass et al. 2014), as dominated by the 3 major Proteobacteria classes (Alpha, Beta, Gamma), with consistent contributions from Actinobacteria, however samples here analysed were notable for a higher proportion of Bacteroidetes than previously described. The particular assessment of the hindgut community may be the cause of this variation, indicating a localised community and would require further assessment to differentiate from the rest of the host population. There was, however, significant variation between sites and the abundance of these major phyla were not consistent i.e. Gammaproteobacteria relative abundance per site ranged from 7.7% (PTB_polluted) to 88% (CWM_polluted). This was not indicative of the polluted state of the sites, where site pairs often demonstrated high similarity. Conversely, the geographic distance between sites did not correlate, signifying the underlying environmental characteristics were more directive in establishing community structure. Lineage was not a significant determinant of the bacterial community or a factor in taxonomic family association.

As was expected, the Comamonadaceae family (which exclusively represented the *Verminephrobacter* symbiont) was universally identified in all 118 individuals although in significantly lower abundance at the DGC arsenic contaminated site as has been previously described (Chapter 3; Pass et al. 2014). This contributes to the

ubiquitous symbiont concept of the species (Lund et al. 2009; Lund et al. 2014) but again questions the essential nature when the eukaryotic host is observed living in its absence. Laboratory studies have demonstrated reduced fecundity and later arrival at sexual maturity under low nutrient conditions of aposymbiotic earthworms from the related species *Aporrectodea tuberculata* (Lund et al. 2010). However, no *in situ* studies analysing the host life history have been performed and analysis is required to determine if wild organisms suffer these failings.

Only two other families were observed in every individual; Flavobacteriaceae (Phylum: Bacteroidetes) and Mycobacteriaceae (Phylum: Actinobacteria). The Flavobacteriaceae is typical of the soil environment and is commonly observed in the rhizosphere (Johansen et al. 2009; Kolton et al. 2012). The family is highly diverse and ubiquitously chemoorganotrophes, generally adept at digesting polysaccharides and proteins; substrates which are known to be secreted into the alimentary canal by the host in the hind gut (Wüst et al. 2011). Although typically aerobic, some exist as facultative anaerobes and capnophiles which requires further exploration to clarify the pervasive abundance (here averaging 15%) in the anaerobic conditions of the earthworm gut (Campbell 2014). This has potential benefits to the host in degrading the soil surface organic matter which *L. rubellus* feeds upon but may only be representative of the transient soil community. The environmentally pervasive monophyletic Mycobacteriaceae family which was ubiquitous across all UK earthworm microbiomes relates to the mycobacterium genus which, although most renowned as a number of human pathogens (Lory 2014), is common in natural soils and waters (Iivanainen et al. 1997; Niva et al. 2006). It has previously been identified in *L. rubellus* and *L. terrestris* from individuals living in pasture fields of infected ruminants and has been noted as a potential vector for transmission of the disease although they are unlikely to be a significant factor (Fischer et al. 2003).

A number of other commonly associated taxa were identified which are key in defining the core *L. rubellus* microbiome of UK individuals and in total 37 families were found in >90% of *L. rubellus* bacterial communities. Of these however, only 6 families were present at >0.1% relative abundance per sample, indicating the wide diversity in the associated microbiota. In addition to those above, there were Cytophagaceae, Nocardioidaceae and Hyphomicrobiaceae. Although these associations are indicative of a highly varied community and certainly greatly influenced by the soil community, they are also important targets for further analysis

when determining *L. rubellus* associated species in future analyses (See Chapter 6 and Discussion).

The largest impacting environmental factor on the intra-sample diversity of UK soil microbial communities is the pH of the soil substrate (Rousk et al. 2010; Griffiths et al. 2011) however this was not observed in this study. This is likely due to the consistent circumneutral pH of the earthworm gut lumen (Wüst et al. 2011; Drake & Horn 2007) which will consistently buffer the transient community regardless of the external conditions. However, the identified ecotypes did slightly relate with differing soil pH indicating that the factor is not truly immaterial. Furthermore, a number of bacterial families associated with the pH of the external soil environment, although in combination with the additional soil characteristics. Realistically, the heterogeneity of soil makes it difficult to associate free-living invertebrates with the physiochemical measurements made due to the ability of earthworms to avoid stress (Kappler & Brune 1999). For this reason, it was more appropriate to relate the microbial community to the metal abundance calculations from the body tissue of the individuals. This allowed more accurate association due to the accurate portrayal of the earthworm interactions with contaminants. Conversely however, *L. rubellus* are known to bioaccumulate metals from the soil (Andre et al. 2010; Morgan et al. 2007) which may unfairly represent the abundance of certain accumulates in the host. A number of bacterial families identified with tissue metal abundance although co-influence of metals and other factors (e.g. pH) appears likely. *Legionella* demonstrated the strongest positive association to a measured factor; arsenic abundance in the *L. rubellus* tissue. This genus regularly inhabits protozoan free-living organisms which can act parasitically on the host (Diederen 2008) and potentially the heightened stress of extreme arsenic contamination may have weakened host viability allowing this proliferation. The Acidobacteria family of Koribacteraceae (*Candidatus Koribacter versatilis*) was one which contravened the soil pH association of the majority of the described correlating taxa, increasing abundance with higher soil acidity and Al/Fe load. However, given the generally observed effects upon Acidobacteria of the more neutral gut of the host, this is likely residual from higher abundance in ingested soils as is observed in this candidate taxa (Campbell 2014).

Due to the technological limitations of the Ion Torrent (IT) platform which was employed in this analysis, it was determined only to assess the bacterial OTUs at the family level. While this resulted in a reduced accuracy of taxonomic identification,

the increased number of individuals and the depth of sequencing which could be achieved was essential to evaluate a nationwide population such as was undertaken to a statistically accurate degree. There has been established an error frequency for the relatively novel IT platform which is yet to be modelled (Bragg et al. 2013; Quail et al. 2012), and mirrors the issues faced in the early stages for analysing data from the 454 platform (Claesson et al. 2010). While these were mitigated through 'denoising' software for 454 sequence data (Quince et al. 2009; Quince et al. 2011), platform-specific error correction methods were not available for IT data and more generic error mitigation was here performed using the UPARSE pipeline. Despite employing UPARSE as the highest quality control standard available, the combination of uncertain error rate and the relatively short amplicon (150bp after quality truncation versus ~550bp with 454 FLX+ platform) deemed it appropriate to maintain taxonomic identification at the family level so as not to false-report on genera or species through misidentification.

5.5 Conclusion

A core bacterial community appears at the family taxonomic level within the *Lumbricus rubellus* hindgut sourced from a range of UK controlled and polluted sites. Although likely originating principally from the soil community, previous studies indicate particularly the increased earthworm abundance of Actinobacteria, of which the Mycobacteriaceae family has been consistently identified throughout polluted and control sites. Flavobacteriaceae (Bacteroidetes) were also present alongside Mycobacteriaceae and in greater abundance than previous whole-worm assessments, perhaps relating to the hindgut focus of this study. Higher resolution of these groups is essential to understanding the association with *L. rubellus*.

Distinct groupings (ecotypes) appeared which differentiated *L. rubellus* microbiomes at the whole community level, and the impact of metal contamination was key in differentiating earthworm populations. However soil moisture and organic content were driving forces in diversity and richness likely because of their impact upon bacterial growth, more so than the surrounding soil pH.

6

The Voxel Worm: Spatial Characterisation of the Earthworm microbiome

6.1 Chapter Introduction

The microbial communities associated with the earthworm have been shown as highly concurrent between individuals from different sites and species yet distinct from the surrounding soil environment, and consist predominantly of Proteobacteria and Actinobacteria. Also regularly observed at lower abundance levels are Bacteroidetes, Acidobacteria, Firmicutes, Chloroflexi and Cyanobacteria (Chapters 3, 4 & 5). To date, high throughput sequencing of earthworm microbial communities has only been performed on whole, *in situ,* earthworms and no analysis has been performed on the host-associated community in the absence of the soil bolus. Furthermore, there has been indication of species localisation within the earthworm gut both through classic techniques (Jolly et al. 1993; Byzov et al. 2009; Thakuria et al. 2010) and now from amplicon sequencing analysis of the *L. rubellus* hindgut (Chapter 3).

The most extensive illustration of microbial localisation within an individual species has been produced as the Human Microbiome Project (HMP) (The Human Microbiome Consortium 2012). Here, through profiling of different sites within the Human body, microbially distinct communities and the expression pathways associated with particular organs and body locations has been revealed. Spatial microbial localisation is less studied in invertebrates although the rewards for localisation of functionally rewarding microbial species still apply. Most noted are the specialised hindgut communities observed in species such as termites (Warnecke et al. 2007), beetles (Egert et al. 2003; Egert et al. 2005), and gut symbionts in *ocedax* polychaetes (Verna et al. 2010; Goffredi et al. 2007). A high level of heterogeneity can be observed in microbial communities in small spatial areas for example, as is frequently observed in soils (Nunan et al. 2002; Ettema & Wardle 2002; Roesch et al. 2007). Investigation of the interactions between transient and associated bacteria are key but must also be accompanied by understanding the localisation of bacteria to particular body areas, tissues and organs.

A number of bacterial species are known to be vertically transmitted symbionts, being passed through generations via the cocoons of various animal species. The *Verminephrobacter* symbiont of the earthworm is actively transferred to the cocoon through specialist organs and is hence after isolated in the second loop of the nephridia (Davidson & Stahl 2008; Dulla et al. 2012). This symbiont is known to benefit the host especially in low nutrient environments and its absence reduces fitness and fecundity (Lund et al. 2010; Lund et al. 2014).

This chapter describes the high resolution microbiomic profiling of *L. rubellus* with the intent to identify communities associated with body organs and tissues and to establish the impact upon the transient community. Voxel (volumetric pixel) refers to the smallest possible three-dimensional unit of a sample, and through 'voxelation' of the whole earthworm there is the potential to understand the spatial localisation of microbial species to an unprecedented scale.

6.2 Methodological Approach

6.2.1 Site Description and Experimental Design

Earthworms and soil samples were obtained from Dinas Powys a previously studied site absent of significant contamination or stress effects (Corp & Morgan 1991) (Centre: Latitude: 51.4412852, Longitude: -3.2325891). Earthworms were originally visually identified as *Lumbricus rubellus* followed by confirmation with COI barcode sequencing for both individuals and cocoons (Chapter 2.1.2.4.3). Replicate soil samples were collected from the epigeic level (Chapter 2.1.1.2). Additional collected soil was dried at 80ºC overnight during which all earthworm individuals were depurated on wet filter paper. Soil was sieved (1mm), rehydrated, and divided into three containers (10x20x10 cm) to which 5 adult *L. rubellus* were added respectively. After 24 hours organic manure was sieved (1 mm) and 1 cm was added to the surface of the soil containers. Mesocosms were incubated at 14ºC for 28 days. On days 14 and 21 earthworms were removed and soils were hand sieved for cocoons which were kept in distilled H_20 at 4ºC, after which the earthworms were returned. On day 28, 5 individuals were flash frozen and stored in RNAlater ICE (Life Technologies, UK), whilst 5 were depurated for 72 hours individually before undergoing the same process. Triplicate post-experimental soil samples were frozen until required and casts were collected daily from the depurating individuals.

Only individuals of *L. rubellus* Lineage A were analysed and determined on the basis of COI sequencing (Chapter 2.1.2.4.3). Fresh and depurated individuals were preserved in RNAlater ICE prior to dissection which generated a sequential series of 24 anatomically defined whole segment assemblages (SA1-SA12, Gut (g) or Body Wall (b)) to the specification presented in Figure 6.1. Cocoons were surface-sterilised by submersion in 12% NaOCl for 45 seconds followed by 95% EtOH for 5 minutes and thoroughly washed with double distilled H_20 as described in Coelom et al. (2012). DNA extraction for all samples was performed to the modified Qiagen Blood and Tissue protocol (Chapter 2.1.2.3.1), however, a precursor step of 10% v/v Proteinase K (56ºC for 20 minutes) was performed to initially overcome the RNAlater ice preservative process before starting the bead beating protocol.

PCR amplification of the 357F-907R amplicon region of 16S rRNA was performed with unique barcodes included in primer design for post-sequencing sample identification followed by submission of 266 samples plus 53 blank controls (randomly interspersed during PCR) for Illumina MiSeq 2x 300 paired end sequencing (Chapter 2.1.3.3) returned approximately 17.1 million reads.

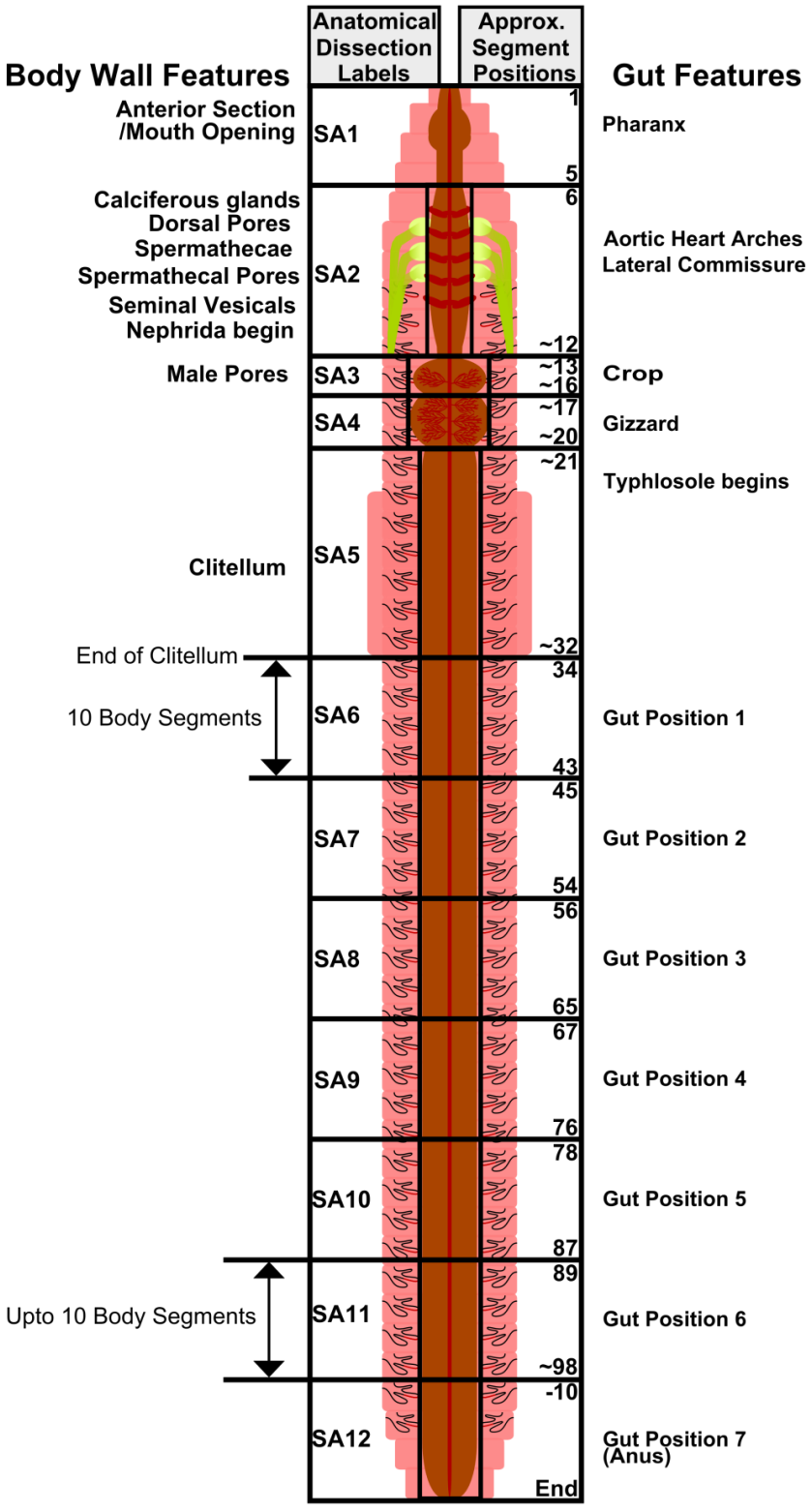


Figure 6.1 - Schematic representation of *L. rubellus* showing voxelation dissection positions. If internal organs varied from canonical position, segment number was not adhered to and the organ was extracted intact (only relevant at positions denoted by tilde (~)). The posterior section (SA12) was always 10 segments and SA11 allowed natural variance. Diagram not to scale. Source: (Sims & Gerard 1985)

6.2.2 Bioinformatic Methodology

The informatics processing steps performed are described in detail with source references in the Chapter 2 and sample descriptions in Appendix 1d. Approximately 17.1 million reads were produced a single Miseq sequencing run with 8.65 million being retained after quality filtering. Each amplicon sample was subsampled to 5,000 reads which was the most common minimum level while retaining three replicates of all samples. A brief overview of the steps taken follows:

Stage	Chapter	Software	Description
Input Data	2.1.3.3	MiSeq 2x 300 paired end sequencing	17.1 million Forward 17.8 million Reverse
Merging and Quality Filtering	2.2.1.2	USEARCH	16.0 million joined. Minimum length 400bp Max expected errors = 5 (Note: 3 failed flow positions were observed in all reads and accounted for) Singletons were removed
Denoising	2.2.1.1	NA	Not applicable to miseq data
Contamination Filtering	2.2.1.2	NA	No significant contamination detected across 53 Blank samples
Subsampling (normalisation)	2.2.2	Qiime	Randomly subsample to minimum sample sequence count
OTU generation	2.2.2	UCLUST	OTUs were generated at 0.97
Taxonomic Annotation	2.2.2	UCLUST & BLAST, Greengenes database	UCLUST annotation of OTUs identical to reference OTUS. Blast annotation at <0.001 E value for novel OTUs
Phylogenetic Alignment	2.2.2	FastTree	For downstream use in diversity analysis
Alpha Diversity Analysis	2.2.3	Qiime	Shannon, Chao1, Observed Species
Beta Diversity Analysis	2.2.3	Qiime/Unifrac	Jackknifed UPGMA analysis
Functional Inference Analysis	2.2.3	Picrust, KEGG	Identification of changes in functional gene/pathway abundances

6.3 Results

6.3.1 Changes to Diversity and Richness along the Gut Transect

During transit along the length of the host gut an initial substantial reduction in diversity and richness was observed (Figure 6.2). In the gut of Fresh (*in situ*) individuals, indices then increased along the gut length to more closely represent those observed in the surrounding soil. The gut of depurated individuals demonstrated higher diversity and richness in the first half of the earthworm body, indicative of the influence of transient bacterial communities on the total community structure. At SA10 (Figure 6.2) there was a significant decrease in measured diversity and richness, and additionally changes to taxonomic structure (Figure 6.3) which may relate to the opening of the typhlosole specialised gut section into the main gut lumen.

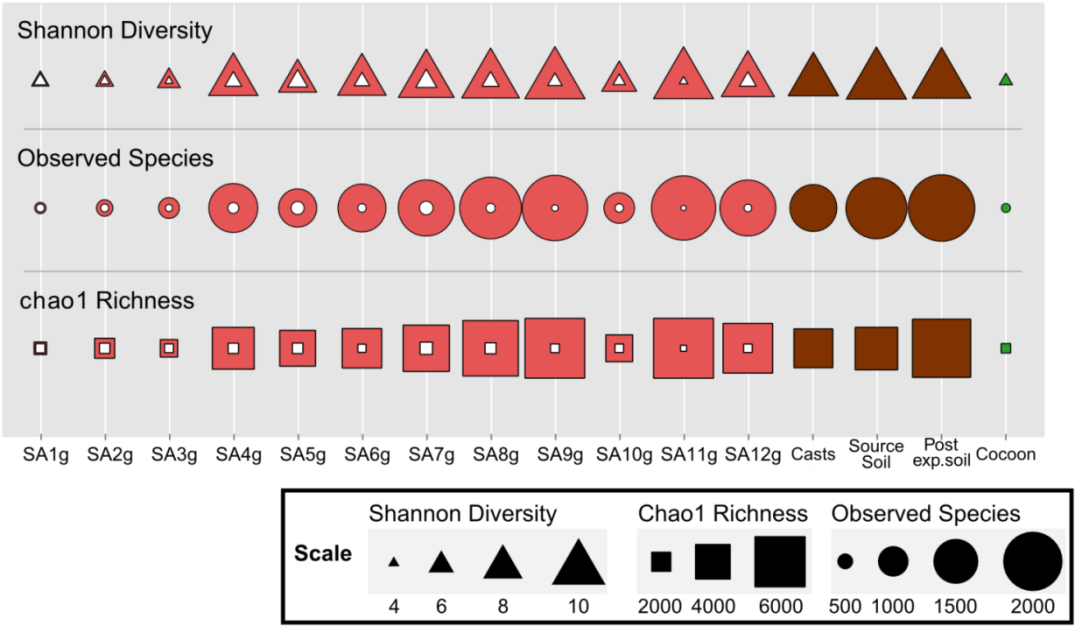


Figure 6.2 - Pictographic representation of diversity and richness estimates along the earthworm gut alongside soil and cast indices. Earthworm samples labelled with body position (outer shape: fresh, inner white: depurated). Metrics rapidly reduce in the anterior sections of the earthworm when compared to soils but rises in the gut-proper.

6.3.2 Positional differences in Phylum level structure

Changes in microbial community can be observed along the earthworm length both in the gut and body components with distinct variation being observed for fresh and depurated samples (Figures 6.3a, b, c & d). In each case, there is a distinct difference between the pre-clitellum anterior ('head') section of the earthworm, and the longer body portion which houses the remaining gut. Notable is the proportion of OTUs which were unable to be annotated, indicating a number of novel species.

6.3.2.1 Gut Community

At the phylum level, the fresh gut community from the gut-proper is consistent through to the cast material which the earthworms egest and is highly similar to that of the soil community. After 30 days of earthworm activity on their native (albeit initially oven-dried) soil, the community consisted predominantly of Proteobacteria (37.5%), Actinobacteria (15.2%), Firmicutes (4.4%), Chloroflexi (10.3%) and Acidobacteria (15.2%). This closely resembles the structure observed in the source soil prior to drying, and the intestinal region of the gut of fresh earthworms. In the depurated gut of earthworms, there is a significant reduction in Actinobacteria, Firmicutes, Chloroflexi and Acidobacteria in comparison to fresh gut contents, demonstrating the adhesion to the gut wall of Proteobacteria, Spirochaetes and Ternicutes by the microbial community. The relative abundance of Gammaproteobacteria is high in the early region of both depurated and fresh earthworm guts, and is greatest in SA2g and SA3g (Highest Fresh abundance: 47.2%, Depurated: 69.5%). In the fresh earthworm this is rapidly overcome in the intestinal gut (SA4g onwards) resulting in low relative abundance whereas there is a smaller reduction in abundance along the length of depurated worms (Figure 6.4b). Notable is the resurgence at SA10g, coinciding with the reduced diversity and richness in this section (Figure 6.2).

6.3.2.2 Body Community

The bacterial community of the earthworm body is dominated by the *Verminephrobacter* symbiont. The Betaproteobacteria fraction of the earthworm microbial community (Figures 6.3c and 6.3d) is >99% composed of the Comamonadaceae family (of which *Verminephrobacter* is a constituent). Although taxonomically diverged from reference sequences from other Lumbricidae, it is closely related to uncultured *Verminephrobacter* sequences obtained from other *L. rubellus* sequencing efforts (note, the symbiont of *L. rubellus* specifically has not been previously characterised).

