Article (refereed) - postprint

Pass, Daniel Antony; Morgan, Andrew John; Read, Daniel S.; Field, Dawn; Weightman, Andrew J.; Kille, Peter. 2015. **The effect of anthropogenic arsenic contamination on the earthworm microbiome**. *Environmental Microbiology*, 17 (6). 1884-1896. 10.1111/1462-2920.12712

The effect of anthropogenic arsenic contamination on the earthworm microbiome.

Pass, D. A.[1]*, Morgan, A.J.[1], Read, D. S. [2], Field, D.[2], Weightman, A. J.[1] and Kille, P.[1]

[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.

Subject category: Microbe-microbe and microbe-host interactions

Running Title: The Earthworm microbiome

* Corresponding author: Daniel Antony Pass. School of Biosciences, Cardiff University, P.O. Box 915, Cardiff, CF10 3TL, UK.

Email: daniel.antony.pass@gmail.com, Phone: 02920876680

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.12712

Abstract

Earthworms are globally distributed and perform essential roles for soil health and microbial structure. We have investigated the effect of an anthropogenic contamination gradient on the bacterial community of the keystone ecological species *Lumbricus rubellus* through utilising 16S rRNA pyrosequencing for the first time to establish the microbiome of the host and surrounding soil.

The earthworm-associated microbiome differs from the surrounding environment which appears to be a result of both filtering and stimulation likely linked to the altered environment associated with the gut micro-habitat (neutral pH, anoxia and increased carbon substrates). We identified a core earthworm community comprising Proteobacteria (~50%) and Actinobacteria (~30%), with lower abundances of Bacteroidetes (~6%) and Acidobacteria (~3%). In addition to the known earthworm symbiont (*Verminephrobacter sp.*) we identified a potential host-associated Gammaproteobacteria species (*Serratia sp.*) which was absent from soil yet observed in most earthworms.

Although a distinct bacterial community defines these earthworms, clear family- and species-level modification were observed along an arsenic and iron contamination gradient. Several taxa observed in uncontaminated control microbiomes are suppressed by metal/metalloid field exposure, including eradication of the hereto ubiquitously associated *Verminephrobacter* symbiont, which raises implications to its functional role in the earthworm microbiome.

Keywords: microbiome, earthworm, symbiotic, host-associated, 16S rRNA pyrosequnecing.

Introduction

In one square metre of a favourable soil environment roughly one litre of soil is contained within an earthworm population's gut where 4-10% of total soil is consumed annually (Drake & Horn 2007). Extrapolation 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 (1L M^{-2} of favourable UK soil (Barr *et al.*, 1978)) and therefore the egested material clearly represents the major constituent of soil.

Consequently, the global impact exerted by earthworms on the soil environment is vast and is integral to its microbial structure and physiochemical properties. The gut environment differs greatly from the surrounding soil as a result of a number of factors including exposure to anoxia and pH neutralisation (Drake & Horn 2007). Additionally, levels of organic carbon are higher in the gut than the surrounding soil due to the secretion of intestinal mucus producing a 'priming' effect (Brown *et al.* 2000). This can stimulate significantly an increase in the abundance of methanogenic, fermentative, and nitrate-reducing bacteria (Depkat-Jakob *et al.*, 2012, 2013). The transit time of ingested soil to eventual egestion is rapid, reported to range from 6-8 hours for *Lumbricus rubellus* (Daniel & Anderson 1992) to 2-16 hours for other earthworm species (Brown *et al.* 2000), raising the question of the extent of change which could occur in the microbial community during transit.

Host-associated microbiota is increasingly understood to contribute to an individual's phenotype. The host's impact on its microbiota and, in turn, the impact of the microbiota on the host can be observed in species at all taxonomic levels, including humans (Li *et al.*, 2008; Ley *et al.*, 2008). This 'two way street' forms the basis of the observed mutualism which can play an important role in the host organism's environmental interactions. Invertebrate examples of this mutualism include cellulose and xylan digestive processes in wood-feeding termites (Warnecke *et al.*, 2007), collagenolytic activity in *Osedax* boneworms (Goffredi *et al.*, 2007), and immune system potentiation in *Drosophila* (Teixeira *et al.*, 2008) and tsetse flies (Weiss *et al.*, 2012). The location of

such symbionts varies, including as organ-associated species (e.g. *Verminephrobacter* in earthworm species found in the nephridia (Pandazis, 1931; Schramm *et al.*, 2003)), or gut-bound structures which promote biofilm-like congregations, increasing microbial load and functional capacity (Hackstein & Stumm, 1994). A microbial community which could reduce host stress would be highly beneficial, and host-microbial symbiosis could therefore be seen as either an endpoint (i.e. an important component of the host) or a stepping-stone in invertebrate evolution which buffers the individual from external stress and enables the host population to encroach on environments otherwise inhospitable. If either, or both, hypotheses are correct this would exert strong selective pressure for the host to accommodate microbes which reduce the toxicity of environmental stressors.