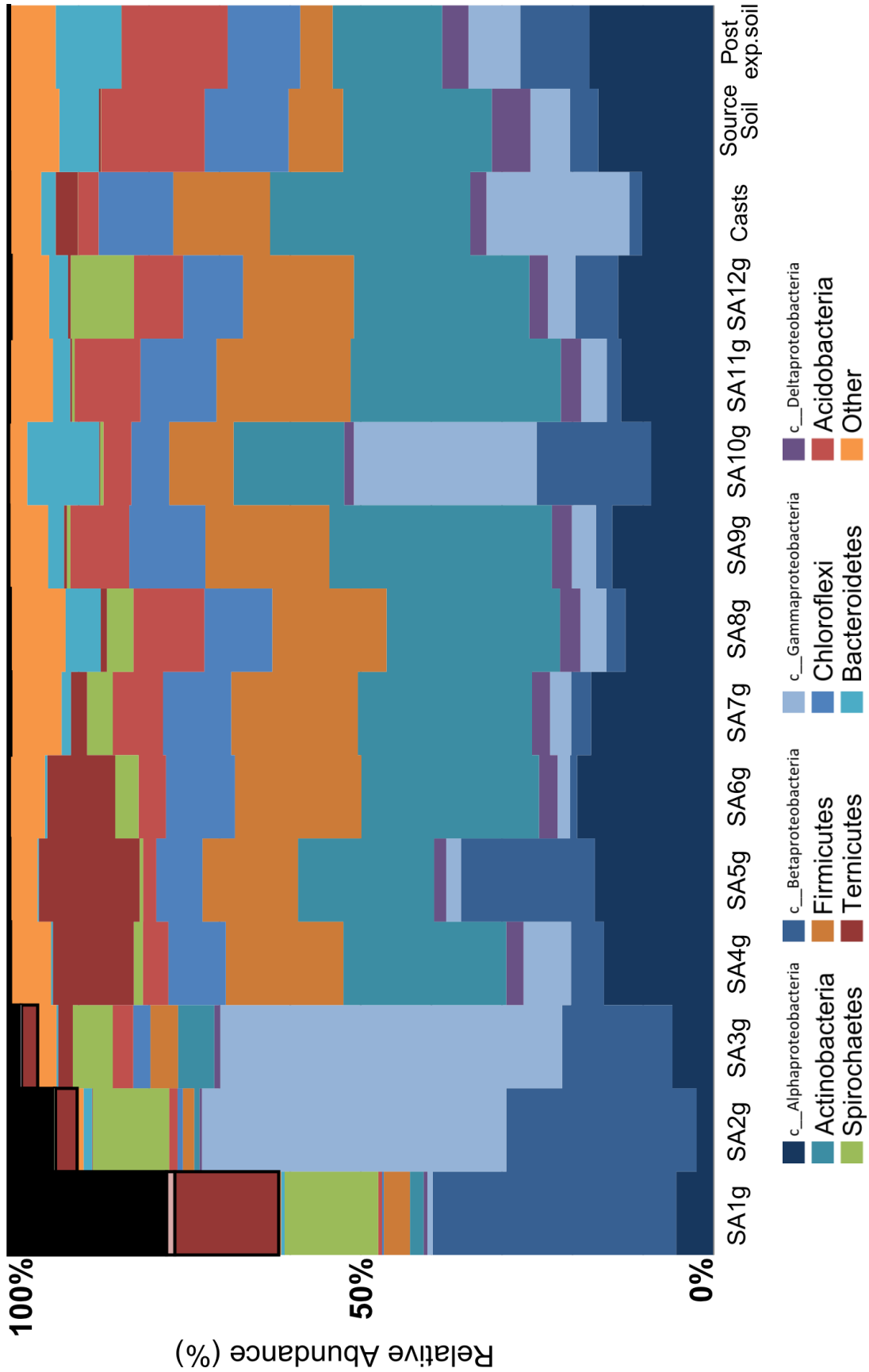


Figure 6.3 - **(a) Relative phyla abundance of the fresh earthworm gut. Proteobacteria classes individually represented (blue). Full black areas are unknown species. Black-boxed areas are manually annotated species which were beyond accurate identification.**

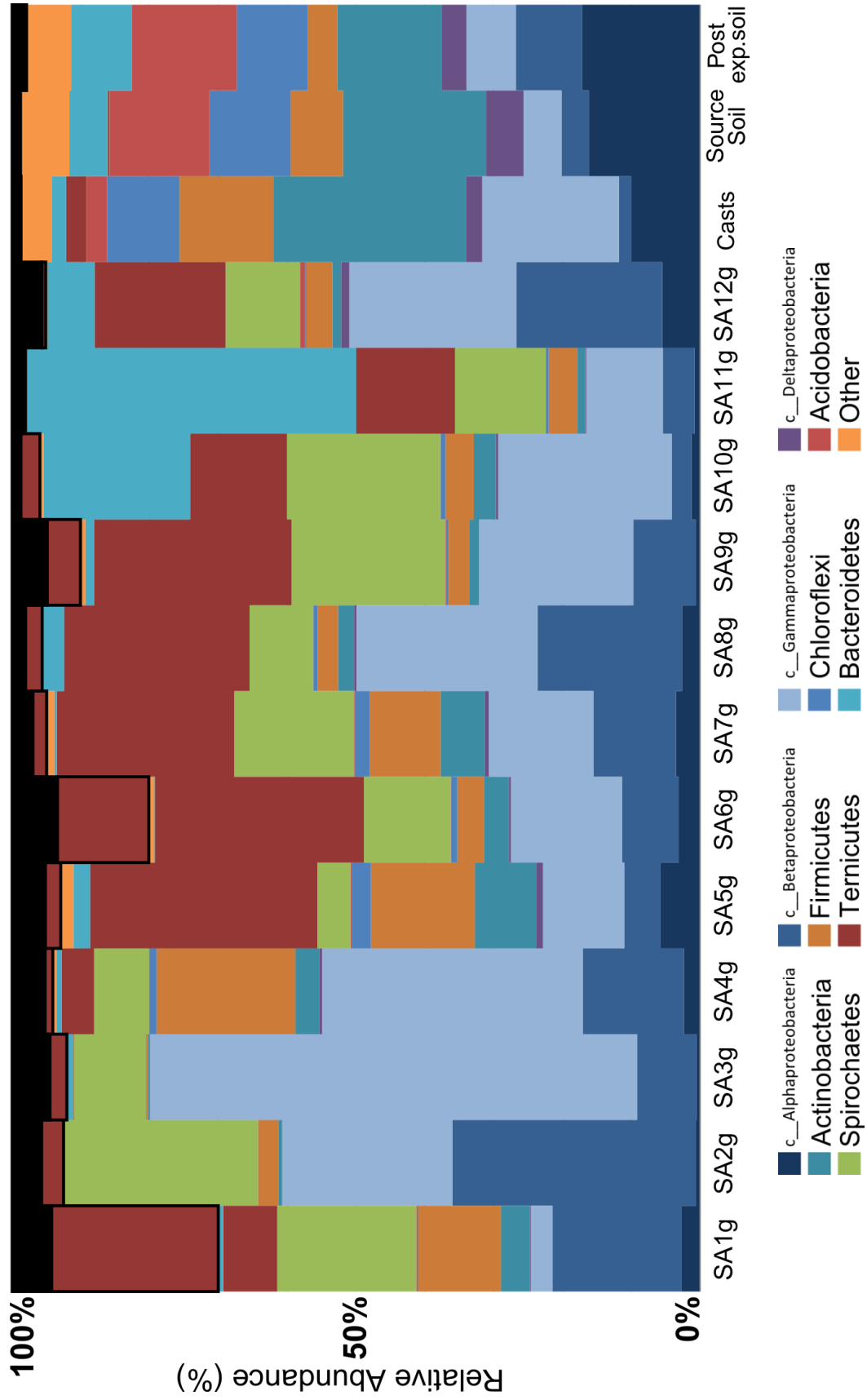


Figure 6.3 (b) Relative phyla abundance of the depurated earthworm gut. Proteobacteria classes individually represented (blue). Full black areas are unknown species. Black-boxed areas are manually annotated species which were beyond identification.

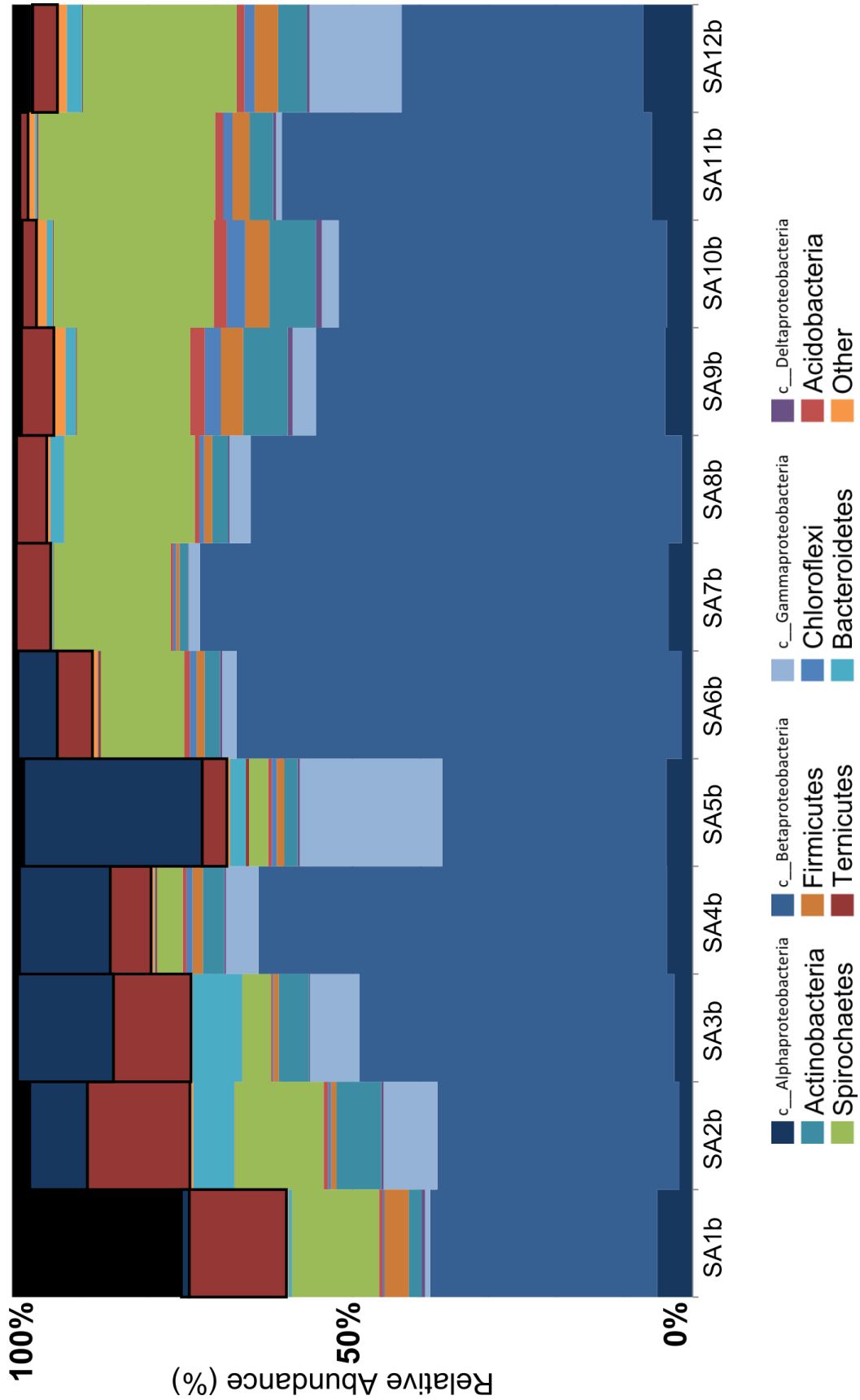


Figure 6.3 I Relative phyla abundance of the fresh earthworm body. Proteobacteria classes individually represented (blue). Full black areas are unknown species. Black-boxed areas are manually annotated species which were beyond identification.

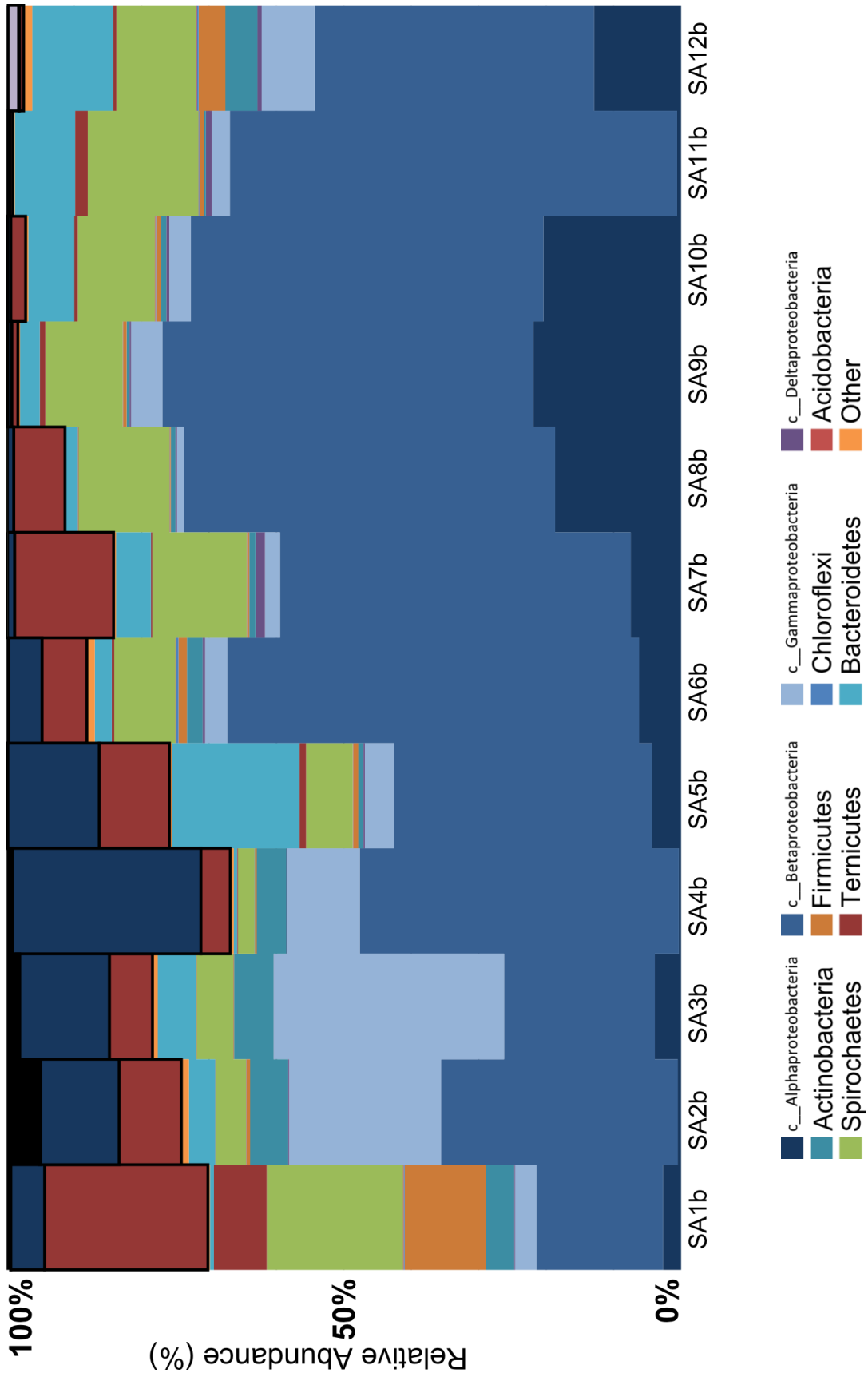


Figure 6.3 (d) Relative phyla abundance of the depurated earthworm body. Proteobacteria classes individually represented (blue). Full black areas are unknown species. Black-boxed areas are manually annotated species which were beyond identification.

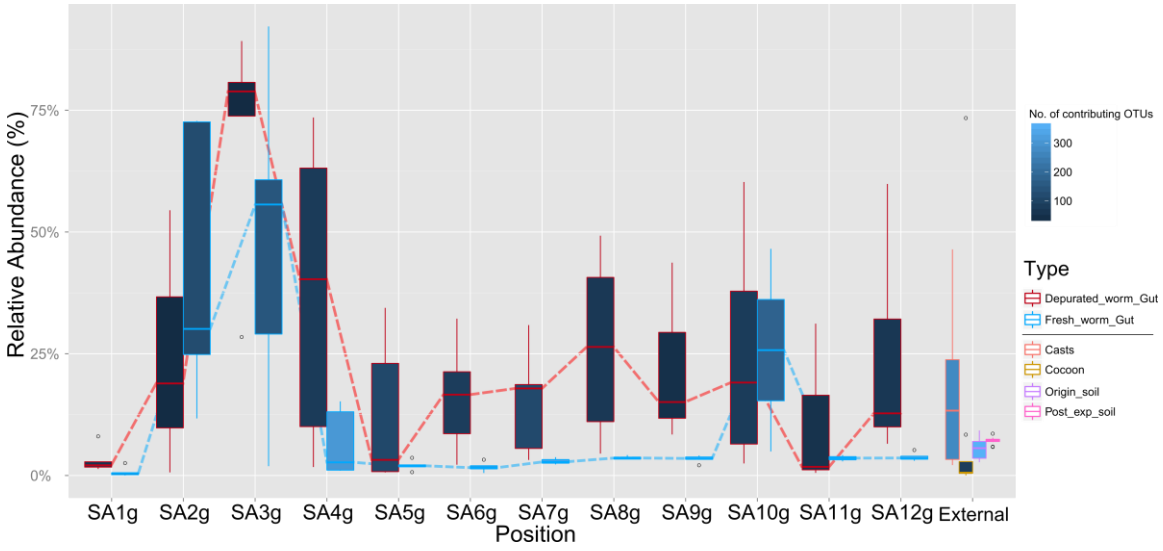


Figure 6.4 - Gammaproteobacteria abundance in the gut. The class was enriched in the anterior sections of the gut but rapidly decreases to below soil levels in the main gut. A localised enrichment was observed at section 10 which coincides with diversity measures at this position and may relate to the termination of the typhlosole.

6.3.3 Spatial positioning of key abundant OTUs

Distribution of numerous OTUs were found to be specific to particular areas or tissues within the earthworm host. Some OTUs such as *Verminephrobacter* varied in proportion but this was likely due to reported abundances being proportional to the community and highlights a potential disadvantage of this approach.

A distinctly abundant OTU associated specifically with the crop organ (position 3) in the gut samples (15.5% fresh gut; 44.8% depurated gut), indicating a gut-wall association with the host beyond the transient material (Figure 6.5). There was also observed a localised abundance in the body samples of the crop area, potentially representing a reservoir in the body wall for this species. This Enterobacterales (Family) species could not be more accurately defined however the high abundance and unambiguous body positioning is of key interest in exploring digestive mutualism in the earthworm.

The presence of Spirochaetes is observable throughout earthworm samples, but most prevalent in the body tissues (is absent from soil samples, Figures 6.3). This phyla which accounts for up to 24.9% in some samples exclusively contains the genera *Borrelia*, of which a number of species spread via tick-borne parasite implicating *L. rubellus* as a possible transmission vector for Borreliosis disease. Species level characterisation was not possible with the sequence data produced, however, the OTU distinctly associated phylogenetically with the clade *Borrelia* (Figure 6.6).

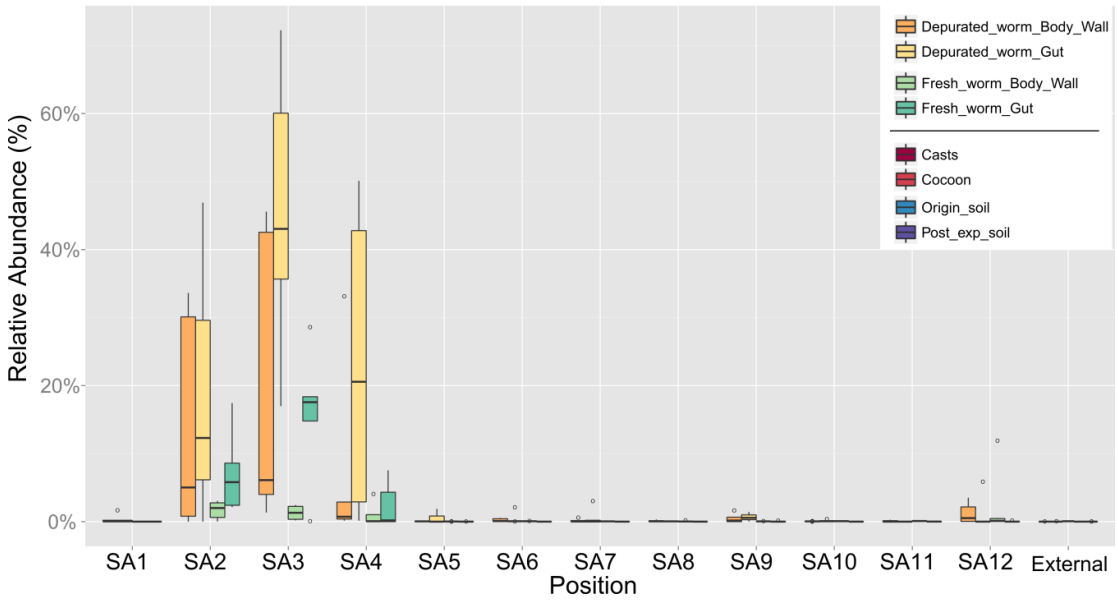


Figure 6.5 - Spatial distribution of novel Enterobacteriaceae OTU. Most abundant in the gut of the earthworm, the abundance was higher in depurated individuals, indicating gut wall adherence.

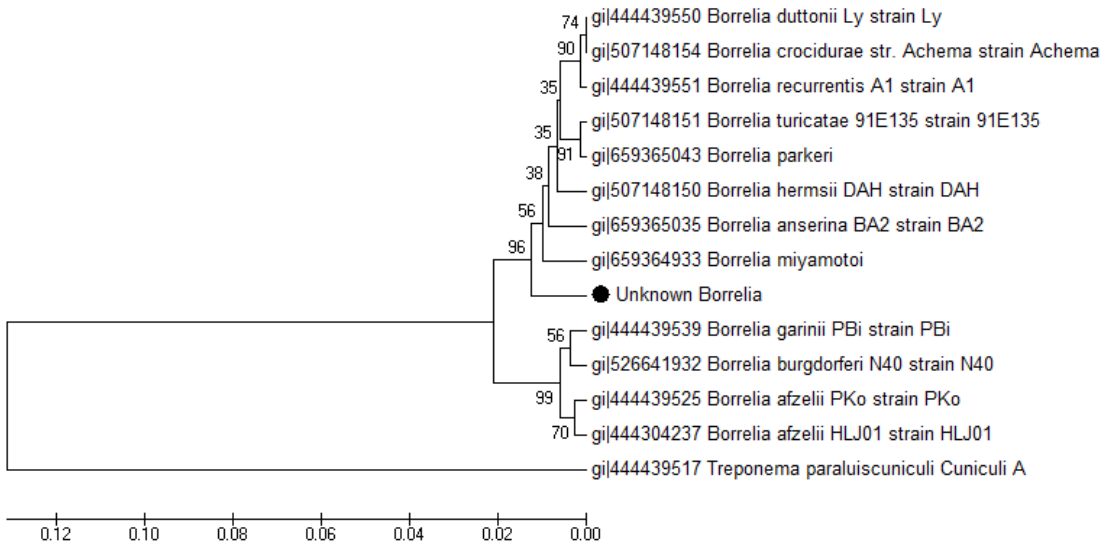


Figure 6.6 - Phylogenetic position of novel *Borrelia* species. Reference sequences obtained from cultured refseq species. Tree generated with UPGMA method (1000 bootstraps). The novel sequence nested within the genus but varied from reference sequences at this amplicon region.

A high proportion of the microbiota in the anterior sections of the body was below annotation parameters but when assessed phylogenetically up to ~25% of some fresh samples could be identified as Alphaproteobacteria (Figure 6.3). This was a single highly abundant OTU which was taxonomically identified as Rickettsiales (order) but no closely related references were available. The positioning of the species was highly localised to the reproductive areas of the host i.e. clitellum and spermathecal pores/seminal vesicles (Figure 6.7).

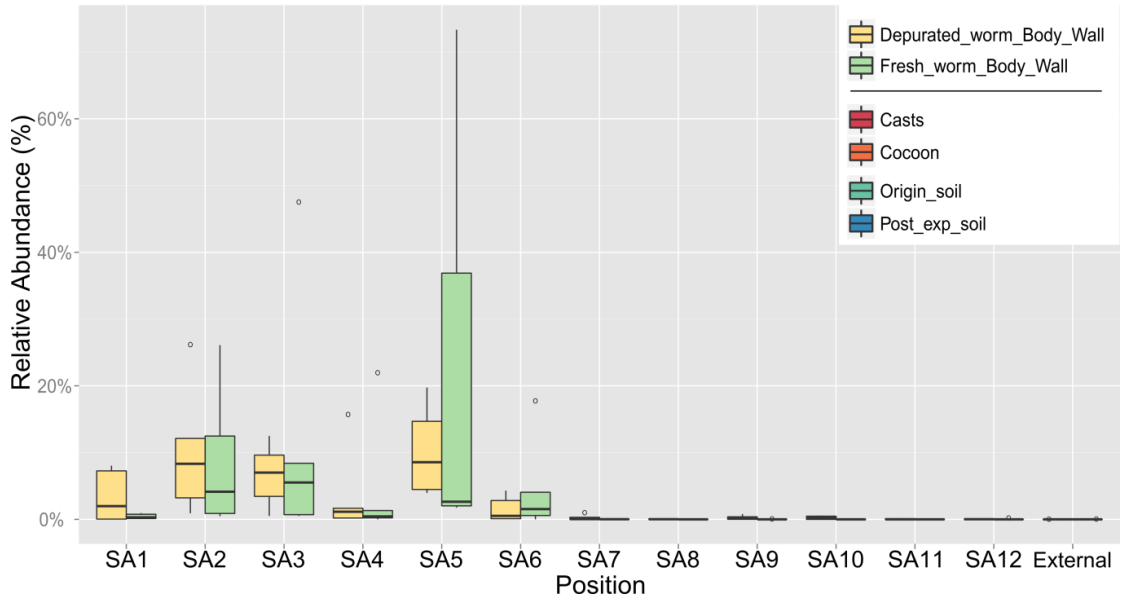


Figure 6.7 - Spatial abundance of the novel Rickettsiales species shows localisation to the reproductive body organs in both fresh and depurated individuals. High abundance at clitellum and seminal vesicals may associate with the body tissues, but more likely is hyperparasitism through the Monocysitis agillis earthworm parasite.

6.3.4 Longitudinal variation in functional capacitive communities

Given the major changes along the length of the earthworm in community structure, there was the potential for variation in the functional capacity in the microbial metagenome. The abundance of functional pathways in the sample communities was inferred from the species abundance using the picrust software package (Langille et al. 2013). Methane metabolism pathway genes along the length of the earthworm gut were higher in fresh individuals than depurated, indicating association of the pathway to the transient soil community, (Figure 6.8). Approximately 1% of genes in the metagenome gut community formed part of the pathway. The associated proportion increased in abundance towards the hindgut and was represented equally in the casts, but was higher than the post experimental soil community. Large variation was found in SA5g (post-gizzard) perhaps representing transition into the gut-proper, and SA10g (hindgut) which corresponds with the non-concurrent diversity and richness measured in this sample.

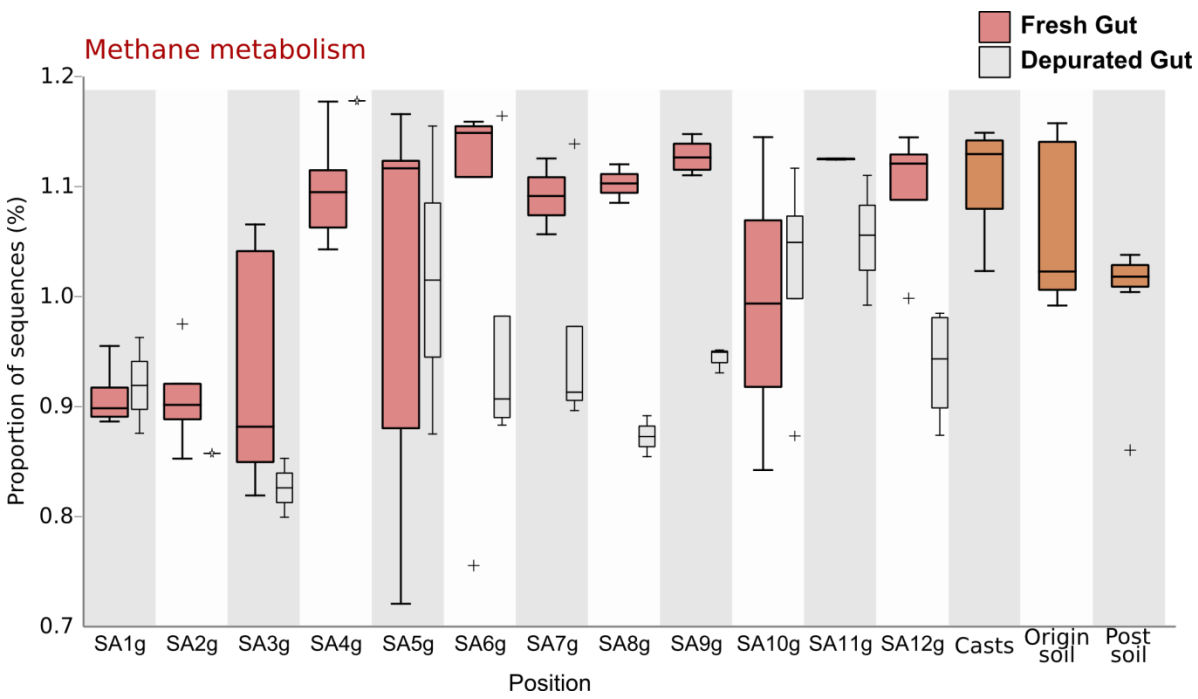


Figure 6.8 - Predicted KEGG pathway relative abundance for methane metabolism in the *L. rubellus* gut. Red: Fresh; Grey: Depurated; Brown: Soil material. Species capable of these mechanisms were highest in the post-clitellum fresh gut which was consistent in the gut but reduced in the post experimental soil community.

The potential for microbially-mediated biotic and xenobiotic degradation in the earthworm has previously been noted in Chapter 4 but not explored beyond whole-individual. These suggested pathways demonstrated higher association to the body wall tissues of *L. rubellus* rather than in the transient or gut wall associated community (Figure 6.9), which supports the hypothesis of a beneficial microbial functionality. A subdivision in several major degradation pathways can be observed which associate with particular localisations in the host. Higher abundance of genes involved in the KEGG benzoate degradation pathway were apparent in the early midgut region in fresh individuals, although depurated organisms had a much higher level of variation (Figure 6.10).

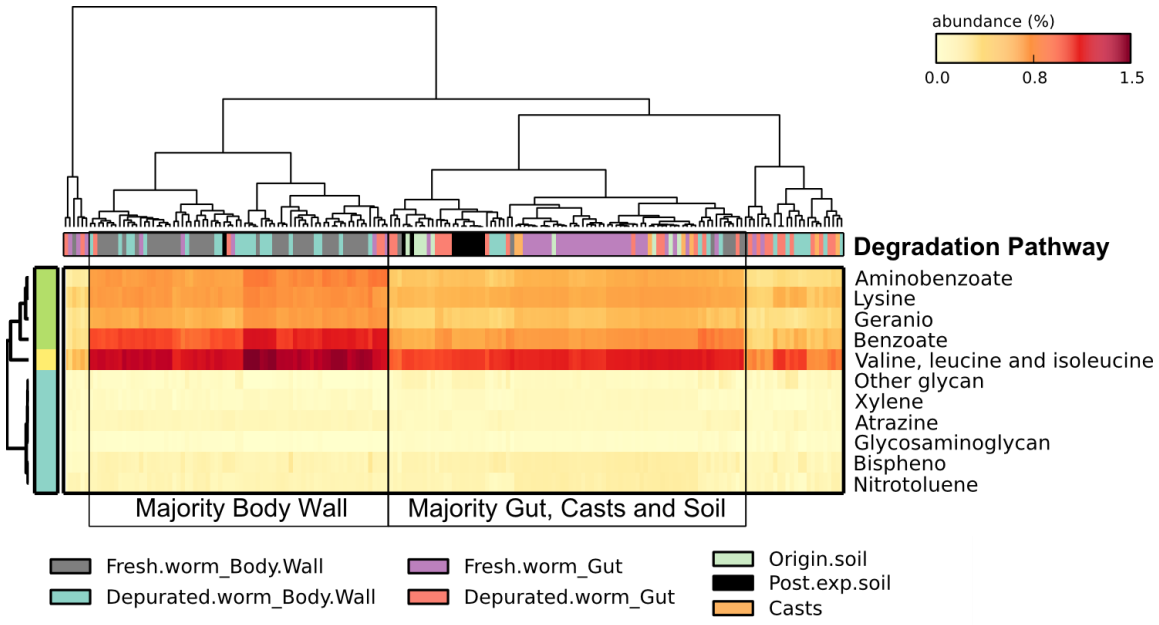


Figure 6.9 - Differential abundance of major predicted KEGG degradation pathways. UPGMA clustering (0.75 association) separated body wall samples distinctly from other samples, demonstrating the higher abundance of species with degradation capabilities than the gut and soil communities. Type association boxes were manually assigned to significantly branching clades.

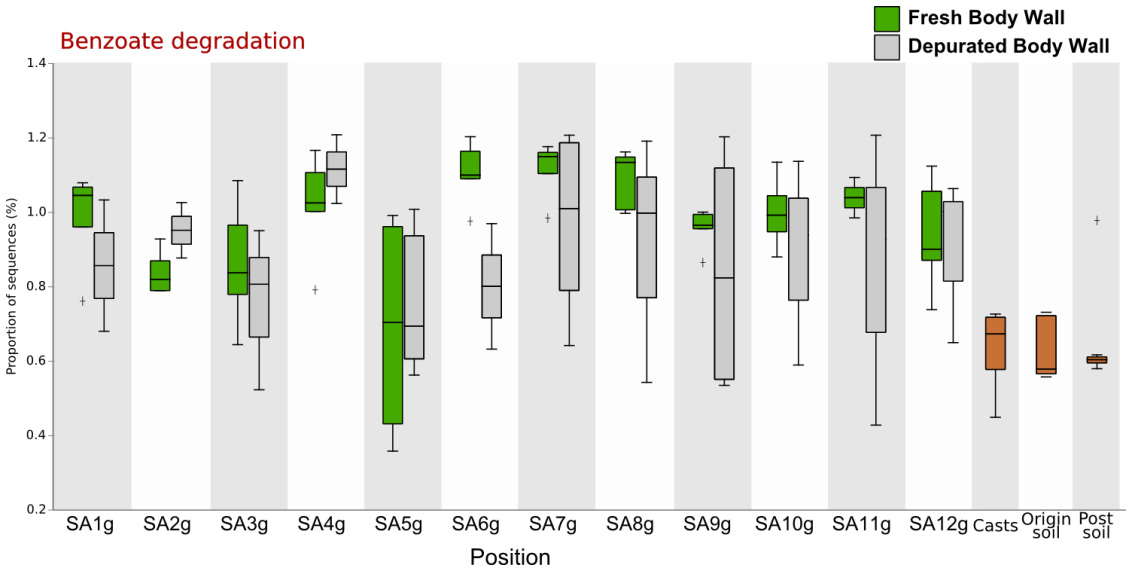


Figure 6.10 - Predicted KEGG pathway relative abundance for Benzoate degradation localisation in the Earthworm Body Wall. Genes encoding this pathway are predicted at a higher abundance than the external cast/soil material or the gut bolus.

6.3.5 Vertical Bacterial Transmission

Obligate symbiosis is commonly observed in bacterial species that are passed on through vertical transmission which, in the case of the earthworm is achieved via the cocoons of the host. Analysis of cocoons generated during the 30 day incubation was performed to investigate the community transmitted from parent to offspring to further understand the species which associate with the earthworm itself.

The relative proportion of internal cocoon bacterial species was compared to the host-associated community to identify vertically transmitted species however, only the *Verminephrobacter* symbiont was also identified in the adult *L. rubellus* at significant abundance (Figure 6.11). The cocoon microbiome comprised 12% *Verminephrobacter* which was the second most abundant after the highly abundant (29.5%) *Pedobacter* OTU (Family: Sphingobacteriaceae). *Pedobacter* was not observed throughout the adult earthworm samples in a high abundance, with its greatest presence accounting for 1.1% in depurated SA3 (which includes the male seminal pores), but minor abundance (<0.5%) in all other samples.

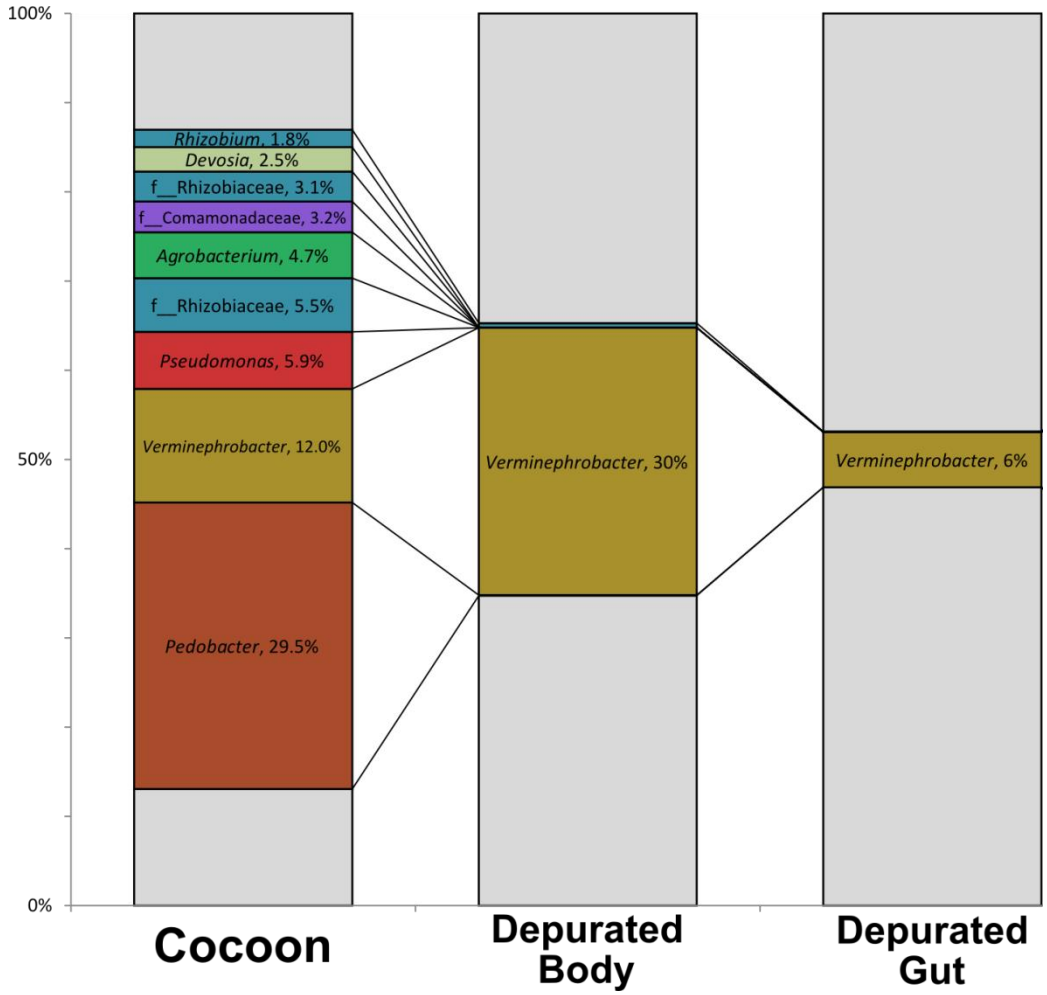


Figure 6.11 - Major OTUs observed in the cocoon present at >1.5% relative abundance and their abundance in the body and gut of the earthworm. Grey sections account for the remainder of the community. Only the *Verminephrobacter* symbiont was present in the adult *L. rubellus* community at high abundance. This was localised in the body as expected, but an amount also was detectable in the gut community.

6.4 Discussion

This analysis is believed to be the highest resolution study of microbial communities within a single eukaryotic species to date, with each individual divided and assessed at 24 spatially distinct body locations in both natural (*in situ*) and depurated conditions. In a species which has previously been described as unlikely to have an indigenous microbial community (Egert et al. 2004), the results presented strongly support that a distinct microbiome associates with the host, and that the earthworm has a definitive effect on the transient microbial populations as well as retaining a community associated with its own body tissues.

The major phylum abundances were found to be in agreement with previous sequence based analysis of earthworm microbiomes (Chapter 3, Pass et al. 2014) when observed at whole-individual level (predominantly of Proteobacteria, Actinobacteria, Firmicutes, Chloroflexi and Acidobacteria). The dominance of the *Verminephrobacter* symbiont in the body-wall communities was also as expected (Lund et al. 2009; Lund et al. 2010). However, it was clear at the highest taxonomic level there are significant tissue- and organ-specific bacterial communities beyond those which have previously been described, and that the presence of the gut bolus vastly over-represents the chronically-associated microbiota. Despite the highly soil-similar gut community, there was observed a distinct profile in the anterior sections of the host in depurated and fresh individuals. Also notably, the profile of the early gut and hindgut (section 10) were similar which may relate to the opening of the typhlosole (a specialised discrete digestive tract that runs along the length of the gut) which typically opens into the hindgut at approximately ¾ length of the earthworm body (Edwards 2004). Species localisation was particularly evident in the crop and gizzard segments which hosted a highly abundant Enterobacteriaceae species which could not be annotated more accurately due to its absence from reference databases, highlighting its novelty. Identification of such species and the varied spatial localisation of high level taxonomic groups makes it clear that whole-organism microbial analysis can be ineffective and potentially misleading when attempting to describe the association of bacteria with an eukaryotic host. A taxon which is 50% of total abundance of the crop organ in fresh *in situ* earthworms cannot be observed in a whole-organism sequencing analysis because of the relative minority overall, however the biological connotations of this abundance are profound and should not be dismissed. The inherent bias of all High Throughput Sequencing methods must also be noted, in that all samples are equally represented regardless of sequence abundance (i.e. scaled to 100%). It is therefore dangerous to suggest

that a body wall sample displaying particular dominant taxa is of more phenotypic importance when in actuality the overall bacterial abundance may be significantly lower than a fresh gut sample, and has minor interactions with the relevant 'section'. To enable a more accurate description of the localised communities, quantitative measurements would greatly augment the results here presented.

In the gut there was observed a distinct difference between fresh and depurated profiles, where the fresh gut represented a community intersecting the soil and the depurated gut. This supports the theory that the earthworm has gut-wall associated species which are not purely representative of the soil. Depurated gut samples are less populated by the Alphaproteobacteria, Actinobacteria, Firmicutes and Chloroflexi which form a major component (>50%) of fresh gut and soil samples. Ternicutes, Gammaproteobacteria and Bacteroidetes retain the highest proportion through adherence to the *L. rubellus* gut wall and are present in the absence of the soil bolus. The Gammaproteobacteria (Enterobacteriaceae), which were in highest abundance in the crop/gizzard organs, were maintained throughout the gut and remained in the depurated individuals, indicating a strong association with the host. Although accurate identification was not possible due to the phylogenetic distance from reference databases, these species, amongst others, are targets for further study of earthworm gut symbionts through quantification, cultivation and accurate mapping to determine whether they are truly associated with the gut wall and their potential functional role. Enterobacteriaceae have previously been noted as a 'fermenter' species in the related earthworm species *L. terrestris* (Wüst et al. 2011) and the taxon contains a number of symbionts found in insect species e.g. *Serratia symbotica* (Lamelas et al. 2011; Weiss et al. 2012) which has been postulated in *L. rubellus* previously (Pass et al. 2014).

The body wall tissues showed little differentiation between depurated and fresh earthworms indicating their isolation from the environmental conditions of the gut passage. The nephridia is known to be a closed organ (Davidson & Stahl 2008) and the *Verminephrobacter* symbiont represented >98% of the Betaproteobacteria which dominated these tissues. The remaining community here observed may be contained within this organ or be associated with the other body tissues such as the spermathacae, seminal vesicles, clitellum or coelom (excretory fluid containing organ which runs adjacent to the gut or the reproductive organs in the anterior sections) however, manual dissection could not distinguish between tissue specific microbiomes and a higher spatial resolution techniques, such as *in situ* hybridisation,

would be needed to resolve the precise cellular association of the microbial community.

The identified Ricketssiales species was highly surprising due to the highly specific localisation and the taxonomic separation from reference sequences. A number of other species from this order (e.g. the renowned *Wolbachia* genus) have been described as symbionts/parasites in insect species (Teixeira et al. 2008; Moya et al. 2008; Coelom et al. 2012) which may elucidate on the co-occurrence with the earthworm. The actual mechanism behind this association may be in the context of the common earthworm parasite *Monocystis* sp. which infects the sexual organs of the host, although transmits through the soil rather than sexually (Field & Michiels 2006; Field & Michiels 2005). The proliferation areas for this protozoan parasite are concomitant with the identified Rickettsiales species and the *Monocystis* may represent a more realistic host for the novel species. The potential for hyper-parasitism or symbiotic contribution requires further investigation to understand this complex system.

Although there are multiple targets that may confer beneficial effects to the earthworm host or the environmental services which it is noted for, the high abundance of a *Borrelia* genus related OTU raises implications for *L. rubellus* acting as a transmission vector for a tick-borne borreliosis-like disease which the genus is frequently associated with. As has been identified above, the abundance of this species is apparent only due to the high resolution sequencing analysis that has been performed, as a whole-organism approach would render this a minor component. Whether the disease is prevalent in widespread *L. rubellus* and/or other earthworm species is of key interest and requires further study, however there is the potential for this infection to have disproportionately proliferated due to the close proximity of individuals during the incubation phase of the study.

It is evident that the internalised earthworm microbial community is distinctly different to that of the surrounding soil. The taxonomic profile of the mid to hind-gut is highly similar to the surrounding soil and to the cast material which it egests, however in the anterior sections of the gut there are large rapid changes observed on the ingested transient community prior to this structure returning. It therefore follows that the cast material which the earthworm excretes forms the major constituent of soil (Drake & Horn 2007; Pass et al. 2014). As such the action of the earthworm on the transient soil may define the microbial community profile of their environment rather than the commonly assumed inverse; that the earthworm

microbiome is representing solely the ingested material. Such an effect coincides with the description of earthworms as 'ecosystem engineers' (Lavelle et al. 2006) and can further explain the beneficial effects which they have on maintaining a healthy soil environment (Blouin et al. 2013; Darwin 1896).

Given the major shifts and spatial localisation of taxonomic groups, it follows that the functional actions of the microbiome changes between the anatomical spatial assemblages. This was assessed using functional annotation from taxonomic assignment i.e. metagenome inference (Langille et al. 2013). While a crude approximation, it gives key insights into the underlying effects that may be occurring from the taxonomic changes.

The related *L. terrestris* species is known to increase the abundance of 'fermenter' species (Wüst et al. 2011) which coincides with methane (Depkat-Jakob et al. 2012) and N_2O (Horn et al. 2003; Ihssen et al. 2003) emission by earthworms and contributes to their role in greenhouse gas interactions (Lubbers et al. 2013). Methane metabolism associated pathways were found at their highest presence in the mid- to hindgut of the fresh earthworm at a level similar to the soil, and at a reduced level in the anterior sections and depurated gut. This likely occurs due to the relative suppression of the transient material during the initial ingestion, and the pathways being associated with the transient material rather than the gut wall-adhered species of the depurated samples. This suggests that the methane producing attributes which earthworms are sometimes associated with are not inherent to the earthworm-bacteria complex itself (i.e. bound to the gut tissues), but instead relates to the transient microbiota being stimulated in the specific chemical environment of the earthworm gut.

Earthworms have previously been shown to increase the presence of a number of bacterial species known to degrade both organic matter and complex anthropogenic compounds e.g. herbicides (Liu et al. 2011). The genomic capacity of the microbiome for a number of degradation pathways was found to be at the highest in the body wall tissues of the earthworm, with the most abundant predicted to be involved in benzoate degradation. As earthworm species inhabit polluted environments, these pathways would be of significant advantage to the earthworm when inhabiting such environments (Morgan et al. 2007; Sizmur & Hodson 2009). It is worth noting that the inferred metagenome process has a number of assumptions and likely conflates functionality based upon only a portion of the genetic pathway being observed. Whilst the pathway identified may be improbable given the environmental

considerations it is suggestive for a homologous system and provides additional evidence when considering the greater potential of minor taxonomic changes. The genes involved in the identified benzoate pathway for example, are mainly concerned with aromatic hydrocarbon degradation.

A number of novel OTUs which could not be accurately identified from reference databases were identified primarily in the anterior sections of the earthworm and although detected/present below annotation quality parameters, they could be identified predominantly as Alphaproteobacteria or Ternicutes. These species were observed mainly in the body wall tissue samples and may represent hereto unknown host-associated species, and were also present at a high relative abundance (up to ~30%) in both fresh and depurated body-wall tissues. However, it is unlikely that these species are vertically transmitted due to their absence from the cocoon community profiles. The occurrence, frequency or vector of bacterial transfer/acquirement from the soil into the host is unknown, however environmentally acquired bacteria have been confirmed as symbionts in insects encountering novel food sources (Hansen & Moran 2014), or newly hatched cocoons (Coelom et al. 2012).

The ubiquitous symbiont *Verminephrobacter* has previously been identified in the cocoon (Lund et al. 2009) with identification of biological channels though which they are actively introduced (Dulla et al. 2012). It was therefore expected to be the most abundant or sole species present in the sealed cocoon. Although distinct presence was observed, it was only the second highest taxonomic group, being in lesser abundance than a singular *Pedobacter* species OTU. The *Pedobacter* (Bacteroidetes) has been previously observed in earthworm nephridia at minor levels (Davidson et al. 2013) perhaps demonstrating chronic association such as that known to occur with *Verminephrobacter*, however this *Pedobacter* species was not observed at high abundances in other samples present within the soil or earthworm and creates uncertainty over its role within the formative stages of earthworm development.

6.5 Conclusion

The microbial voxelation of *Lumbricus rubellus* revealed an underlying microbial structure that was previously unprecedented in any invertebrate. Detection of several species would not have been possible at the whole-organism resolution which microbiomic studies are typically undertaken. A large number of species novel to the earthworm microbiome have been identified and their exact location in the host allude to potential functionality. The presence of a gut wall-associated community, which remains after depuration of the gut cavity, may indicate symbiotic potential. The novel observation of an unknown *Borellia* species demands further investigation in order to understand what role this earthworm species could have as a reservoir for a microbial species so closely linked to disease. However the absence of any species from the host in the cocoon other than the known *Verminephrobacter* species makes it unlikely that the novel species identified are vertically transmitted symbionts and must therefore be horizontally acquired.

7
Final Discussion

7.1 Introduction

Application of the latest High Throughput Sequencing (HTS) approaches to the earthworm bacterial microbiome was undertaken with the intention to interrogate the host-community system at an unprecedented level, to determine the correlation between the observable environmental effects of earthworms and the little understood host-microbial community interactions. This thesis focused predominantly upon the earthworm *Lumbricus rubellus* as a model for annelids and other soil invertebrates due to their prevalence and distribution in the United Kingdom, their predilection for contaminated environments, and the quantity of literature available on this popular research species. The assessment of the Azorean *Amynthas gracilis* was revealing firstly on the impact of environmental stressors, and in the wider context of cross-species correlations. A number of novel outcomes have been discovered which lay foundations for a wealth of new research to elucidate upon important species correlations and potential symbiosis where it was previously not believed to exist. In summary:

- The earthworm microbiome demonstrated a distinct community structure consistent at the phylum level yet unique from soil due to earthworm-enriched Actinobacteria and reduced Acidobacteria abundance.
- *L. rubellus* demonstrated a core community which arose from a number of host-associated species including the previously described *Vermineprobacter* symbiont; the novel crop associated gut Enterobacteraceae species, and enrichment of Flavobacteriaceae in the hindgut.
- Environmental stressors including heavy metal pollution reliably influenced the host microbiota, proportionally shifting the community profile and impacting the core community.
- High environmental contamination (particularly arsenic) eradicated the *Verminephrobacter* symbiont from *L. rubellus* which may impact host viability.
- Classical microbiological techniques (culture/molecular) are inadequate for accurately profiling the community and omit high amounts of diversity.
- Assessing the microbiome of an organism as a single unit obscures organ and tissue localised, highly abundant species to below detection limits.

7.2 Core community

Through deep High Throughput Sequencing analysis of the earthworm microbiome there is convincing evidence of a microbial community associated with the host. There is significant determination from the soil microbial community which it inhabits and ingests, however the combination of host-specific species and the effect which the earthworm gut microhabitat has on transient soil community results in a microbiome which is distinguished from the external surroundings. Earthworms

exhibit a number of roles on their immediate habitat e.g. organic matter degradation (Li et al. 2002; Jana et al. 2010) and the global environment e.g. increased N_20 emission (Horn et al. 2003; Drake et al. 2006). The accurate identification of the core microbiome which associates with this ubiquitous global invertebrate has provided a number of key targets for investigating the microbial role in assisting or inhibiting these processes. Furthermore, the high variability in the microbial community determined by site and contaminant implores further study into establishing the causative effect of this change on the host (perhaps detrimental), or the acquired functionality which has arisen from the novel associations.

There existed a consistent phylum level structure associated within the earthworm across at least two species (*L. rubellus* (Chapters 3,5 & 6) and *Amynthas gracilis (*Chapter 4)). Despite belonging to different taxonomic families and the geographic separation of the Azores-sourced *A. gracilis* (2,195 km from the UK), the microbiomic shift from the soil community was markedly similar, clearly demonstrating the consistent increase in Actinobacteria and reduction in Acidobacteria in comparison to the respective soil communities (Figure 7.1). While a regularly detected taxa of the rhizosphere (Rosenberg et al. 2014; Niva et al. 2006) it appears that the increased abundance of Actinobacteria versus natural soil populations is a novel result in earthworm microbial study (beyond passing observations (Wüst et al. 2011; Singleton et al. 2003)) and demonstrates the improvements which have been achieved by the HTS approach. No results suggested vertical transmission of bacterial species other than the already described *Verminephrobacter* symbiont (Betaproteobacteria) (Pinel et al. 2008; Lund et al. 2010) and it is most likely that the observed abundance originated from selective stimulation of the transient material due to physical conditions of the microhabitat and the organic metabolites secreted by the host into the gut lumen (Horn et al. 2003; Drake & Horn 2007).

A number of central genera represent the structure of this phylum; Mycobacterium, Streptomyces and additionally the Nocardiodiaceae Family, particularly in abundance in the mid- and hindgut (Figure 7.2). Mycobacterium is common in natural soils and waters (Iivanainen et al. 1997; Niva et al. 2006) and has previously been identified in *L. rubellus* inhabiting pasture fields of infected ruminants (Fischer et al. 2003). Streptomyces is also commonly identified in abundance from soil communities and their notoriety as antibiotic synthesisers (Kampfer et al. 2014) draws parallels with the antibacterial and antifungal effects for which earthworms have been historically studied (Pan et al. 2003; Khomiakov et al. 2007; Vasanthi & Singh 2013).

Nocardiodiaceae are perhaps of most interest due to the penchant of the family for metabolism of complex compounds including organic matter and toxic pollutants through aromatic carbon degradation (Toth & Borsodi 2014) and have been described as selected for in the termite hindgut which may draw parallels with *L. rubellus* in the potential function (Fall et al. 2007).

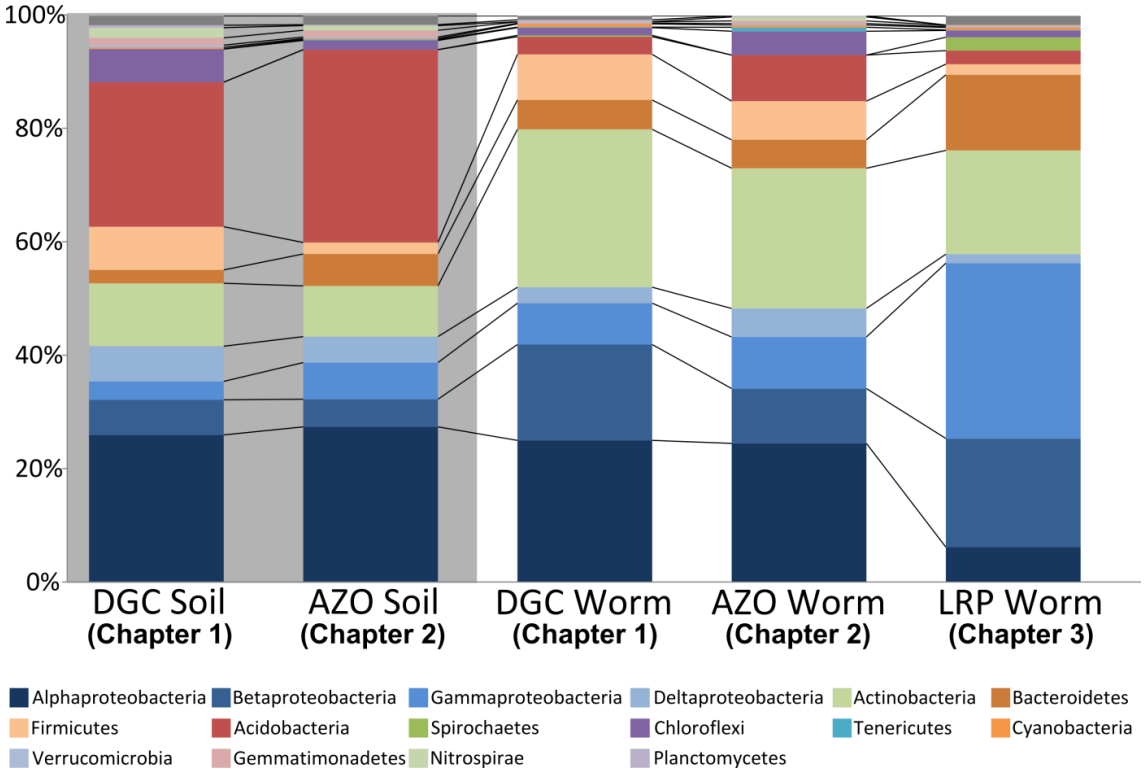


Figure 7.1 - Phylum level profile of all earthworms from each study. Proteobacteria is subdivided into class for detail. Independent earthworm and soil structures were largely similar, and demonstrate clearly the consistent abundance changes in Actinobacteria and Acidobacteria. The high-level overview obscures more significant changes such as for Betaproteobacteia where all soil species are lost and the *Verminephrobacter* symbiont dominates this clade in the earthworm.

The high abundance of Bacteroidetes, particularly flavobacterium in the hindgut of *L. rubellus* was found to be consistent in UK wide individuals and spatially confirmed to this area at a higher proportion in depurated earthworms, indicating gut wall adherence and potential involvement in the digestion of organic matter in symbiotic association with the earthworm host. Flavobacterium is typical of the soil environment and is commonly observed in the rhizosphere (Johansen et al. 2009; Kolton et al. 2012). It is chemoorganotrophic and adept at digesting polysaccharides and proteins which coincides with those secreted into the hindgut by the earthworm host (Wüst et al. 2011). Additionally, members exist as facultative anaerobes

(Bernardet et al. 1996) which would allow exploitation of the earthworm gut habitat as has been observed.

7.3 Whole organism versus spatial 'voxelation' for analysis of host-associated microbial communities

A remarkable finding was the amount of hidden diversity present within the host microbiome when visualising at the whole organism level. Species which represented a significant proportion of organ-associated microbiota e.g. the crop-associated Enterobacteraceae species (50%+ abundance in this digestive organ), were typically not detected to a noteworthy abundance and overshadowed by more dominating taxonomic clades. Similarly, the detection of a novel *Borrellia* sp. would not be possible in an 'averaged' organism, which raises concern given the prevalence of diseases vectors in species of this genus. When applying high resolution microbial profiling of individual body sections it became apparent that huge resolution was being lost and that despite the significant improvements in detection of bacterial presence through technological advancements, simple changes to experimental design can significantly improve assessing biotic systems and understanding the underlying structure. It is fitting to draw parallels with the concept of homogenising a whole human individual and determining that the resultant microbiome is representative of the nuanced system, when it is evident that communities both spatially distinct and chronically transient are in effect. With every subdivision of a whole sample the resolution is exponentially increased and this 'voxelation' was central to some of the most important discoveries of this thesis.

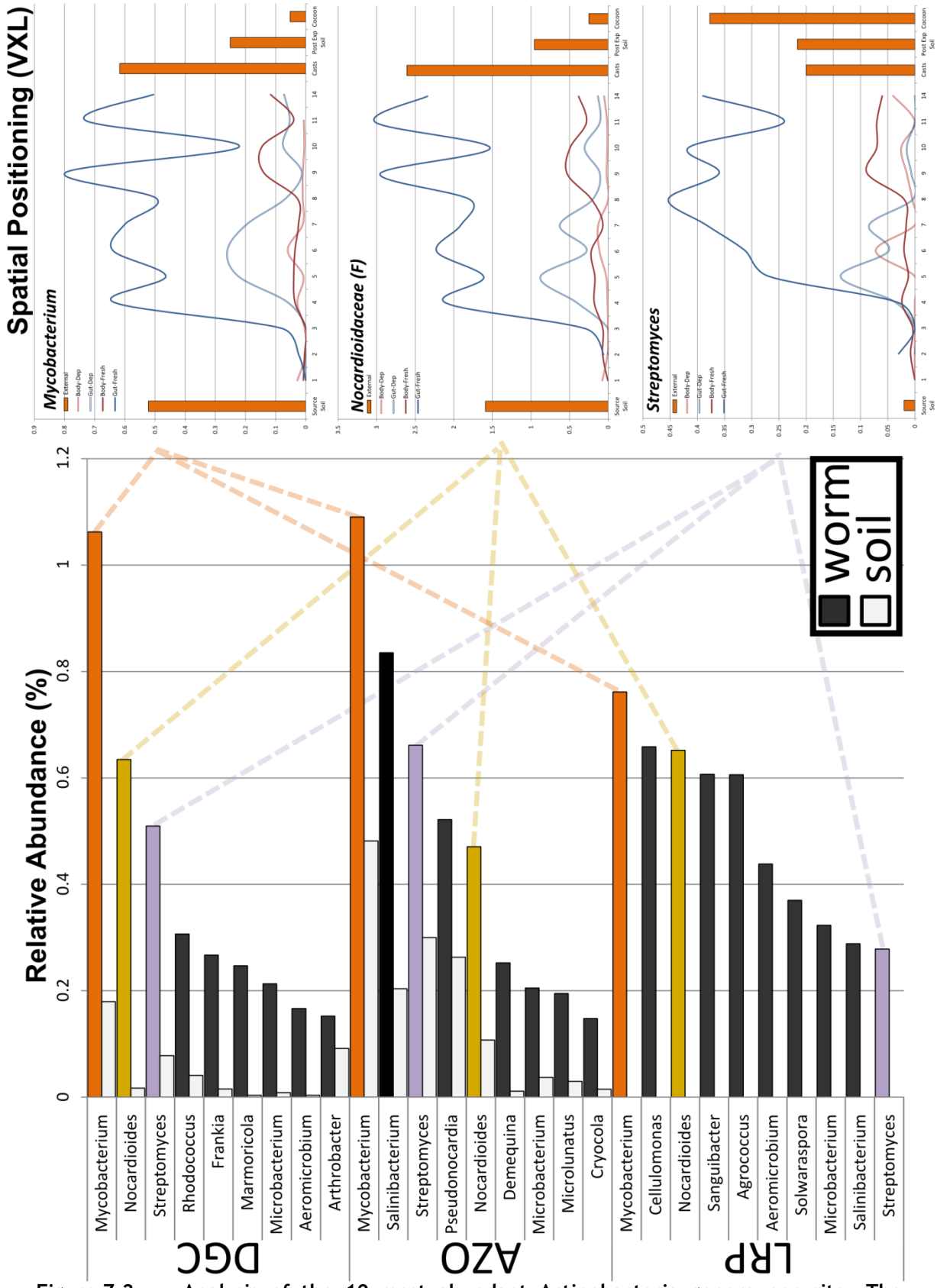


Figure 7.2 - Analysis of the 10 most abundant Actinobacteria genera per site. The earthworm consistently demonstrated an increased abundance of the Actinobacteria phylum beyond the soil. Genera are charted beside the soil (DGC and LRP) and spatial localisation within the host projected from voxelation analysis.

7.4 Improvements to technologies for microbial community assessment

7.4.1 Advantages of High Throughput Sequencing over classical techniques

Previous studies assessing the whole community profile of the earthworm microbiome were largely based upon classical microbiological techniques (Parle 1963), *in situ* observation i.e. fluorescent in situ hybridization (FISH) (Singleton et al. 2003) or molecular techniques such as Denaturing Gradient Gel Electrophoresis (DGGE) (Knapp et al. 2008), terminal-restriction fragment length polymorphism (T-RFLP) (Egert et al. 2004) and 16S clone library sequencing (Furlong et al. 2002). These techniques largely surmise little evidence of a distinct earthworm-associated community, but that there are observable differences between the soil and the earthworm gut community. DGGE analysis performed within this study (Chapter 3) confirmed the variability of the earthworm microbiota and its variance from the soil (Appendix 2) but was consistent with previous findings that the approach cannot accurately distinguish species (Knapp et al. 2008).

HTS approaches to bacterial community profiling are commonly focused upon a variable gene region of 16S rRNA (Huse et al. 2008; Hugenholtz 2002) which is the common target for the DGGE gel technique and improves upon clone library analysis by the huge increase in number of species identified. At the minimum achieved sequencing level of this thesis (Chapter 3), around 3,000 high quality 16S amplicon sequences were acquired per sample which is an order of magnitude higher than previous sequencing of the earthworm microbiome. This enabled the most accurate description of the community to date and confirmed that both phylum and species level differences exist between the soil and the host bacterial populations.

Some significant caveats exist for high throughput amplicon sequencing. Primarily, the presence of the 16S rRNA gene does not correspond to an active population as there is no distinction between active or senescent/dead cells. This could give rise to erroneous detection, however longitudinal analysis of the earthworm gut (Chapter 6) demonstrated that communities can alter abundance within minor time and spatial distance which suggests that the active population is highly dynamic and is unlikely to be carrying 'dead weight'. The presence of a taxonomic group does not indicate the functional activity of that clade both in regards to the senescence and the genomically encoded functional pathways. Mobile genetic elements may result in functionality which is not observed in reference genomes used for describing the community, and are likely to occur in populations chronically exposed to environmental stressors, and encourages caution when suggesting the functional

output of the bacterial population. Finally, sequencing approaches to community determination is only semi-quantitative, which must be considered when describing the microbiome. Quantitative measurements would greatly augment the sequencing results to enable a more accurate understanding of the microbiomic dynamics, such as 16S qPCR for accurately determining the community abundance, or targeting functional genes to determine the community activity and output.

7.4.2 Contrasting the sequencing technologies applied in this study

Three different High Throughput Sequencing technologies were utilised in generating the data for this study which allowed for direct comparison between the platforms (Figure 7.3). Over the duration of this thesis the sequencing costs of newer technologies fell, allowing for sequencing to be performed on the newer platforms for approximately the same costs. The region of the 16S rRNA gene which was sequenced (V3/V4) was consistent (Figure 2.1 (Chapter 2 – Methods)), however the benefit in number of samples and sequencing depth which Ion Torrent data or MiSeq platforms could provide over 454 technology was a viable exchange for the reduced resolution achievable from the significantly shorter length (~150bp after QC versus ~450bp from 454 FLX+ and ~500bp MiSeq Paired end (2x 300bp)). Concerning the experimental design, it was apparent that greater biological significance could be achieved by applying the improved sequencing yield to a larger number of samples rather than purely to improving the resolution on a smaller sample number. This was evident from the results generated on the Ion Torrent platform (Chapter 5) which greatly developed the understanding of the *L. rubellus* microbiome despite the reduced sequence quality. However, the increased length of MiSeq has once more altered the comparison and the combination of sequence length, quality and read count delivers significant enhancement over other options at this time.

Chapter	Sequencing Technology	Total number of amplicon reads	Number of samples	Lowest sample seq. count
3- DGC	454 Titanium/FLX+ (2x)	~1.2 Million	56	~3000
4- AZO	454 FLX++	~0.43 Million	56	~4,500
5- LRP	Ion Torrent PGM (4x)	~7.7 Million	170	~20,000
6- VXL	Illumina MiSeq (Paired end)	~17.1 Million	254	~5,000

Figure 7.3 - Table describing the generation of sequencing data in this thesis study. Due to sequencing variation when running the instrument there is often a large deviation between the expected average sequence count per sample.

7.4.3 Determining function in microbial communities

High Throughput Sequencing analysis of microbial communities has three approaches which present differing cost/benefits and apply to different requirements. This body of work has employed sequencing of microbial amplicons (sometimes referred to as metagenetics (Bik et al. 2012; Lodge et al. 2012)) and benefits from high resolution of community structure for limited cost, but is largely ignorant of functional capabilities of the biological system as no genomic information is available. The process of metagenomics has greater relevance to functional capabilities of the community through ostensibly producing whole genome constructs for the microbial population that allows determination of biological pathways (Woyke et al. 2006; Gill et al. 2006). However, assembly of the metagenome requires a large depth of sequencing to produce a modest number of genomes in comparison to amplicon sequencing, where each read is assumed to account for the presence of a species. Metatranscriptomics represents the most accurate assessment of microbial community functional output through direct determination of expressed genes, however this approach can provide limited information on the microbes contributing to the observed effects (Marchetti et al. 2012; Stewart et al. 2012).

For assessing the hereto unknown microbial community of the earthworm, 16S rRNA amplicon sequencing provided the depth of information required to describe the system, determine causative environmental effects on the community and identify potential targets for further investigation. In attempting to understand the wider connotations of community changes in response to environmental stressors or longitudinal location a novel method of metagenomic inference was employed (Chapters 4 & 6). The PICRUSt pipeline (Langille et al. 2013) attempts to bridge the gap between amplicon sequencing analysis and full metagenomics through deduction of the probable genomic capacity of the community. Assignment of an amplicon sequenced OTU to a high quality reference genome allowed extrapolation of the functional capacity of that OTU and the associated abundance. Ancestral State Reconstruction of the reference dataset allowed sequences for which no reference existed to be functionally annotated through determining common genomic pathways between the most recent common ancestors. Although not a substitute for independent metagenomic or metatranscriptomic assessment of the sample, suggestions which arise from the functional inference raise useful connotations which can direct future study and improve the value of the dataset when often there are several thousand independent species being represented.

7.5 Final Conclusions

Presented is the highest resolution sequencing analysis of host-associated bacteria to date of any single eukaryotic species, including spatial localisation of the entire *Lumbricus rubellus* organism and the impact of a wide range of anthropogenic contaminants and environmental stressors on the basal microbiomic community. A core bacterial community has been described which is distinct from the surrounding soil and a number of novel species have been associated with the earthworm crop, body wall and hindgut. This disputes long held claims that the earthworm has limited or no impact on ingested soil bacteria and demonstrates that the internal properties of the host impart significant effect on the transient material, demanding further analysis to determine potential symbiotic functionality. However, while a biologically important community has been described, the significant impact of anthropogenic contamination on the host microbiome must be considered given the observed eradication of the *Verminephrobacter* symbiont in response to the host's exposure to arsenic and the potential subsequent implications on host health. The impact which the earthworm has on the transient bacterial community is distinct and many of the ecological effects which it has historically been associated with are beginning to become understood. Findings strongly support further analysis of this essential ecosystem engineer to determine which of the phenotypic functions for which it is renowned for are attributable to the eukaryotic host and which are contributed to by the unique microbial community.

8
Bibliography

Alexandrino, M., Macías, F., Costa, R., Gomes, N.C.M., Canário, A.V.M. & Costa, M.C., 2011. A bacterial consortium isolated from an Icelandic fumarole displays exceptionally high levels of sulfate reduction and metals resistance. *Journal of hazardous materials*, 187(1-3), pp.362–70.

Allen, S.E., 1989. *Chemical analysis of ecological materials* 2nd Edition, Oxford: Blackwell Scientific Publications.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3), pp.403–10.

Anderson, T. & Joergensen, R.G., 1997. Relationship between SIR and FE estimates of microbial biomass C in deciduous forest soils at different pH. *Soil Biology and Biochemistry*, 29(7), pp.1033–1042.

Andre, J., Stürzenbaum, S.R., Kille, P., Morgan, a. J. & Hodson, M.E., 2010. 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 and Biochemistry*, 42(9), pp.1566–1573.

Andrews, S., 2014. FastQC. *Barbraham Institute*.

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D. a & Gordon, J.I., 2005. Host-bacterial mutualism in the human intestine. *Science (New York, N.Y.)*, 307(5717), pp.1915–20.

Barr, C.J., Bunce, R.G.H., Smart, S. & Whittaker, H.A., 1978. Countryside Survey 1978 vegetation plot data.

Bednar, A.J., Garbarino, J.R., Ranville, J.F. & Wildeman, T.R., 2005. Effects of iron on arsenic speciation and redox chemistry in acid mine water. *Journal of Geochemical Exploration*, 85(2), pp.55–62.

Bernard, L., Chapuis-Lardy, L., Razafimbelo, T., Razafindrakoto, M., Pablo, A.-L., Legname, E., Poulain, J., Brüls, T., O'Donohue, M., Brauman, A., Chotte, J.-L. & Blanchart, E., 2012. Endogeic earthworms shape bacterial functional communities and affect organic matter mineralization in a tropical soil. *The ISME journal*, 6(1), pp.213–22.

Bernardet, J.-F., Segers, P., Vancanneyt, M., Berthe, F., Kersters, K. & Vandamme, P., 1996. Cutting a Gordian Knot: Emended Classification and Description of the Genus *Flavobacterium*, Emended Description of the Family Flavobacteriaceae, and Proposal of *Flavobacterium hydatis* nom. Nov. (Basonym, *Cytophaga aquatilis* Strohl and. *International Journal of Systematic Bacteriology*, 46(85), pp.128–148.

Bhorgin, A.J. & Uma, K., 2014. Antimicrobial activity of Earthworm Powder (*Lampito mauritii*). , 3(1), pp.437–443.

Bik, E.M., Long, C.D., Armitage, G.C., Loomer, P., Emerson, J., Mongodin, E.F., Nelson, K.E., Gill, S.R., Fraser-Liggett, C.M. & Relman, D. a, 2010. Bacterial diversity in the oral cavity of 10 healthy individuals. *The ISME journal*, 4(8), pp.962–974.

Bik, H.M., Sung, W., De Ley, P., Baldwin, J.G., Sharma, J., Rocha-Olivares, A. & Thomas, W.K., 2012. Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. *Molecular Ecology*, 21, pp.1048–1059.

Blouin, M., Hodson, M.E., Delgado, E. a., Baker, G., Brussaard, L., Butt, K.R., Dai, J., Dendooven, L., Peres, G., Tondoh, J.E., Cluzeau, D. & Brun, J.-J., 2013. A review of earthworm impact on soil function and ecosystem services. *European Journal of Soil Science*, 64(2), pp.161–182.

Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P. & Tyson, G.W., 2012. Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nature Methods*, 9(5), pp.425–426.

Bragg, L.M., Stone, G., Butler, M.K., Hugenholtz, P. & Tyson, G.W., 2013. Shining a Light on Dark Sequencing: Characterising Errors in Ion Torrent PGM Data. *PloS computational biology*, 9(4), p.e1003031.

Bright, M. & Bulgheresi, S., 2010. A complex journey: transmission of microbial symbionts. *Nature reviews. Microbiology*, 8(3), pp.218–30.

Brown, G.G., Barois, I. & Lavelle, P., 2000. Regulation of soil organic matter dynamics and microbial activity in the drilosphere and the role of interactions with other edaphic functional domains. *European Journal of Soil Biology*, 36, p.177−198.

Brown, G.G. & Doube, B.M., 2004. Functional interactions between earthworms, microorganisms, organic matter, and plants. In E. C.A., ed. *Earthworm ecology*. Boca Raton, FL, USA: CRC Press LLC, pp. 213–239.

Brucker, R.M. & Bordenstein, S.R., 2013. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus Nasonia. *Science*, 341(6200), pp.667–9.

Burtelow, A.E., Bohlen, P.J. & Groffman, P.M., 1998. Influence of exotic earthworm invasion on soil organic matter, microbial biomass and denitrification potential in forest soils of the northeastern United States. *Applied Soil Ecology*, 9, pp.197–202.

Byzov, B. a., Nechitaylo, T.Y., Bumazhkin, B.K., Kurakov, a. V., Golyshin, P.N. & Zvyagintsev, D.G., 2009. Culturable microorganisms from the earthworm digestive tract. *Microbiology*, 78(3), pp.360–368.

Campbell, B.J., 2014. The Family Acidobacteriaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson, eds. *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. Springer Berlin Heidelberg, pp. 405–415.

Cao, X., Ma, L.Q. & Shiralipour, A., 2003. Effects of compost and phosphate amendments on arsenic mobility in soils and arsenic uptake by the hyperaccumulator, *Pteris vittata*. *Environmental Pollution*, 126(2), pp.157–167.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, 7(5), pp.335–6.

Chandler, J.A. & Turelli, M., 2014. Comment on " The hologenomic basis of speciation: Gut bacteria cause hybrid lethality in the genus *Nasonia*." *Science.*, 345(6200), p.1011.

Chaston, J. & Goodrich-Blair, H., 2010. Common trends in mutualism revealed by model associations between invertebrates and bacteria. *FEMS microbiology reviews*, 34(1), pp.41–58.

Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P. & O'Toole, P.W., 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic acids research*, 38(22).

Corp, N. & Morgan, a J., 1991. Accumulation of heavy metals from polluted soils by the earthworm, Lumbricus rubellus: can laboratory exposure of "control" worms reduce biomonitoring problems? *Environmental pollution (Barking, Essex : 1987)*, 74, pp.39–52.

Coyle, P., Philcox, J.C., Carey, L.C. & Rofe, A.M., 2002. Metallothionein: the multipurpose protein. *Cellular and Molecular Life Sciences (CMLS)*, 59(4), pp.627–647.

Curtis, T.P., Sloan, W.T. & Scannell, J.W., 2002. Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16), pp.10494–10499.

Daniel, O. & Anderson, J.M., 1992. Microbial biomass and activity in contrasting soil materials after passage through the gut of the earthworm Lumbricus rubellus Hoffmeister. *Soil Biology and Biochemistry*, 24(5), pp.465–470.

Darwin, C., 1896. *The Formation of Vegetable Mould, Through the Action of Worms: With Observations on Their Habits*, D. Appleton.

Davidson, S.K., Powell, R. & James, S., 2013. A global survey of the bacteria within earthworm nephridia. *Molecular phylogenetics and evolution*, 67(December), pp.188–200.

Davidson, S.K. & Stahl, D.A., 2008. Selective recruitment of bacteria during embryogenesis of an earthworm. *The ISME journal*, 2(5), pp.510–8.

Day, G.M., 1950. Influence of earthworms on soil microorganisms. *Soil Science*, 69(3).

Delzenne, N.M., Neyrinck, A.M., Bäckhed, F. & Cani, P.D., 2011. Targeting gut microbiota in obesity: effects of prebiotics and probiotics. *Nature Reviews Endocrinology*, 7(11), pp.639–646.

Depkat-Jakob, P.S., Brown, G.G., Tsai, S.M., Horn, M.A. & Drake, H.L., 2013. Emission of nitrous oxide and dinitrogen by diverse earthworm families from Brazil and resolution of associated denitrifying and nitrate-dissimilating taxa. *FEMS microbiology ecology*, 83(2), pp.375–91.

Depkat-Jakob, P.S., Hunger, S., Schulz, K., Brown, G.G., Tsai, S.M. & Drake, H.L., 2012. Emission of methane by *Eudrilus eugeniae* and other earthworms from Brazil. *Applied and environmental microbiology*, 78(8), pp.3014–9.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*, 72(7), pp.5069–72.

Diederen, B.M.W., 2008. *Legionella* spp. And Legionnaires' disease. *Journal of Infection*, 56, pp.1–12.

Donnelly, R. & Harper, G., 2013. Nuclear DNA recapitulates the cryptic mitochondrial lineages of *Lumbricus rubellus* and suggests the existence of cryptic species in an ecotoxological soil sentinel. *Biological Journal of …*, 110(4), pp.780–795.

Donnelly, R.K., Harper, G.L., Morgan, a. J., Pinto-Juma, G. a. & Bruford, M.W., 2014. Mitochondrial DNA and morphological variation in the sentinel earthworm species *Lumbricus rubellus*. *European Journal of Soil Biology*, 64, pp.23–29.

Drake, H., Schramm, A. & Horn, M., 2006. Earthworm gut microbial biomes: their importance to soil microorganisms, denitrification, and the terrestrial production of the greenhouse gas N2O. In *Intestinal Microorganisms of Soil Invertebrates*. Springer-Verlag, pp. 65–87.

Drake, H.L. & Horn, M.A., 2007. As the worm turns: the earthworm gut as a transient habitat for soil microbial biomes. *Annual review of microbiology*, 61, pp.169–89.

Dulla, G.F.J., Go, R. a, Stahl, D. a & Davidson, S.K., 2012. *Verminephrobacter eiseniae* type IV pili and flagella are required to colonize earthworm nephridia. *The ISME journal*, 6(6), pp.1166–75.

Dumonceaux, T.J., Schellenberg, J., Goleski, V., Hill, J.E., Jaoko, W., Kimani, J., Money, D., Ball, T.B., Plummer, F. a & Severini, A., 2009. Multiplex detection of bacteria associated with normal microbiota and with bacterial vaginosis in vaginal swabs by use of oligonucleotide-coupled fluorescent microspheres. *Journal of clinical microbiology*, 47(12), pp.4067–77.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)*, 26(19), pp.2460–1.

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods*, 10(10), pp.996–8.

Edwards, C.A., 2004. *Earthworm Ecology*, Taylor & Francis.

Edwards, C.A. & Fletcher, K.E., 1988. Interactions between earthworms and microorganisms in organic-matter breakdown. *Agriculture, Ecosystems & Environment*, 24, pp.235–247.

Egert, M., Marhan, S., Wagner, B., Scheu, S. & Friedrich, M.W., 2004. Molecular profiling of 16S rRNA genes reveals diet-related differences of microbial communities in soil, gut, and casts of *Lumbricus terrestris* (Oligochaeta: Lumbricidae). *FEMS microbiology ecology*, 48(2), pp.187–97.

Egert, M., Stingl, U., Bruun, L.D., Pommerenke, B., Brune, A. & Friedrich, M.W., 2005. Structure and topology of microbial communities in the major gut compartments of *Melolontha melolontha* larvae (Coleoptera: Scarabaeidae). *Applied and Environmental Microbiology*, 71(8), pp.4556–4566.

Egert, M., Wagner, B., Lemke, T., Brune, A. & Friedrich, M.W., 2003. Microbial Community Structure in Midgut and Hindgut of the Humus-Feeding Larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Applied and Environmental Microbiology*, 69(11), pp.6659–6668.

Ettema, C. & Wardle, D., 2002. Spatial soil ecology. *Trends in ecology & evolution*, 17(4), pp.177–183.

Fall, S., Hamelin, J., Ndiaye, F., Assigbetse, K., Aragno, M., Chotte, J.L. & Brauman, A., 2007. Differences between bacterial communities in the gut of a soil-feeding termite (*Cubitermes niokoloensis*) and its mounds. *Applied and Environmental Microbiology*, 73(16), pp.5199–5208.

Farmer, J.J., Davis, B.R., Hickman-Brenner, F.W., McWhorter, A., Huntley-Carter, G.P., Asbury, M.A., Riddle, C., Wathen-Grady, H.G., Elias, C. & Fanning, G.R., 1985.

Biochemical identification of new species and biogroups of Enterobacteriaceae isolated from clinical specimens. *Journal of clinical microbiology*, 21(1), pp.46–76.

Feller, C., Brown, G.G., Blanchart, E., Deleporte, P. & Chernyanskii, S.S., 2003. Charles Darwin, earthworms and the natural sciences: various lessons from past to future. *Agriculture, Ecosystems & Environment*, 99(1-3), pp.29–49.

Field, D., Tiwari, B., Booth, T., Houten, S., Swan, D., Bertrand, N. & Thurston, M., 2006. Open software for biologists: from famine to feast. *Nature biotechnology*, 24(7), pp.801–3.

Field, S.G. & Michiels, N.K., 2006. Acephaline gregarine parasites (*Monocystis* sp.) are not transmitted sexually among their lumbricid earthworm hosts. *The Journal of parasitology*, 92(2), pp.292–7.

Field, S.G. & Michiels, N.K., 2005. Parasitism and growth in the earthworm *Lumbricus terrestris*: fitness costs of the gregarine parasite *Monocystis* sp. *Parasitology*, 130(4), pp.397–403.

Fischer, O.. a., Matlova, L., Bartl, J., Dvorska, L., Svastova, P., du Maine, R., Melicharek, I., Bartos, M. & Pavlik, I., 2003. Earthworms (Oligochaeta, Lumbricidae) and mycobacteria. *Veterinary Microbiology*, 91(4), pp.325–338.

Frey, S.D., 2006. Spatial Distribution of Soil Organisms. In E. A. Paul, ed. *Soil Microbiology, Ecology and Biochemistry*. Soil Microbiology and Biochemistry Series. Elsevier Science, pp. 283–302.

Fuchs, G., Boll, M. & Heider, J., 2011. Microbial degradation of aromatic compounds – from one strategy to four. *Nature reviews. Microbiology*, 9(11), pp.803–16.

Furlong, M.A., Singleton, D.R., Coleman, D.C. & Whitman, W.B., 2002. Molecular and Culture-Based Analyses of Prokaryotic Communities from an Agricultural Soil and the Burrows and Casts of the Earthworm *Lumbricus rubellus*. *Applied and environmental microbiology*, 68(3), pp.1265–1279.

García, J.A. & Fragoso, C., 2002. Growth, reproduction and activity of earthworms in degraded and amended tropical open mined soils: Laboratory assays. *Applied Soil Ecology*, 20, pp.43–56.

Geider, R.J., Delucia, E.H., Falkowski, P.G., Finzi, A.C., Grime, J.P., Grace, J., Kana, T.M., Roche, J.L.A., Long, S.P., Osborne, B.A., Platt, T., Prentice, I.C., Raven, J.A., Schlesinger, W.H., Smetacek, V. & Stuart, V., 2001. Primary productivity of planet earth : biological determinants and physical constraints in terrestrial and aquatic habitats. *Global Change Biology*, 7, pp.849–882.

Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M. & Nelson, K.E., 2006. Metagenomic analysis of the human distal gut microbiome. *Science (New York, N.Y.)*, 312(7), pp.1355–1359.

Goffredi, S.K., Johnson, S.B. & Vrijenhoek, R.C., 2007. Genetic diversity and potential function of microbial symbionts associated with newly discovered species of *Osedax* polychaete worms. *Applied and environmental microbiology*, 73(7), pp.2314–23.

Gómez-Brandón, M., Aira, M., Lores, M. & Domínguez, J., 2011. Epigeic earthworms exert a bottleneck effect on microbial communities through gut associated processes. J. A. Gilbert, ed. *PloS one*, 6(9), p.e24786.

Gougoulias, C., Clark, J.M. & Shaw, L.J., 2014. The role of soil microbes in the global carbon cycle: tracking the below-ground microbial processing of plant-derived carbon for manipulating carbon dynamics in agricultural systems. *Journal of the science of food and agriculture*, 94(12), pp.2362–71.

Grice, E.A., Kong, H.H., Conlan, S., Deming, C.B., Davis, J., Young, A.C., Program, N.C.S., Bouffard, G.G., Blakesley, R.W., Murray, P.R., Green, E.D., Turner, M.L. & Segre, J.A., 2009. Topographical and Temporal Diversity of the Human Skin Microbiome. *Science* , 324 (5931), pp.1190–1192.

Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M. & Whiteley, A.S., 2011. The bacterial biogeography of British soils. *Environmental microbiology*, 13(6), pp.1642–54.

Hackstein, J.H. & Stumm, C.K., 1994. Methane production in terrestrial arthropods. *Proceedings of the National Academy of Sciences of the United States of America*, 91(12), pp.5441–5.

Hansen, A.K. & Moran, N. a, 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Molecular ecology*, 23(6), pp.1473–96.

Hathaway, J.J.M., Garcia, M.G., Balasch, M.M., Spilde, M.N., Stone, F.D., Dapkevicius, N.E., Lurdes, M. De, Amorim, I.R., Gabriel, R., Borges, P. a. V. & Northup, D.E., 2014. Comparison of Bacterial Diversity in Azorean and Hawai'ian Lava Cave Microbial Mats. *Geomicrobiology Journal*, 31(3), pp.205–220.

Heemsbergen, D.A., Berg, M.P., Loreau, M., van Hal, J.R., Faber, J.H. & Verhoef, H.A., 2004. Biodiversity effects on soil processes explained by interspecific functional dissimilarity. *Science (New York, N.Y.)*, 306(5698), pp.1019–1020.

Héry, M., Singer, A.C., Kumaresan, D., Bodrossy, L., Stralis-Pavese, N., Prosser, J.I., Thompson, I.P. & Murrell, J.C., 2008. Effect of earthworms on the community structure of active methanotrophic bacteria in a landfill cover soil. *The ISME journal*, 2(1), pp.92–104.

Horn, M.A., Schramm, A. & Drake, H.L., 2003. The earthworm gut: An ideal habitat for ingested N2O-producing microorganisms. *Applied and Environmental Microbiology*, 69(3), pp.1662–1669.

Hugenholtz, P., 2002. Exploring prokaryotic diversity in the genomic era. *Genome biology*, 3(2), pp.1–8.

Huse, S.M., Dethlefsen, L., Huber, J. a., Welch, D.M., Relman, D. a. & Sogin, M.L., 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PloS Genetics*, 4(11).

Ihssen, J., Horn, M.A., Matthies, C., Gößner, A., Schramm, A. & Drake, H.L., 2003. N2O-producing microorganisms in the gut of the earthworm Aporrectodea caliginosa are indicative of ingested soil bacteria. *Applied and Environmental Microbiology*, 69(3), pp.1655–1661.

Iivanainen, E.E.K., Martikainen, P.J., Räisänen, M.L. & Katila, M.L., 1997. Mycobacteria in boreal coniferous forest soils. *FEMS microbiology ecology*, 23, pp.325–332.

Imperio, T., Viti, C. & Marri, L., 2008. *Alicyclobacillus pohliae* sp. nov., a thermophilic endospore-forming bacterium isolated from geothermal soil of the north-west slope of Mount Melbourne (Antarctica). *International Journal of Systematic and Evolutionary Microbiology*, 58, pp.221–225.

Jana, U., Barot, S., Blouin, M., Lavelle, P., Laffray, D. & Repellin, A., 2010. Earthworms influence the production of above- and belowground biomass and the expression of genes involved in cell proliferation and stress responses in *Arabidopsis thaliana*. *Soil Biology and Biochemistry*, 42(2), pp.244–252.

Johansen, J.E., Nielsen, P. & Binnerup, S.J., 2009. Identification and potential enzyme capacity of flavobacteria isolated from the rhizosphere of barley (Hordeum vulgare L.). *Canadian journal of microbiology*, 55(3), pp.234–41.

Johnson, D.B., Okibe, N. & Roberto, F.F., 2003. Novel thermo-acidophilic bacteria isolated from geothermal sites in Yellowstone National Park: Physiological and phylogenetic characteristics. *Archives of Microbiology*, 180, pp.60–68.

Jolly, J.M., Lappin-Scott, H.M., Anderson, J.M. & Clegg, C.D., 1993. Scanning electron microscopy of the gut microflora of two earthworms: *Lumbricus terrestris* and *Octolasion cyaneum*. *Microbial Ecology*, 26(3), pp.235–245.

Kampfer, P., Glaeser, S.P., Parkes, L., Keulen, G. van & Dyson, P., 2014. The Family Streptomycetaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson, eds. *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. Springer Berlin Heidelberg, pp. 889–1010.

Kappler, a. & Brune, a., 1999. Influence of gut alkalinity and oxygen status on mobilization and size-class distribution of humic acids in the hindgut of soil-feeding termites. *Applied Soil Ecology*, 13, pp.219–229.

Kemmitt, S., Wright, D., Goulding, K. & Jones, D., 2006. pH regulation of carbon and nitrogen dynamics in two agricultural soils. *Soil Biology and Biochemistry*, 38(5), pp.898–911.

Khomiakov, N. V, Kharin, S. a., Nechitaĭlo, T.I., Golyshin, P.N., Kurakov, a. V., Byzov, B. a. & Zviagintsev, D.G., 2007. Reaction of microorganisms to the digestive fluid of the earthworms. *Mikrobiologiia*, 76(1), pp.55–65.

Kille, P., Andre, J., Anderson, C., Ang, H.N., Bruford, M.W., Bundy, J.G., Donnelly, R., Hodson, M.E., Juma, G., Lahive, E., Morgan, a. J., Stürzenbaum, S.R. & Spurgeon, D.J., 2013. DNA sequence variation and methylation in an arsenic tolerant earthworm population. *Soil Biology and Biochemistry*, 57, pp.524–532.

Knapp, B. a, Seeber, J., Podmirseg, S.M., Meyer, E. & Insam, H., 2008. Application of denaturing gradient gel electrophoresis for analysing the gut microflora of *Lumbricus rubellus* Hoffmeister under different feeding conditions. *Bulletin of entomological research*, 98(3), pp.271–9.

Kolton, M., Green, S.J., Harel, Y.M., Sela, N., Elad, Y. & Cytryn, E., 2012. Draft genome sequence of *Flavobacterium* sp. strain F52, isolated from the rhizosphere of bell pepper (*Capsicum annuum* L. Cv. Maccabi). *Journal of Bacteriology*, 194(19), pp.5462–5463.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K. & Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology*, 79(17), pp.5112–20.

Lamelas, A., Gosalbes, M.J., Manzano-Marín, A., Peretó, J., Moya, A. & Latorre, A., 2011. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. P. M. Richardson, ed. *PloS genetics*, 7(11), p.e1002357.

Langdon, C.J., Morgan, A.J., Charnock, J.M., Semple, K.T. & Lowe, C.N., 2009. As-resistance in laboratory-reared F1, F2 and F3 generation offspring of the earthworm Lumbricus rubellus inhabiting an As-contaminated mine soil. *Environmental pollution*, 157(11), pp.3114–9.

Langdon, C.J., Piearce, T.G., Meharg, A.A. & Semple, K.T., 2003. Interactions between earthworms and arsenic in the soil environment: a review. *Environmental Pollution*, 124(3), pp.361–373.

Langdon, C.J., Piearce, T.G., Meharg, A.A. & Semple, K.T., 2001. Survival and behaviour of the earthworms *Lumbricus rubellus* and *Dendrodrilus rubidus* from arsenate-contaminated and non-contaminated sites. *Soil Biology and Biochemistry*, 33(9), pp.1239–1244.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J. a, Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G. & Huttenhower, C., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology*, 31(9), pp.814–21.

Lapanje, A., Zrimec, A., Drobne, D. & Rupnik, M., 2010. Long-term Hg pollution-induced structural shifts of bacterial community in the terrestrial isopod (*Porcellio scaber*) gut. *Environmental Pollution*, 158(10), pp.3186–93.

Larkin, M. a., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P. a., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, a., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), pp.2947–2948.

Lassalle, F., Lassegues, M. & Roch, P., 1988. Protein analysis of earthworm coelomic fluid IV. Evidence, activity induction and purification of Eisenia fetida andrei lysozyme (Annelidae). *Comparative Biochemistry and Physiology*, 91(1), pp.187–192.

Lavelle, P., Decaëns, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P. & Rossi, J.-P., 2006. Soil invertebrates and ecosystem services. *European Journal of Soil Biology*, 42, pp.S3–S15.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R. & Gordon, J.I., 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature reviews. Microbiology*, 6(10), pp.776–788.

Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., Zhang, Y., Shen, J., Pang, X., Zhang, M., Wei, H., Chen, Y., Lu, H., Zuo, J., Su, M., Qiu, Y., Jia, W., Xiao, C., Smith, L.M., Yang, S., Holmes, E., Tang, H., Zhao, G., Nicholson, J.K., Li, L. & Zhao, L., 2008. Symbiotic gut microbes modulate human metabolic phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*, 105(6), pp.2117–22.

Li, X., Fisk, M.C., Fahey, T.J. & Bohlen, P.J., 2002. Influence of earthworm invasion on soil microbial biomass and activity in a northern hardwood forest. *Soil Biology and Biochemistry*, 34(12), pp.1929–1937.

Liebeke, M., Bruford, M.W., Donnelly, R.K., Ebbels, T.M.D., Hao, J., Kille, P., Lahive, E., Madison, R.M., Morgan, A.J., Pinto-juma, G.A., Spurgeon, D.J., Svendsen, C. & Bundy, J.G., 2014. Identifying biochemical phenotypic differences between cryptic species. *Biology …*, 10(9).

Liu, X., Hu, C. & Zhang, S., 2005. Effects of earthworm activity on fertility and heavy metal bioavailability in sewage sludge. *Environment international*, 31(6), pp.874–9.

Liu, Y.-J., Zaprasis, A., Liu, S.-J., Drake, H.L. & Horn, M.A., 2011. The earthworm *Aporrectodea caliginosa* stimulates abundance and activity of phenoxyalkanoic acid herbicide degraders. *The ISME journal*, 5(3), pp.473–85.

Lodge, D.M., Turner, C.R., Jerde, C.L., Barnes, M. a., Chadderton, L., Egan, S.P., Feder, J.L., Mahon, A.R. & Pfrender, M.E., 2012. Conservation in a cup of water: Estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, 21, pp.2555–2558.

Lory, S., 2014. The Family Mycobacteriaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson, eds. *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. Springer Berlin Heidelberg, pp. 570–575.

Lozupone, C. & Knight, R., 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Society*, 71(12), pp.8228–8235.

Lubbers, I.M., van Groenigen, K.J., Fonte, S.J., Six, J., Brussaard, L. & van Groenigen, J.W., 2013. Greenhouse-gas emissions from soils increased by earthworms. *Nature Climate Change*, 3(3), pp.187–194.

Lund, M.B., Davidson, S.K., Holmstrup, M., James, S., Kjeldsen, K.U., Stahl, D.A. & Schramm, A., 2009. Diversity and host specificity of the *Verminephrobacter*-earthworm symbiosis. *Environmental microbiology*, 12, pp.2142–2151.

Lund, M.B., Holmstrup, M., Lomstein, B.A., Damgaard, C. & Schramm, A., 2010. Beneficial effect of *Verminephrobacter* nephridial symbionts on the fitness of the earthworm Aporrectodea tuberculata. *Applied and environmental microbiology*, 76(14), pp.4738–43.

Lund, M.B., Kjeldsen, K.U. & Schramm, A., 2014. The earthworm-*Verminephrobacter* symbiosis: an emerging experimental system to study extracellular symbiosis. *Frontiers in microbiology*, 5(3), p.128.

Marchetti, a., Schruth, D.M., Durkin, C. a., Parker, M.S., Kodner, R.B., Berthiaume, C.T., Morales, R., Allen, a. E. & Armbrust, E. V., 2012. PNAS Plus: Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proceedings of the National Academy of Sciences*, 109(6), pp.E317–E325.

Mardis, E.R., 2008. Next-generation DNA sequencing methods. Annual review of genomics and human genetics, 9, pp.387–402.

Morecroft, M.D., Bealey, C.E., Beaumont, D. a., Benham, S., Brooks, D.R., Burt, T.P., Critchley, C.N.R., Dick, J., Littlewood, N. a., Monteith, D.T., Scott, W. a., Smith, R.I., Walmsley, C. & Watson, H., 2009. The UK Environmental Change Network: Emerging trends in the composition of plant and animal communities and the physical environment. *Biological Conservation*, 142(12), pp.2814–2832.

Morgan, A.J., Kille, P. & Stürzenbaum, S.R., 2007. Microevolution and Ecotoxicology of Metals in Invertebrates. *Environmental Science & Technology*, 41(4), pp.1085–1096.

Morgan, A.J. & Mariño, F., 1999. The time-course of metal (Ca, Cd, Cu, Pb, Zn) accumulation from a contaminated soil by three populations of the earthworm, *Lumbricus rubellus*. *Applied Soil Ecology*, 12(2), pp.169–177.

Morgan H., N., 1988. The role of micro-organisms in the nutrition of *Eisenia foetida*. In E. C.A., ed. *Earthworms in waste and environmental management*. The Hague: SPB Academic Publishing, pp. 71–82.

Moya, A., Peretó, J., Gil, R. & Latorre, A., 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nature reviews. Genetics*, 9(3), pp.218–29.

Nahmani, J., Hodson, M.E. & Black, S., 2007. A review of studies performed to assess metal uptake by earthworms. *Environmental pollution (Barking, Essex : 1987)*, 145(2), pp.402–24.

Nechitaylo, T.Y., Yakimov, M.M., Godinho, M., Timmis, K.N., Belogolova, E., Byzov, B.A., Kurakov, A. V, Jones, D.L. & Golyshin, P.N., 2010. Effect of the earthworms *Lumbricus terrestris* and *Aporrectodea caliginosa* on bacterial diversity in soil. *Microbial ecology*, 59(3), pp.574–87.

Niva, M., Hernesmaa, A., Haahtela, K., Salkinoja-salonen, M., Sivonen, K. & Haukka, K., 2006. Actinobacterial communities of boreal forest soil and lake water are rich in mycobacteria. *Boreal Environment Research*, 11(February), pp.45–53.

Norris, T.B., Wraith, J.M., Castnholz, R.W. & McDermott, T.R., 2002. Soil Microbial Community Structure across a Thermal Gradient following a Geothermal Heating Event Soil. *Applied and environmental microbiology*, 68(12), pp.6300–6309.

Novo, M., Morgan, a. J., Kille, P. & Cunha, L., How to thrive on a volcanic caldera: Studies of an extremophile metazoan. *In preperation*.

Nunan, N., Wu, K., Young, I.M., Crawford, J.W. & Ritz, K., 2002. In situ spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil. *Microbial Ecology*, 44(4), pp.296–305.

OECD, 1984. Guidelines for the testing of chemicals. No. 207 Earthworm acute toxicity tests. In Paris, France.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H., 2011. Vegan: Community Ecology Package.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D. a. & Brown, P.O., 2007. Development of the human infant intestinal microbiota. *PloS Biology*, 5(7), pp.1556–1573.

Pan, W., Liu, X., Ge, F. & Zheng, T., 2003. Reconfirmation of antimicrobial activity in the coelomic fluid of the earthworm *Eisenia fetida andrei* by colorimetric assay. *Journal of biosciences*, 28(6), pp.723–731.

Pandazis, G., 1931. Zur Frage der Bakteriensymbiose bei Oligochaeten. *Zentralbl Bakteriol*, 120, pp.440–453.

Parks, D.H., Tyson, G.W., Hugenholtz, P. & Beiko, R.G., 2014. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* , 30 (21), pp.3123–3124.

Parle, J., 1963. Micro-organisms in the intestines of earthworms. *Journal of General Microbiology*, 30(3), pp.1–11.

Pass, D.A., Morgan, A.J., Read, D.S., Field, D., Weightman, A.J. & Kille, P., 2014. The effect of anthropogenic arsenic contamination on the earthworm microbiome. *Environmental microbiology*.

Pinel, N., Davidson, S.K. & Stahl, D.A., 2008. *Verminephrobacter eiseniae* gen. nov., sp. nov., a nephridial symbiont of the earthworm *Eisenia foetida* (Savigny). *International journal of systematic and evolutionary microbiology*, 58(Pt 9), pp.2147–57.

Price, M.N., Dehal, P.S. & Arkin, A.P., 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution*, 26(7), pp.1641–50.

Quail, M., Smith, M.E., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P. & Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics*, 13(1), p.1.

Quince, C., Lanzén, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F. & Sloan, W.T., 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature methods*, 6(9), pp.639–41.

Quince, C., Lanzen, A., Davenport, R.J. & Turnbaugh, P.J., 2011. Removing noise from pyrosequenced amplicons. *BMC bioinformatics*, 12(1), p.38.

R Core Team & R Development Core Team, 2013. R: A Language and Environment for Statistical Computing.

Reeder, J. & Knight, R., 2010. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nature Methods*, 7(9), pp.668–669.

Rigottier-Gois, L., 2013. Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis. *The ISME journal*, 7(7), pp.1256–61.

Rinaldi, a. P., Vandemeulebrouck, J., Todesco, M. & Viveiros, F., 2012. Effects of atmospheric conditions on surface diffuse degassing. *Journal of Geophysical Research*, 117(B11201).

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G. & Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME journal*, 1(4), pp.283–90.

Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. & Thompson, F., 2014. *The Prokaryotes: Actinobacteria*, Springer Berlin Heidelberg.

Rousk, J., Bååth, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., Knight, R. & Fierer, N., 2010. Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME journal*, 4(10), pp.1340–51.

Russell, E.J. & Wild, A., 1988. *Russell's soil conditions and plant growth*, Longman Scientific & Technical.

Sabri, A., Leroy, P., Haubruge, E., Hance, T., Frère, I., Destain, J. & Thonart, P., 2011. Isolation, pure culture and characterization of Serratia symbiotica sp. nov., the R-type of secondary endosymbiont of the black bean aphid Aphis fabae. *International journal of systematic and evolutionary microbiology*, 61(9), pp.2081–8.

Sahm, K., John, P., Nacke, H., Wemheuer, B., Grote, R., Daniel, R. & Antranikian, G., 2013. High abundance of heterotrophic prokaryotes in hydrothermal springs of the Azores as revealed by a network of 16S rRNA gene-based methods. *Extremophiles: life under extreme conditions*, 17(4), pp.649–62.

Coelom, H., Kreutzer, E., Sudakaran, S. & Kaltenpoth, M., 2012. Actinobacteria as essential symbionts in firebugs and cotton stainers (Hemiptera, Pyrrhocoridae). *Environmental microbiology*.

Santrůcková, H., Bird, M.I., Elhottová, D., Novák, J., Picek, T., Simek, M. & Tykva, R., 2005. Heterotrophic fixation of CO_2 in soil. *Microbial ecology*, 49(2), pp.218–25.

Schaefer, M., Petersen, S.O. & Filser, J., 2005. Effects of *Lumbricus terrestris*, *Allolobophora chlorotica* and *Eisenia fetida* on microbial community dynamics in oil-contaminated soil. *Soil Biology and Biochemistry*, 37(11), pp.2065–2076.

Schloss, P.D., Delalibera, I., Handelsman, J. & Raffa, K.F., 2006. Bacteria Associated with the Guts of Two Wood-Boring Beetles: *Anoplophora glabripennis* and *Saperda vestita* (Cerambycidae). *Environmental Entomology*, 35, pp.625–629.

Schönholzer, F., Hahn, D., Zarda, B. & Zeyer, J., 2002. Automated image analysis and in situ hybridization as tools to study bacterial populations in food resources, gut and cast of *Lumbricus terrestris*. *Journal of Microbiological Methods*, 48(1), pp.53–68.

Schramm, A., Davidson, S.K., Dodsworth, J. a., Drake, H.L., Stahl, D.A. & Dubilier, N., 2003. *Acidovorax*-like symbionts in the nephridia of earthworms. *Environmental Microbiology*, 5(9), pp.804–809.

Sechi, P., OrozcoVTerwengel, P., Russo, I.-R., P, K. & MW, B., Paleoclimatic modelling and cryptic mitochondrial diversity in the earthworm *Lumbricus rubellus* in Europe. *Evolution - Submitted*.

Selosse, M.-A., Baudoin, E. & Vandenkoornhuyse, P., 2004. Symbiotic microorganisms, a key for ecological success and protection of plants. *Comptes Rendus Biologies*, 327(7), pp.639–648.

Senderovich, Y. & Halpern, M., 2013. The protective role of endogenous bacterial communities in chironomid egg masses and larvae. *The ISME journal*, 7(11), pp.2147–58.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), pp.2498–504.

Sharma, S., Pradhan, K., Satya, S. & Vasudevan, P., 2005. Potentiality of Earthworms for Waste Management and in Other Uses – A Review. *The Journal of American Science*, 1(1), pp.4–16.

Silva, C., Ferreira, T. & Viveiros, F., 2007. *Radon (222Rn) soil gas measurements at Furnas Volcano (S. Miguel Island, Azores)*,

Sims, R.W. & Gerard, B.M., 1985. *Earthworms: Keys and Notes for the Identification and Study of the Species*, London: Linnean Society of London and the Estuarine and Brackish-Water Sciences Association.

Singleton, D.R., Hendrix, P.F., Coleman, D.C. & Whitman, W.B., 2003. Identification of uncultured bacteria tightly associated with the intestine of the earthworm *Lumbricus rubellus* (Lumbricidae; Oligochaeta). *Soil Biology and Biochemistry*, 35(12), pp.1547–55.

Sizmur, T. & Hodson, M.E., 2009. Do earthworms impact metal mobility and availability in soil?–a review. *Environmental Pollution*, 157(7), pp.1981–9.

Sizmur, T., Watts, M.J., Brown, G.D., Palumbo-Roe, B. & Hodson, M.E., 2011. Impact of gut passage and mucus secretion by the earthworm *Lumbricus terrestris* on mobility and speciation of arsenic in contaminated soil. *Journal of hazardous materials*, 197, pp.169–75.

Sparling, G.P., 1992. Ratio of microbial biomass carbon to soil organic carbon as a sensitive indicator of changes in soil organic matter. *Soil Research*, 30(2), pp.195–207.

Spurgeon, D.J., Lawlor, A., Hooper, H.L., Wadsworth, R., Svendsen, C., Thomas, L.D.K., Ellis, J.K., Bundy, J.G., Keun, H.C. & Jarup, L., 2011. Outdoor and indoor cadmium distributions near an abandoned smelting works and their relations to human exposure. *Environmental Pollution*, 159(12), pp.3425–3432.

Spurgeon, D.J., Weeks, J.M. & Gestel, C.A.M. Van, 2003. A summary of eleven years progress in earthworm ecotoxicology. *Pedobiologia*, 47, pp.588–606.

Stewart, F.J., Ulloa, O. & Delong, E.F., 2012. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology*, 14, pp.23–40.

Stürzenbaum, S.R., Andre, J., Kille, P. & Morgan, a J., 2009. Earthworm genomes, genes and proteins: the (re)discovery of Darwin's worms. *Proceedings of the Royal Society of London. Series B - Biological Sciences*, 276(1658), pp.789–97.

Stürzenbaum, S.R., Winters, C., Galay, M., Morgan, a. J. & Kille, P., 2001. Metal ion trafficking in earthworms. Identification of a cadmium-specific metallothionein. *Journal of Biological Chemistry*, 276(36), pp.34013–34018.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28(10), pp.2731–2739.

Teixeira, L., Ferreira, A. & Ashburner, M., 2008. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PloS biology*, 6(12), p.e2.

Thakuria, D., Schmidt, O., Finan, D., Egan, D. & Doohan, F.M., 2010. Gut wall bacteria of earthworms: a natural selection process. *The ISME journal*, 4(3), pp.357–66.

The Human Microbiome Consortium, 2012. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), pp.207–14.

Tikhonov, V. V., Byzov, B. a., Zavgorodnyaya, Y. a. & Demin, V. V., 2011. Earthworms as modifiers of the structure and biological activity of humic acids. *Biology Bulletin*, 38(1), pp.17–24.

Toth, E.M. & Borsodi, A.K., 2014. The Family Nocardioidaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson, eds. *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. Springer Berlin Heidelberg, pp. 651–700.

Turpeinen, R., Pantsar-Kallio, M., Häggblom, M. & Kairesalo, T., 1999. Influence of microbes on the mobilization, toxicity and biomethylation of arsenic in soil. *The Science of the total environment*, 236(1-3), pp.173–80.

Turroni, F., Peano, C., Pass, D. a, Foroni, E., Severgnini, M., Claesson, M.J., Kerr, C., Hourihane, J., Murray, D., Fuligni, F., Gueimonde, M., Margolles, A., De Bellis, G., O'Toole, P.W., van Sinderen, D., Marchesi, J.R. & Ventura, M., 2012. Diversity of bifidobacteria within the infant gut microbiota. *PloS one*, 7(5), p.e36957.

Vasanthi, K. & Singh, A.J.A.R., 2013. Antimicrobial activity of earthworm (*Eudrilus eugeniae*) paste. *African Journal of Environmental Science and Technology*, 7(August), pp.783–789.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F. & van Sinderen, D., 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and molecular biology reviews*, 71(3), pp.495–548.

Verna, C., Ramette, A., Wiklund, H., Dahlgren, T.G., Glover, A.G., Gaill, F. & Dubilier, N., 2010. High symbiont diversity in the bone-eating worm *Osedax mucofloris* from shallow whale-falls in the North Atlantic. *Environmental microbiology*, 12(8), pp.2355–70.

Walter, J. & Ley, R., 2011. The human gut microbiome: ecology and recent evolutionary changes. *Annual review of microbiology*, 65, pp.411–29.

Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S.G., Podar, M., Martin, H.G., Kunin, V., Dalevi, D., Madejska, J., Kirton, E., Platt, D., Szeto, E., Salamov, A., Barry, K., Mikhailova, N., Kyrpides, N.C., Matson, E.G., Ottesen, E.A., Zhang, X., Hernandez, M., Murillo, C., Acosta, L.G., Rigoutsos, I., Tamayo, G., Green, B.D., Chang, C., Rubin, E.M., Mathur, E.J., Robertson, D.E., Hugenholtz, P. & Leadbetter, J.R., 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, 450(7169), pp.560–565.

Webster, G., Parkes, R.J., Cragg, B. a, Newberry, C.J., Weightman, A.J. & Fry, J.C., 2006. Prokaryotic community composition and biogeochemical processes in deep subseafloor sediments from the Peru Margin. *FEMS microbiology ecology*, 58(1), pp.65–85.

Weiss, B.L., Maltz, M. & Aksoy, S., 2012. Obligate symbionts activate immune system development in the Tsetse fly. *Journal of immunology*, 188(7), pp.3395–403.

Wickham, H., 2009. *Ggplot2: elegant graphics for data analysis*, Springer New York.

Woyke, T., Teeling, H., Ivanova, N.N., Huntemann, M., Richter, M., Gloeckner, F.O., Boffelli, D., Anderson, I.J., Barry, K.W., Shapiro, H.J., Szeto, E., Kyrpides, N.C., Mussmann, M., Amann, R., Bergin, C., Ruehland, C., Rubin, E.M. & Dubilier, N., 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature*, 443(7114), pp.950–955.

Wüst, P.K., Horn, M.A. & Drake, H.L., 2011. Clostridiaceae and Enterobacteriaceae as active fermenters in earthworm gut content. *The ISME journal*, 5(1), pp.92–106.

Wust, P.K., Horn, M.A., Drake, H.L. & Wüst, P.K., 2009. In Situ Hydrogen and Nitrous Oxide as Indicators of Concomitant Fermentation and Denitrification in the Alimentary Canal of the Earthworm *Lumbricus terrestris*. *Applied and environmental microbiology*, 75(7), pp.1852–9.

Yasir, M., Aslam, Z., Kim, S.W., Lee, S.-W., Jeon, C.O. & Chung, Y.R., 2009. Bacterial community composition and chitinase gene diversity of vermicompost with antifungal activity. *Bioresource technology*, 100(19), pp.4396–403.

9
Appendices

Appendix 1 - HTS Sample Sequencing Descriptions

Appendix 1a - Chapter 3 (DGC) Sample Descriptions

Forward Linker Primer Sequence: CCTNCGGGNGGCAGCAG

SampleID	BarcodeSequence	Type	Site
S1A	AACCATAA	SOIL	1
S1B	AACTCATG	SOIL	1
S1C	AACTCATA	SOIL	1
S2A	AACCGCCG	SOIL	2
S2B	AACCATCT	SOIL	2
S2C	AACTCAGT	SOIL	2
S3A	AACTCACA	SOIL	3
S3B	AACCATCG	SOIL	3
S3C	AACTCAGG	SOIL	3
S4A	AACCGCCA	SOIL	4
S4B	AACCATCA	SOIL	4
S4C	AACTCACC	SOIL	4
S5A	AACCGCAT	SOIL	5
S5B	AACCATCC	SOIL	5
S5C	AACTCAGC	SOIL	5
S6A	AACCGCAG	SOIL	6
S6B	AACCATAG	SOIL	6
S6C	AACTCACT	SOIL	6
S7A	AACCGATT	SOIL	7
S7B	AACCATAC	SOIL	7
S7C	AACTCACG	SOIL	7
W12	ATCGATGA	WORM	1
W13	ATCGTACA	WORM	1
W14	AACCGAGA	WORM	1
W15	AACCGATG	WORM	1
W21	ATCTCTGA	WORM	2
W22	ATCGCATA	WORM	2
W23	AACTATTG	WORM	2
W24	AACCATGG	WORM	2
W25	AACCGATC	WORM	2
W31	AACTCAAC	WORM	3
W32	AACTCAAG	WORM	3
W33	ATCGGTAA	WORM	3
W34	AACCATGC	WORM	3
W35	AACCGATA	WORM	3
W41	AACCAGTT	WORM	4
W42	AACCAGTC	WORM	4
W43	ATCGGCTA	WORM	4
W44	AACCATGA	WORM	4

W45	AACTATTC	WORM	4
W52	AACCAGTG	WORM	5
W53	ATCGGAGA	WORM	5
W54	AACCGAGT	WORM	5
W55	AACCGAGG	WORM	5
W62	AACCAGGT	WORM	6
W63	ATCGCTCA	WORM	6
W64	ATCTAGGA	WORM	6
W65	AACCGAGC	WORM	6
W71	ATCTATTA	WORM	7
W72	AACCAGGC	WORM	7
W73	ATCGTGTA	WORM	7
W74	ATCTACCA	WORM	7
W75	ATCGCGAA	WORM	7

Appendix 1b - Chapter 4 (AZO) Sample Descriptions

Forward Linker Primer Sequence: CCTACGGGAGGCAGCAG

SampleID	BarcodeSequence	Type	Source	Exposure
S1A	AATCACAA	worm	AV	AV
S1B	CAGCTCGA	worm	AV	AV
S1C	CGTCTAGA	worm	AV	AV
S2A	CAGGACCA	worm	AV	AV
S2B	CAGGAGGA	worm	AV	AV
S2C	CAGGATTA	worm	AV	AV
S3A	CAGGCCAA	worm	AV	AV
S3B	CAGGCGCA	worm	AV	AV
S3C	CGTCTCTA	worm	AV	AV
S4A	AATTCTTA	worm	RF	AV
S4B	CAGGTCTA	worm	RF	AV
S4C	CAGGTTAA	worm	RF	AV
S5A	CAGTAACA	worm	RF	AV
S5B	AATTGCAA	worm	RF	AV
S5C	CAGTAGTA	worm	RF	AV
S6A	CAGTCGGA	worm	RF	AV
S6B	CAGTCTTA	worm	RF	AV
S6C	CGTAACAA	worm	RF	AV
S7A	CGTAAGCA	worm	AV	RF
S7B	CGTAATGA	worm	AV	RF
S7C	CGTACATA	worm	AV	RF
S8A	CGTACGAA	worm	AV	RF
S8B	CGTACTCA	worm	AV	RF
S8C	AATTGGCA	worm	AV	RF
S9A	CGTAGCTA	worm	AV	RF

S9B	CGTAGTAA	worm	AV	RF
S9C	CGTATACA	worm	AV	RF
S10A	CGTATCGA	worm	RF	RF
S10B	CGTATGTA	worm	RF	RF
S10C	CGTCACCA	worm	RF	RF
S11A	CGTCAGGA	worm	RF	RF
S11B	AATTGTGA	worm	RF	RF
S11C	CGTCCGCA	worm	RF	RF
S12A	CGTCCTGA	worm	RF	RF
S12B	CGTCGATA	worm	RF	RF
S12C	CGTCGGAA	worm	RF	RF
S13A	AATCAGCA	soil	AV	AV
S13B	ACAACCGA	soil	AV	AV
S13C	AATCATGA	soil	AV	AV
S14A	AATCCATA	soil	RF	AV
S14B	AATCCGAA	soil	RF	AV
S14C	AATCCTCA	soil	RF	AV
S15A	AATCGAGA	soil	AV	RF
S15B	ACAATTGA	soil	AV	RF
S15C	AATCGCTA	soil	AV	RF
S16A	ACCAACGA	soil	RF	RF
S16B	ACCAAGTA	soil	RF	RF
S16C	ACACCAGA	soil	RF	RF
S17	AATCGTAA	wormN	Nat	RF
S18	AATCTACA	wormN	Nat	RF
S19	AATCTCGA	wormN	Nat	RF
S20	AATCTGTA	wormN	Nat	AV
S21	AATGACCA	wormN	Nat	AV
S22	AATGAGGA	wormN	Nat	AV
S23	ACCACGGA	soilN	Nat	AV
S24	ACACGACA	soilN	Nat	RF

Appendix 1c - Chapter 5 (LRP) Sample Descriptions

Forward Linker Primer Sequence: CCTACGGGAGGCAGCAG

SampleID	BarcodeSequence	Site	Status
s001	CGTAACGTAATG	DGC_polluted	Polluted
s002	ATGGACCTAGCT	DGC_polluted	Polluted
s004	GCGGTACTACTA	DGC_polluted	Polluted
s006	AGTGACTGTCAA	DGC_polluted	Polluted
s007	ATTAAGCCTGGA	DGC_polluted	Polluted
s008	TGACGCCTCCAA	DGC_polluted	Polluted
s009	CAGGCATAACAT	DGC_polluted	Polluted
s010	GTGTACATAACG	DGC_polluted	Polluted

s011	GCTTGCCAATCG	DGC_polluted	Polluted
s012	CAGCAGTCTTCG	DGC_control	Control
s013	AGGCACAGTAGG	DGC_control	Control
s014	CCTGACACACAC	DGC_control	Control
s015	CTAACTGACGCA	DGC_control	Control
s017	CGCAGATTAGTA	DGC_control	Control
s019	AGGTACGCAATT	DGC_control	Control
s021	ACGACCTACGCT	PTB_polluted	Polluted
s022	GACAATTCCGAA	PTB_polluted	Polluted
s023	AACCACTAACCG	PTB_polluted	Polluted
s024	CCAGAAGTGTTC	PTB_polluted	Polluted
s025	CTCGAGCGTACT	PTB_polluted	Polluted
s026	GATTGAACGCTA	PTB_polluted	Polluted
s030	CTGACCGTTAAG	PTB_polluted	Polluted
s033	CGCCACGTGTAT	PTB_control	Control
s034	CCAACGTAACCA	PTB_control	Control
s035	ACTAATACGCGA	PTB_control	Control
s036	GAGCATTACATG	PTB_control	Control
s037	AATTAGGCGTGT	PTB_control	Control
s040	TCATACAGCCAG	SHP_polluted	Polluted
s041	AAGCACGTCTCA	SHP_polluted	Polluted
s042	CGCTAATCGTGA	SHP_polluted	Polluted
s044	CATCAAGCATAG	SHP_polluted	Polluted
s045	TCCATACCGGAA	SHP_polluted	Polluted
s046	ATCTGACATCGG	SHP_polluted	Polluted
s047	TGCATCGCGTCA	SHP_polluted	Polluted
s049	CGTAACGTAATG	SHP_polluted	Polluted
s050	ATGGACCTAGCT	SHP_polluted	Polluted
s053	GTCATAAGAACC	SHP_control	Control
s054	GCGGTACTACTA	SHP_control	Control
s055	ATTAAGCCTGGA	SHP_control	Control
s061	CGCAGATTAGTA	SCT_polluted	Polluted
s063	ATACAGCATACG	SCT_polluted	Polluted
s064	ATGCTAACCACG	SCT_polluted	Polluted
s065	CACGATGGTCAT	SCT_polluted	Polluted
s067	GGAGTTGAGGTG	SCT_polluted	Polluted
s068	CATACCGTGAGT	SCT_polluted	Polluted
s069	ACGACCTACGCT	SCT_polluted	Polluted
s070	AACCACTAACCG	SCT_control	Control
s071	CCAGAAGTGTTC	SCT_control	Control
s072	CTCGAGCGTACT	SCT_control	Control
s073	GATTGAACGCTA	SCT_control	Control
s074	CTTGACGAGGTT	SCT_control	Control
s077	CGCGCAAGTATT	SCT_control	Control

s078	ACAGACGACGGA	SCT_control	Control
s079	AGCTTCGACAGT	SCT_control	Control
s080	GCACACCTGATA	CWM_polluted	Polluted
s083	ACTAATACGCGA	CWM_polluted	Polluted
s088	CCTACATGAGAC	CWM_polluted	Polluted
s091	ACCGAACAATCC	CWM_polluted	Polluted
s092	CGCTAATCGTGA	CWM_control	Control
s094	CATCAAGCATAG	CWM_control	Control
s095	TCAGCGCCGTTA	CWM_control	Control
s096	TCCATACCGGAA	CWM_control	Control
s099	GATGTATGTGGT	CWM_control	Control
s100	ATGGACCTAGCT	CWM_control	Control
s101	GTCATAAGAACC	CWM_control	Control
s103	CGTGCCGCTTAA	SND_ECN	ECN
s104	CCTTATAGAAGG	SND_ECN	ECN
s105	AGTGACTGTCAA	SND_ECN	ECN
s106	TGACGCCTCCAA	SND_ECN	ECN
s107	CAGGCATAACAT	SND_ECN	ECN
s109	ATGTAATAGGCC	SND_ECN	ECN
s110	CAGCAGTCTTCG	SND_ECN	ECN
s112	CTAACTGACGCA	SND_ECN	ECN
s113	ATACAGCATACG	AHT_ECN	ECN
s115	AGGTACGCAATT	AHT_ECN	ECN
s116	AACGACACGCTT	AHT_ECN	ECN
s117	CAAGCAGGTGAG	AHT_ECN	ECN
s118	CAACCACTCGGT	AHT_ECN	ECN
s119	AGTCAATGGCCT	AHT_ECN	ECN
s120	AGGAACCAGACG	AHT_ECN	ECN
s121	GGAGTTGAGGTG	AHT_ECN	ECN
s123	ACGACCTACGCT	AHT_ECN	ECN
s124	GTACCGTTGCAA	AHT_ECN	ECN
s125	GATTGAACGCTA	PDW_ECN	ECN
s126	TCCTGAACACAG	PDW_ECN	ECN
s129	CGGCGCATTATA	PDW_ECN	ECN
s130	CTGACCGTTAAG	PDW_ECN	ECN
s132	AGCTTCGACAGT	PDW_ECN	ECN
s133	GCACACCTGATA	PDW_ECN	ECN
s134	TCTTGCGGAGTC	PDW_ECN	ECN
s136	ACCACACGTAGT	DRA_ECN	ECN
s138	ACTAATACGCGA	DRA_ECN	ECN
s139	AATTAGGCGTGT	DRA_ECN	ECN
s140	TAATCGGTGCCA	DRA_ECN	ECN
s143	CCTACATGAGAC	DRA_ECN	ECN
s144	TGCTCCGTAGAA	DRA_ECN	ECN

s145	ACCGAACAATCC	DRA_ECN	ECN
s146	CGCTAATCGTGA	DRA_ECN	ECN
s149	GATGTATGTGGT	AMT_polluted2	Polluted
s150	CAATACGACCGT	AMT_polluted2	Polluted
s151	ATGGACCTAGCT	AMT_polluted2	Polluted
s156	GCGGTACTACTA	AMT_polluted2	Polluted
s157	CTGCAAGCCTGT	AMT_polluted2	Polluted
s158	CGTGCCGCTTAA	AMT_polluted2	Polluted
s168	CTAACTGACGCA	AMT_control	Control
s171	AGGTACGCAATT	SHP_control	Control
s173	AGGAACCAGACG	SHP_control	Control
s174	GGAGTTGAGGTG	SHP_control	Control
s175	ACTCACAGGAAT	SHP_control	Control
s176	AACCACTAACCG	DGC_polluted	Polluted
s177	CCAGAAGTGTTC	DGC_polluted	Polluted
s179	TCCTGAACACAG	PTB_polluted	Polluted
s180	CTTGACGAGGTT	PTB_polluted	Polluted
s187	ACCGAACAATCC	AMT_polluted1	Polluted
s189	CGCTAATCGTGA	AMT_polluted1	Polluted
s191	CATCAAGCATAG	AMT_polluted1	Polluted
s192	TCCATACCGGAA	AMT_polluted1	Polluted
s193	TGCATCGCGTCA	AMT_polluted1	Polluted

Appendix 1d - Chapter 6 (VXL) Sample Description

Forward Linker Primer Sequence: TATGGTAATTGGCCTACGGGAGGCAGCAG

Reverse Linker Primer Sequence: AGTCAGTCAGGGCCGTCAATTCMTTTRAGT

SampleID	Worm Pos	Sample no	Type	Worm Sample type	Barcode Sequence	Reverse barcode
S001	1A	1	Depurated.worm	Anterior	ATCGTACG	AACTCTCG
S002	2B	1	Depurated.worm	Body.Wall	ATCGTACG	ACTATGTC
S003	3B	1	Depurated.worm	Body.Wall	ATCGTACG	AGTAGCGT
S004	4B	1	Depurated.worm	Body.Wall	ATCGTACG	CAGTGAGT
S005	5B	1	Depurated.worm	Body.Wall	ATCGTACG	CGTACTCA
S006	6B	1	Depurated.worm	Body.Wall	ATCGTACG	CTACGCAG
S007	7B	1	Depurated.worm	Body.Wall	ATCGTACG	GGAGACTA
S008	8B	1	Depurated.worm	Body.Wall	ATCGTACG	GTCGCTCG
S009	14B	1	Depurated.worm	Body.Wall	ATCGTACG	GTCGTAGT
S010	2G	1	Depurated.worm	Gut	ATCGTACG	TAGCAGAC
S011	3G	1	Depurated.worm	Gut	ATCGTACG	TCATAGAC
S012	4G	1	Depurated.worm	Gut	ATCGTACG	TCGCTATA
S013	5G	1	Depurated.worm	Gut	ACTATCTG	AACTCTCG

S014	6G	1	Depurated.worm	Gut	ACTATCTG	ACTATGTC
S015	7G	1	Depurated.worm	Gut	ACTATCTG	AGTAGCGT
S016	8G	1	Depurated.worm	Gut	ACTATCTG	CAGTGAGT
S017	14G	1	Depurated.worm	Gut	ACTATCTG	CGTACTCA
S025	1A	2	Depurated.worm	Anterior	TAGCGAGT	AACTCTCG
S026	2B	2	Depurated.worm	Body.Wall	TAGCGAGT	ACTATGTC
S027	3B	2	Depurated.worm	Body.Wall	TAGCGAGT	AGTAGCGT
S028	4B	2	Depurated.worm	Body.Wall	TAGCGAGT	CAGTGAGT
S029	5B	2	Depurated.worm	Body.Wall	TAGCGAGT	CGTACTCA
S030	6B	2	Depurated.worm	Body.Wall	TAGCGAGT	CTACGCAG
S031	7B	2	Depurated.worm	Body.Wall	TAGCGAGT	GGAGACTA
S032	8B	2	Depurated.worm	Body.Wall	TAGCGAGT	GTCGCTCG
S033	9B	2	Depurated.worm	Body.Wall	TAGCGAGT	GTCGTAGT
S034	10B	2	Depurated.worm	Body.Wall	TAGCGAGT	TAGCAGAC
S035	11B	2	Depurated.worm	Body.Wall	TAGCGAGT	TCATAGAC
S036	14B	2	Depurated.worm	Body.Wall	TAGCGAGT	TCGCTATA
S037	2G	2	Depurated.worm	Gut	CTGCGTGT	AACTCTCG
S038	3G	2	Depurated.worm	Gut	CTGCGTGT	ACTATGTC
S039	4G	2	Depurated.worm	Gut	CTGCGTGT	AGTAGCGT
S040	5G	2	Depurated.worm	Gut	CTGCGTGT	CAGTGAGT
S041	6G	2	Depurated.worm	Gut	CTGCGTGT	CGTACTCA
S042	7G	2	Depurated.worm	Gut	CTGCGTGT	CTACGCAG
S043	8G	2	Depurated.worm	Gut	CTGCGTGT	GGAGACTA
S044	9G	2	Depurated.worm	Gut	CTGCGTGT	GTCGCTCG
S045	10G	2	Depurated.worm	Gut	CTGCGTGT	GTCGTAGT
S046	11G	2	Depurated.worm	Gut	CTGCGTGT	TAGCAGAC
S047	14G	2	Depurated.worm	Gut	CTGCGTGT	TCATAGAC
S049	1A	3	Depurated.worm	Anterior	TCATCGAG	AACTCTCG
S050	2B	3	Depurated.worm	Body.Wall	TCATCGAG	ACTATGTC
S051	3B	3	Depurated.worm	Body.Wall	TCATCGAG	AGTAGCGT
S052	4B	3	Depurated.worm	Body.Wall	TCATCGAG	CAGTGAGT
S053	5B	3	Depurated.worm	Body.Wall	TCATCGAG	CGTACTCA
S054	6B	3	Depurated.worm	Body.Wall	TCATCGAG	CTACGCAG
S055	7B	3	Depurated.worm	Body.Wall	TCATCGAG	GGAGACTA
S056	8B	3	Depurated.worm	Body.Wall	TCATCGAG	GTCGCTCG
S057	9B	3	Depurated.worm	Body.Wall	TCATCGAG	GTCGTAGT
S058	10B	3	Depurated.worm	Body.Wall	TCATCGAG	TAGCAGAC
S059	3G	3	Depurated.worm	Gut	TCATCGAG	TCATAGAC
S060	14B	3	Depurated.worm	Body.Wall	TCATCGAG	TCGCTATA
S061	4G	3	Depurated.worm	Gut	CGTGAGTG	AACTCTCG
S062	5G	3	Depurated.worm	Gut	CGTGAGTG	ACTATGTC
S063	6G	3	Depurated.worm	Gut	CGTGAGTG	AGTAGCGT
S064	7G	3	Depurated.worm	Gut	CGTGAGTG	CAGTGAGT
S065	8G	3	Depurated.worm	Gut	CGTGAGTG	CGTACTCA

S066	9G	3	Depurated.worm	Gut	CGTGAGTG	CTACGCAG
S067	10G	3	Depurated.worm	Gut	CGTGAGTG	GGAGACTA
S068	14G	3	Depurated.worm	Gut	CGTGAGTG	GTCGCTCG
S073	1A	4	Depurated.worm	Anterior	GGATATCT	AACTCTCG
S074	2B	4	Depurated.worm	Body.Wall	GGATATCT	ACTATGTC
S075	3B	4	Depurated.worm	Body.Wall	GGATATCT	AGTAGCGT
S076	4B	4	Depurated.worm	Body.Wall	GGATATCT	CAGTGAGT
S077	5B	4	Depurated.worm	Body.Wall	GGATATCT	CGTACTCA
S078	6B	4	Depurated.worm	Body.Wall	GGATATCT	CTACGCAG
S079	7B	4	Depurated.worm	Body.Wall	GGATATCT	GGAGACTA
S080	8B	4	Depurated.worm	Body.Wall	GGATATCT	GTCGCTCG
S081	9B	4	Depurated.worm	Body.Wall	GGATATCT	GTCGTAGT
S082	10B	4	Depurated.worm	Body.Wall	GGATATCT	TAGCAGAC
S083	11B	4	Depurated.worm	Body.Wall	GGATATCT	TCATAGAC
S084	14B	4	Depurated.worm	Body.Wall	GGATATCT	TCGCTATA
S085	2G	4	Depurated.worm	Gut	GACACCGT	AACTCTCG
S086	3G	4	Depurated.worm	Gut	GACACCGT	ACTATGTC
S087	4G	4	Depurated.worm	Gut	GACACCGT	AGTAGCGT
S088	5G	4	Depurated.worm	Gut	GACACCGT	CAGTGAGT
S089	6G	4	Depurated.worm	Gut	GACACCGT	CGTACTCA
S090	8G	4	Depurated.worm	Gut	GACACCGT	CTACGCAG
S091	9G	4	Depurated.worm	Gut	GACACCGT	GGAGACTA
S092	10G	4	Depurated.worm	Gut	GACACCGT	GTCGCTCG
S093	11G	4	Depurated.worm	Gut	GACACCGT	GTCGTAGT
S094	14G	4	Depurated.worm	Gut	GACACCGT	TAGCAGAC
S095	7G	4	Depurated.worm	Gut	GACACCGT	TCATAGAC
S097	1A	5	Depurated.worm	Anterior	CTACTATA	AACTCTCG
S098	2B	5	Depurated.worm	Body.Wall	CTACTATA	ACTATGTC
S099	3B	5	Depurated.worm	Body.Wall	CTACTATA	AGTAGCGT
S100	4B	5	Depurated.worm	Body.Wall	CTACTATA	CAGTGAGT
S101	5B	5	Depurated.worm	Body.Wall	CTACTATA	CGTACTCA
S102	6B	5	Depurated.worm	Body.Wall	CTACTATA	CTACGCAG
S103	7B	5	Depurated.worm	Body.Wall	CTACTATA	GGAGACTA
S104	8B	5	Depurated.worm	Body.Wall	CTACTATA	GTCGCTCG
S105	9B	5	Depurated.worm	Body.Wall	CTACTATA	GTCGTAGT
S106	10B	5	Depurated.worm	Body.Wall	CTACTATA	TAGCAGAC
S107	11B	5	Depurated.worm	Body.Wall	CTACTATA	TCATAGAC
S108	NA	3d	Post.exp.soil	na	CTACTATA	TCGCTATA
S109	3G	5	Depurated.worm	Gut	CGTTACTA	AACTCTCG
S110	4G	5	Depurated.worm	Gut	CGTTACTA	ACTATGTC
S111	5G	5	Depurated.worm	Gut	CGTTACTA	AGTAGCGT
S112	6G	5	Depurated.worm	Gut	CGTTACTA	CAGTGAGT
S113	7G	5	Depurated.worm	Gut	CGTTACTA	CGTACTCA
S114	8G	5	Depurated.worm	Gut	CGTTACTA	CTACGCAG

S115	9G	5	Depurated.worm	Gut	CGTTACTA	GGAGACTA
S116	10G	5	Depurated.worm	Gut	CGTTACTA	GTCGCTCG
S117	11G	5	Depurated.worm	Gut	CGTTACTA	GTCGTAGT
S118	14G	5	Depurated.worm	Gut	CGTTACTA	TAGCAGAC
S119	14B	5	Depurated.worm	Body.Wall	CGTTACTA	TCATAGAC
S120	NA	3e	Post.exp.soil	na	CGTTACTA	TCGCTATA
S121	NA	1a	Post.exp.soil	na	AGAGTCAC	AACTCTCG
S122	NA	1b	Post.exp.soil	na	AGAGTCAC	ACTATGTC
S123	NA	1c	Post.exp.soil	na	AGAGTCAC	AGTAGCGT
S124	NA	2a	Post.exp.soil	na	AGAGTCAC	CAGTGAGT
S125	NA	2b	Post.exp.soil	na	AGAGTCAC	CGTACTCA
S126	NA	2c	Post.exp.soil	na	AGAGTCAC	CTACGCAG
S127	NA	3a	Post.exp.soil	na	AGAGTCAC	GGAGACTA
S128	NA	3b	Post.exp.soil	na	AGAGTCAC	GTCGCTCG
S129	NA	3c	Post.exp.soil	na	AGAGTCAC	GTCGTAGT
S130	3B	1	Fresh.worm	Body.Wall	AGAGTCAC	TAGCAGAC
S132	NA	3f	Post.exp.soil	na	AGAGTCAC	TCGCTATA
S133	NA	1a	Casts	na	TACGAGAC	AACTCTCG
S134	NA	1b	Casts	na	TACGAGAC	ACTATGTC
S135	NA	1c	Casts	na	TACGAGAC	AGTAGCGT
S136	NA	2a	Casts	na	TACGAGAC	CAGTGAGT
S137	NA	2b	Casts	na	TACGAGAC	CGTACTCA
S138	NA	2c	Casts	na	TACGAGAC	CTACGCAG
S139	NA	3a	Casts	na	TACGAGAC	GGAGACTA
S140	NA	3b	Casts	na	TACGAGAC	GTCGCTCG
S141	NA	3c	Casts	na	TACGAGAC	GTCGTAGT
S144	NA	1d	Casts	na	TACGAGAC	TCGCTATA
S145	1A	1	Fresh.worm	Anterior	ACGTCTCG	AACTCTCG
S146	2G	1	Fresh.worm	Gut	ACGTCTCG	ACTATGTC
S147	3G	1	Fresh.worm	Gut	ACGTCTCG	AGTAGCGT
S148	4G	1	Fresh.worm	Gut	ACGTCTCG	CAGTGAGT
S149	5G	1	Fresh.worm	Gut	ACGTCTCG	CGTACTCA
S150	6G	1	Fresh.worm	Gut	ACGTCTCG	CTACGCAG
S151	7G	1	Fresh.worm	Gut	ACGTCTCG	GGAGACTA
S152	8G	1	Fresh.worm	Gut	ACGTCTCG	GTCGCTCG
S153	9G	1	Fresh.worm	Gut	ACGTCTCG	GTCGTAGT
S154	14G	1	Fresh.worm	Gut	ACGTCTCG	TAGCAGAC
S155	2B	1	Fresh.worm	Body.Wall	ACGTCTCG	TCATAGAC
S156	NA	1e	Casts	na	ACGTCTCG	TCGCTATA
S159	4B	1	Fresh.worm	Body.Wall	TCGACGAG	AGTAGCGT
S160	5B	1	Fresh.worm	Body.Wall	TCGACGAG	CAGTGAGT
S161	6B	1	Fresh.worm	Body.Wall	TCGACGAG	CGTACTCA
S162	7B	1	Fresh.worm	Body.Wall	TCGACGAG	CTACGCAG
S163	8B	1	Fresh.worm	Body.Wall	TCGACGAG	GGAGACTA

S164	9B	1	Fresh.worm	Body.Wall	TCGACGAG	GTCGCTCG
S165	14B	1	Fresh.worm	Body.Wall	TCGACGAG	GTCGTAGT
S167	14B	2	Fresh.worm	Body.Wall	TCGACGAG	TCATAGAC
S168	NA	1d	Origin.soil	na	TCGACGAG	TCGCTATA
S169	1A	2	Fresh.worm	Anterior	GATCGTGT	AACTCTCG
S170	2B	2	Fresh.worm	Body.Wall	GATCGTGT	ACTATGTC
S171	3B	2	Fresh.worm	Body.Wall	GATCGTGT	AGTAGCGT
S172	4B	2	Fresh.worm	Body.Wall	GATCGTGT	CAGTGAGT
S173	5B	2	Fresh.worm	Body.Wall	GATCGTGT	CGTACTCA
S174	6B	2	Fresh.worm	Body.Wall	GATCGTGT	CTACGCAG
S175	7B	2	Fresh.worm	Body.Wall	GATCGTGT	GGAGACTA
S176	8B	2	Fresh.worm	Body.Wall	GATCGTGT	GTCGCTCG
S177	9B	2	Fresh.worm	Body.Wall	GATCGTGT	GTCGTAGT
S178	10B	2	Fresh.worm	Body.Wall	GATCGTGT	TAGCAGAC
S179	11B	2	Fresh.worm	Body.Wall	GATCGTGT	TCATAGAC
S180	NA	1e	Origin.soil	na	GATCGTGT	TCGCTATA
S181	2G	2	Fresh.worm	Gut	GTCAGATA	AACTCTCG
S182	3G	2	Fresh.worm	Gut	GTCAGATA	ACTATGTC
S183	4G	2	Fresh.worm	Gut	GTCAGATA	AGTAGCGT
S185	5G	2	Fresh.worm	Gut	GTCAGATA	CGTACTCA
S186	6G	2	Fresh.worm	Gut	GTCAGATA	CTACGCAG
S187	7G	2	Fresh.worm	Gut	GTCAGATA	GGAGACTA
S188	8G	2	Fresh.worm	Gut	GTCAGATA	GTCGCTCG
S189	9G	2	Fresh.worm	Gut	GTCAGATA	GTCGTAGT
S190	10G	2	Fresh.worm	Gut	GTCAGATA	TAGCAGAC
S191	11G	2	Fresh.worm	Gut	GTCAGATA	TCATAGAC
S192	NA	1f	Origin.soil	na	GTCAGATA	TCGCTATA
S193	1A	3	Fresh.worm	Anterior	CTACTATA	AAGTCGAG
S194	2B	3	Fresh.worm	Body.Wall	CTACTATA	ATACTTCG
S195	3B	3	Fresh.worm	Body.Wall	CTACTATA	AGCTGCTA
S196	4B	3	Fresh.worm	Body.Wall	CTACTATA	CATAGAGA
S197	5B	3	Fresh.worm	Body.Wall	CTACTATA	CGTAGATC
S198	6B	3	Fresh.worm	Body.Wall	CTACTATA	CTCGTTAC
S199	7B	3	Fresh.worm	Body.Wall	CTACTATA	GCGCACGT
S200	8B	3	Fresh.worm	Body.Wall	CTACTATA	GGTACTAT
S201	9B	3	Fresh.worm	Body.Wall	CTACTATA	GTATACGC
S202	10B	3	Fresh.worm	Body.Wall	CTACTATA	TACGAGCA
S203	14B	3	Fresh.worm	Body.Wall	CTACTATA	TCAGCGTT
S204	2G	3	Fresh.worm	Gut	CTACTATA	TCGCTACG
S205	3G	3	Fresh.worm	Gut	CGTTACTA	AAGTCGAG
S206	4G	3	Fresh.worm	Gut	CGTTACTA	ATACTTCG
S207	5G	3	Fresh.worm	Gut	CGTTACTA	AGCTGCTA
S208	6G	3	Fresh.worm	Gut	CGTTACTA	CATAGAGA
S209	7G	3	Fresh.worm	Gut	CGTTACTA	CGTAGATC

S210	8G	3	Fresh.worm	Gut	CGTTACTA	CTCGTTAC
S211	9G	3	Fresh.worm	Gut	CGTTACTA	GCGCACGT
S212	10G	3	Fresh.worm	Gut	CGTTACTA	GGTACTAT
S213	14G	3	Fresh.worm	Gut	CGTTACTA	GTATACGC
S216	14B	4	Fresh.worm	Body.Wall	CGTTACTA	TCGCTACG
S217	1A	4	Fresh.worm	Anterior	AGAGTCAC	AAGTCGAG
S218	2B	4	Fresh.worm	Body.Wall	AGAGTCAC	ATACTTCG
S219	3B	4	Fresh.worm	Body.Wall	AGAGTCAC	AGCTGCTA
S220	4B	4	Fresh.worm	Body.Wall	AGAGTCAC	CATAGAGA
S221	5B	4	Fresh.worm	Body.Wall	AGAGTCAC	CGTAGATC
S222	6B	4	Fresh.worm	Body.Wall	AGAGTCAC	CTCGTTAC
S223	7B	4	Fresh.worm	Body.Wall	AGAGTCAC	GCGCACGT
S224	8B	4	Fresh.worm	Body.Wall	AGAGTCAC	GGTACTAT
S225	9B	4	Fresh.worm	Body.Wall	AGAGTCAC	GTATACGC
S226	10B	4	Fresh.worm	Body.Wall	AGAGTCAC	TACGAGCA
S227	11B	4	Fresh.worm	Body.Wall	AGAGTCAC	TCAGCGTT
S228	12B	4	Fresh.worm	Body.Wall	AGAGTCAC	TCGCTACG
S229	14B	4	Fresh.worm	Body.Wall	TACGAGAC	AAGTCGAG
S230	2G	4	Fresh.worm	Gut	TACGAGAC	ATACTTCG
S231	3G	4	Fresh.worm	Gut	TACGAGAC	AGCTGCTA
S232	4G	4	Fresh.worm	Gut	TACGAGAC	CATAGAGA
S233	5G	4	Fresh.worm	Gut	TACGAGAC	CGTAGATC
S234	6G	4	Fresh.worm	Gut	TACGAGAC	CTCGTTAC
S235	7G	4	Fresh.worm	Gut	TACGAGAC	GCGCACGT
S236	8G	4	Fresh.worm	Gut	TACGAGAC	GGTACTAT
S237	9G	4	Fresh.worm	Gut	TACGAGAC	GTATACGC
S238	10G	4	Fresh.worm	Gut	TACGAGAC	TACGAGCA
S239	11G	4	Fresh.worm	Gut	TACGAGAC	TCAGCGTT
S240	14G	4	Fresh.worm	Gut	TACGAGAC	TCGCTACG
S252	NA	1f	Post.exp.soil	na	ACGTCTCG	TCGCTACG
S264	NA	2d	Post.exp.soil	na	TCGACGAG	TCGCTACG
S265	NA	1a	Origin.soil	na	GATCGTGT	AAGTCGAG
S266	NA	1b	Origin.soil	na	GATCGTGT	ATACTTCG
S267	NA	1c	Origin.soil	na	GATCGTGT	AGCTGCTA
S268	NA	2a	Origin.soil	na	GATCGTGT	CATAGAGA
S269	NA	2b	Origin.soil	na	GATCGTGT	CGTAGATC
S270	NA	2c	Origin.soil	na	GATCGTGT	CTCGTTAC
S271	NA	3a	Origin.soil	na	GATCGTGT	GCGCACGT
S272	NA	3b	Origin.soil	na	GATCGTGT	GGTACTAT
S273	NA	3c	Origin.soil	na	GATCGTGT	GTATACGC
S274	NA	13	Cocoon	na	GATCGTGT	TACGAGCA
S275	NA	1d	Post.exp.soil	na	GATCGTGT	TCAGCGTT
S276	NA	2e	Post.exp.soil	na	GATCGTGT	TCGCTACG
S277	NA	1	Cocoon	na	GTCAGATA	AAGTCGAG

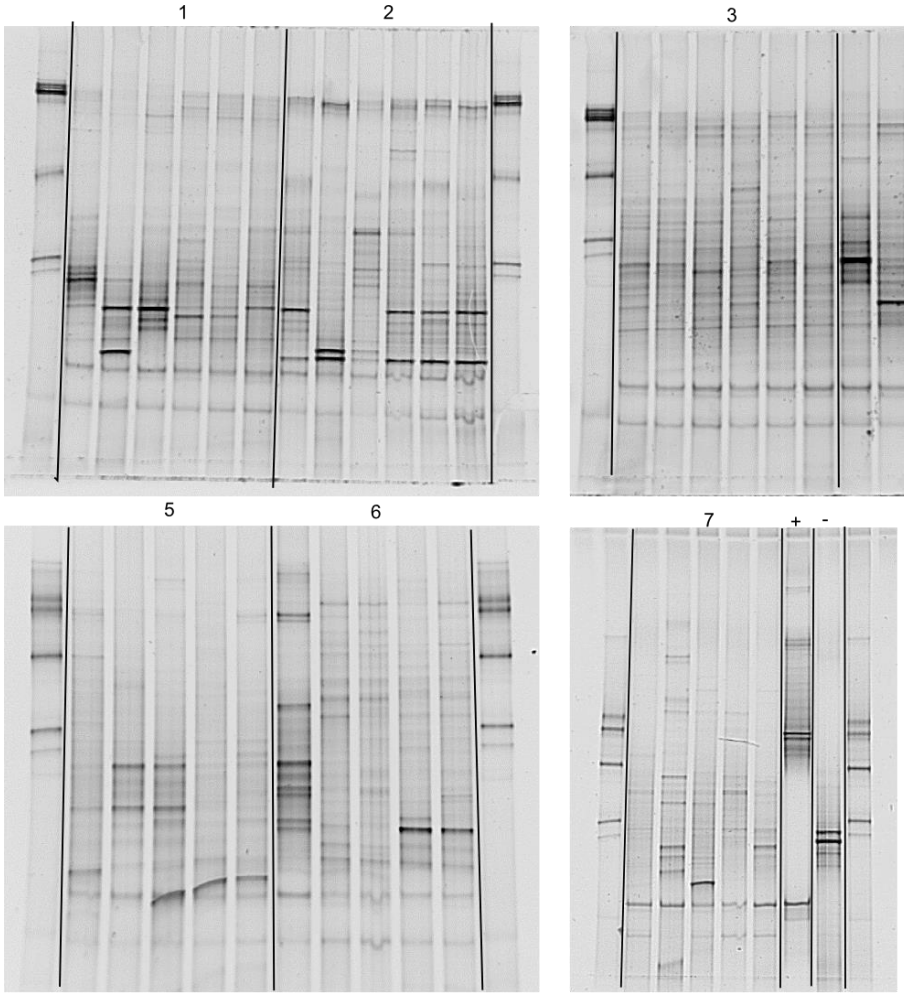
S278	NA	2	Cocoon	na	GTCAGATA	ATACTTCG
S279	NA	3	Cocoon	na	GTCAGATA	AGCTGCTA
S280	NA	4	Cocoon	na	GTCAGATA	CATAGAGA
S281	NA	5	Cocoon	na	GTCAGATA	CGTAGATC
S282	NA	6	Cocoon	na	GTCAGATA	CTCGTTAC
S283	NA	7	Cocoon	na	GTCAGATA	GCGCACGT
S284	NA	8	Cocoon	na	GTCAGATA	GGTACTAT
S285	NA	9	Cocoon	na	GTCAGATA	GTATACGC
S286	NA	10	Cocoon	na	GTCAGATA	TACGAGCA
S287	NA	1e	Post.exp.soil	na	GTCAGATA	TCAGCGTT
S288	NA	2f	Post.exp.soil	na	GTCAGATA	TCGCTACG
S289	1A	7	Fresh.worm	Anterior	ATCGTACG	AAGTCGAG
S290	9B	7	Fresh.worm	Body.Wall	ATCGTACG	ATACTTCG
S291	6G	7	Fresh.worm	Gut	ATCGTACG	AGCTGCTA
S292	NA	1a	Post.exp.soil	na	ATCGTACG	CATAGAGA
S293	NA	3c	Post.exp.soil	na	ATCGTACG	CGTAGATC
S294	NA	1a	Origin.soil	na	ATCGTACG	CTCGTTAC
S295	NA	3c	Origin.soil	na	ATCGTACG	GCGCACGT
S296	1A	4	Depurated.worm	Anterior	ATCGTACG	GGTACTAT
S297	9G	2	Fresh.worm	Gut	ATCGTACG	GTATACGC
S298	1A	2	Depurated.worm	Anterior	ATCGTACG	TACGAGCA
S299	14G	3	Depurated.worm	Gut	ATCGTACG	TCAGCGTT
S300	8G	5	Depurated.worm	Gut	ATCGTACG	TCGCTACG
S301	2B	7	Fresh.worm	Body.Wall	ACTATCTG	AAGTCGAG
S302	10B	7	Fresh.worm	Body.Wall	ACTATCTG	ATACTTCG
S303	7G	7	Fresh.worm	Gut	ACTATCTG	AGCTGCTA
S304	NA	1b	Post.exp.soil	na	ACTATCTG	CATAGAGA
S305	NA	1a	Casts	na	ACTATCTG	CGTAGATC
S306	NA	1b	Origin.soil	na	ACTATCTG	CTCGTTAC
S307	NA	1	Cocoon	na	ACTATCTG	GCGCACGT
S308	8G	1	Depurated.worm	Gut	ACTATCTG	GGTACTAT
S309	10G	2	Fresh.worm	Gut	ACTATCTG	GTATACGC
S310	5G	3	Depurated.worm	Gut	ACTATCTG	TACGAGCA
S311	9B	3	Depurated.worm	Body.Wall	ACTATCTG	TCAGCGTT
S312	7B	5	Depurated.worm	Body.Wall	ACTATCTG	TCGCTACG
S313	3B	7	Fresh.worm	Body.Wall	TAGCGAGT	AAGTCGAG
S314	14B	7	Fresh.worm	Body.Wall	TAGCGAGT	ATACTTCG
S315	8G	7	Fresh.worm	Gut	TAGCGAGT	AGCTGCTA
S316	NA	1c	Post.exp.soil	na	TAGCGAGT	CATAGAGA
S317	NA	1b	Casts	na	TAGCGAGT	CGTAGATC
S318	NA	1c	Origin.soil	na	TAGCGAGT	CTCGTTAC
S319	NA	3	Cocoon	na	TAGCGAGT	GCGCACGT
S320	4B	3	Depurated.worm	Body.Wall	TAGCGAGT	GGTACTAT
S321	10G	4	Fresh.worm	Gut	TAGCGAGT	GTATACGC

S322	6G	3	Depurated.worm	Gut	TAGCGAGT	TACGAGCA
S323	1A	5	Depurated.worm	Anterior	TAGCGAGT	TCAGCGTT
S324	9G	5	Depurated.worm	Gut	TAGCGAGT	TCGCTACG
S325	4B	7	Fresh.worm	Body.Wall	CTGCGTGT	AAGTCGAG
S327	9G	7	Fresh.worm	Gut	CTGCGTGT	AGCTGCTA
S328	NA	2a	Post.exp.soil	na	CTGCGTGT	CATAGAGA
S329	NA	1c	Casts	na	CTGCGTGT	CGTAGATC
S330	NA	2a	Origin.soil	na	CTGCGTGT	CTCGTTAC
S331	NA	5	Cocoon	na	CTGCGTGT	GCGCACGT
S332	7G	3	Depurated.worm	Gut	CTGCGTGT	GGTACTAT
S333	NA	3d	Origin.soil	na	CTGCGTGT	GTATACGC
S334	8G	1	Depurated.worm	Gut	CTGCGTGT	TACGAGCA
S335	NA	1a	Post.exp.soil	na	CTGCGTGT	TCAGCGTT
S336	8B	5	Depurated.worm	Body.Wall	CTGCGTGT	TCGCTACG
S337	5B	7	Fresh.worm	Body.Wall	TCATCGAG	AAGTCGAG
S338	2G	7	Fresh.worm	Gut	TCATCGAG	ATACTTCG
S339	10G	7	Fresh.worm	Gut	TCATCGAG	AGCTGCTA
S340	NA	2b	Post.exp.soil	na	TCATCGAG	CATAGAGA
S341	NA	2a	Casts	na	TCATCGAG	CGTAGATC
S342	NA	2b	Origin.soil	na	TCATCGAG	CTCGTTAC
S343	NA	6	Cocoon	na	TCATCGAG	GCGCACGT
S344	5B	4	Depurated.worm	Body.Wall	TCATCGAG	GGTACTAT
S345	NA	3e	Origin.soil	na	TCATCGAG	GTATACGC
S346	5G	2	Depurated.worm	Gut	TCATCGAG	TACGAGCA
S347	4B	5	Depurated.worm	Body.Wall	TCATCGAG	TCAGCGTT
S348	4B	3	Fresh.worm	Body.Wall	TCATCGAG	TCGCTACG
S349	6B	7	Fresh.worm	Body.Wall	CGTGAGTG	AAGTCGAG
S350	3G	7	Fresh.worm	Gut	CGTGAGTG	ATACTTCG
S351	14G	7	Fresh.worm	Gut	CGTGAGTG	AGCTGCTA
S352	NA	2c	Post.exp.soil	na	CGTGAGTG	CATAGAGA
S353	NA	3c	Casts	na	CGTGAGTG	CGTAGATC
S354	NA	2c	Origin.soil	na	CGTGAGTG	CTCGTTAC
S355	NA	7	Cocoon	na	CGTGAGTG	GCGCACGT
S356	5G	5	Depurated.worm	Gut	CGTGAGTG	GGTACTAT
S357	NA	3f	Origin.soil	na	CGTGAGTG	GTATACGC
S358	8G	2	Depurated.worm	Gut	CGTGAGTG	TACGAGCA
S359	NA	2b	Post.exp.soil	na	CGTGAGTG	TCAGCGTT
S360	5G	4	Fresh.worm	Gut	CGTGAGTG	TCGCTACG
S361	7B	7	Fresh.worm	Body.Wall	GGATATCT	AAGTCGAG
S362	4G	7	Fresh.worm	Gut	GGATATCT	ATACTTCG
S364	NA	3a	Post.exp.soil	na	GGATATCT	CATAGAGA
S366	NA	3a	Origin.soil	na	GGATATCT	CTCGTTAC
S367	NA	8	Cocoon	na	GGATATCT	GCGCACGT
S368	7G	2	Fresh.worm	Gut	GGATATCT	GGTACTAT

S370	7B	3	Depurated.worm	Body.Wall	GGATATCT	TACGAGCA
S371	5G	2	Fresh.worm	Gut	GGATATCT	TCAGCGTT
S372	11G	4	Fresh.worm	Gut	GGATATCT	TCGCTACG
S373	8B	7	Fresh.worm	Body.Wall	GACACCGT	AAGTCGAG
S374	5G	7	Fresh.worm	Gut	GACACCGT	ATACTTCG
S376	NA	3b	Post.exp.soil	na	GACACCGT	CATAGAGA
S378	NA	3b	Origin.soil	na	GACACCGT	CTCGTTAC
S380	8G	2	Fresh.worm	Gut	GACACCGT	GGTACTAT
S382	10G	3	Depurated.worm	Gut	GACACCGT	TACGAGCA
S383	6B	5	Depurated.worm	Body.Wall	GACACCGT	TCAGCGTT

Appendix 2 - Chapter 3 (DGC) Denaturing Gradient Gel Electrophoresis analysis

Appendix 2a - *Lumbricus rubellus* samples

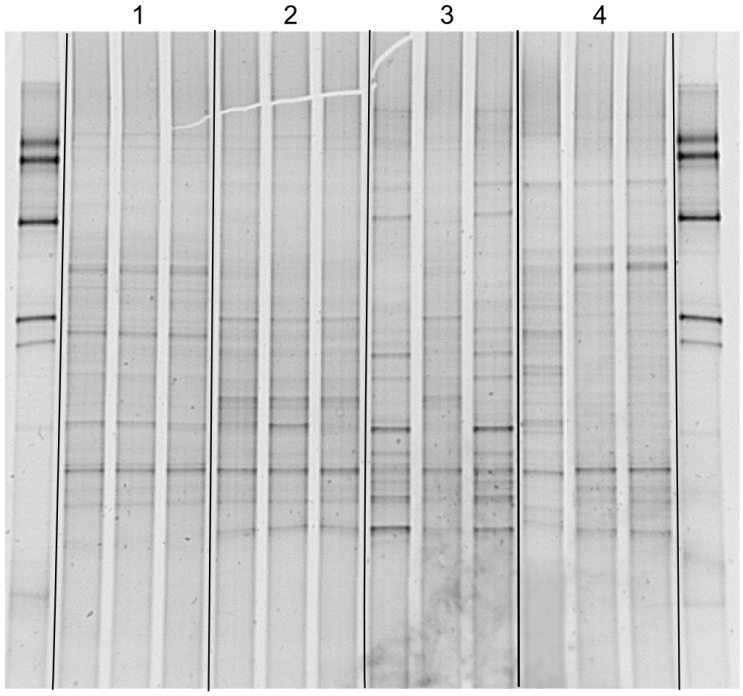


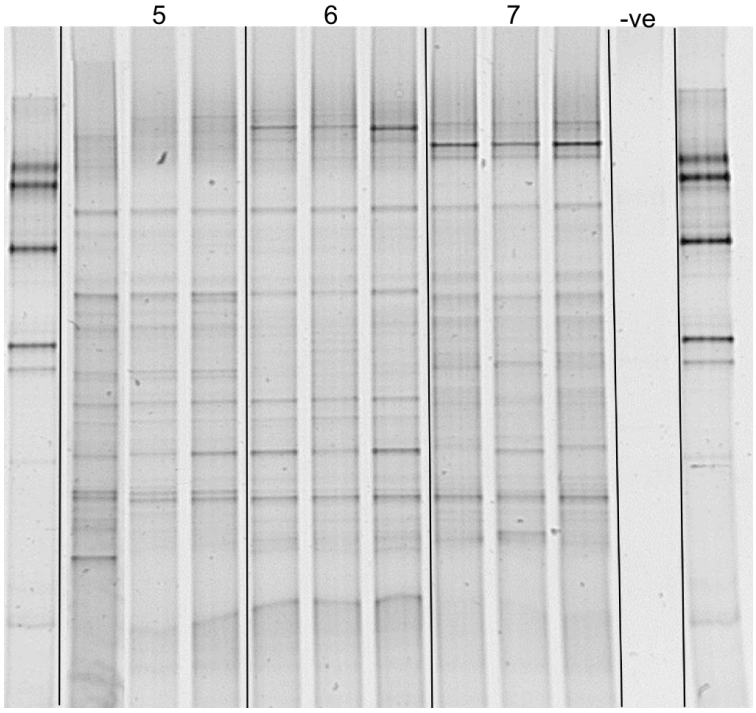
Lumbricus rubellus DGGEs

Label	Samples
1	1-6 Site 1
2	1-6 Site 2
3	1-6 Site 3
4	1-6 Site 4 (On site Control)
5	1-5 Site 5
6	1-5 Site 6
7	1-5 Site 7 (Off Site Control)
+	Positive control
-	Negative Control

DGGE analysis of *L. rubellus* samples from Devon Great Consols study (Chapter 3). Diversity of the bacterial community was confirmed for quality assessment prior to submission for 454 sequencing.

Appendix 2b - Soil samples





Soil Sample DGGEs

Label	Samples
1	1-3 Site 1
2	1-3 Site 2
3	1-3 Site 3
4	1-3 Site 4 (On site Control)
5	1-3 Site 5
6	1-3 Site 6
7	1-3 Site 7 (Off Site Control)
-	Negative Control

DGGE analysis of soil samples from Devon Great Consols study (Chapter 3) Diversity of the bacterial community was confirmed for quality assessment prior to submission for 454 sequencing