Earthworm species ubiquitously host the symbiotic *Acidovorax*-like bacteria *Verminephrobacter* in the osmoregulatory nephridial organ (Pinel *et al.* 2008; Davidson *et al.* 2012) and this vertically transmitted symbiont has diversified with the specific host over significant evolutionary time (62-132 myr; Lund *et al.* (2009)). A role for *Verminephrobacter* in nitrogen and protein recovery was originally posited due to anatomical location and nephridial functionality (Pandazis 1931; Schramm *et al.* 2003); however, this has since been questioned due to an absence of extracellular proteases within the *Verminephrobacter eisinea* genome and on the analysis of aposymbiotically-reared individuals (Lund *et al.*, 2010).

Previous microbial analysis of the related earthworm species *Lumbricus terrestris* by Terminal Restriction Fragment Length Polymorphism (T-RFLP) has demonstrated highly similar microbial profiles in each 'compartment' (transient gut contents, soil, and casts (egested material)) indicative of a soil-derived microbiome (Egert *et al.*, 2004). Whilst the low resolution of T-RFLP analysis was considered a potential limiting factor, the authors concluded that an indigenous microbial community was unlikely. Later research suggests that the majority of microbial activity associated with the earthworm is likely contributed by the transient community being selectively stimulated by the unique environment encountered during transit. Wüst *et al.* (2011) described the role of the gut as an environment which encourages *Clostridia* and *Enterobacteriaceae* 'fermenter' communities through metabolism of mucus- and plant-derived saccharides

resulting in nitrogenous gas production. The earthworm *Eisinea andrei* effects a reduction in soil microbial diversity but an increase in microbial activity through action on the transient community (Gómez-Brandón *et al.*, 2011). Distinct taxonomic groups have been identified at higher abundance in *L. terrestris* and *Apporectodea caliginosa* casts, notably Bacteriodetes species (Nechitaylo *et al.*, 2010) where their role in organic matter breakdown is posited.

Earthworms are sometimes labelled ‘extremophiles’ due to regularly occupying habitats with severe geochemical gradients and high anthropogenic contamination (Morgan *et al.* 2007). The deep-burrowing earthworm species *L. terrestris* increases arsenic mobility in contaminated sites, concurrent with reduction of soil As(V) to As(III) during gut passage (Sizmur *et al.*, 2011). Genetic analysis of *L. rubellus* tolerance to arsenic has been previously undertaken (Langdon *et al.*, 2001, 2009; Kille *et al.*, 2013) suggesting a combination of genetic and epigenetic adaptive strategies. However, the host-associated microbial contribution has never been assessed. In the present study, a disused mine site with a range of arsenic contamination of up to c.x400 higher than the surrounding area was used as a ‘model’ anthropogenically stressed site. This site in the South-West of the United Kingdom has been previously characterised in terms of geochemistry and earthworm genotype (Klinck *et al.*, 2005; Kille *et al.*, 2013) and allows an *in situ* snapshot of the *Lumbricus rubellus* microbiome across a steep gradient where this extremotolerant species is commonly found. The specific aim of the present study was to elucidate both the differences between the microbial population present in the soil and that of the host, and also the impact of extreme stress on this community using High Throughput Sequencing to examine the microbiome of an ecologically-relevant earthworm species to a level of detail and resolution not previously published for any terrestrial oligochaete.

Results

The Basal Earthworm Microbiome

The observed taxonomic profiles and community structure represented 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 As mine site, were analysed and compared with the combined soil microbial composition (Figure 1). For the earthworms Proteobacteria is 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 are 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 is 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 was 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 abundance contributed 2.8% relative proportion to 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%) was 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%).

Host vs Habitat

In total 26,618 OTUs were generated at 97% homology linkage with 15,723 OTUs originating from a single sequence (singletons) after normalisation (expected with this technique due to high variability in the soil environment (Griffiths *et al.*, 2011)). Supplementary Figure 3a shows OTU generation and diversity measures at 97%, 94% and 88%.

Principal Co-ordinate Analysis of Unifrac (Lozupone & Knight 2005) distances showed bacterial communities to differ between soil and host-resident microbiota (Figure 2a).

The largest differences were phylum level shifts where relative abundance of Acidobacteria reduced, and Actinobacteria increased from soil to *L. rubellus* however Figure 2b describes the family level abundance shifts in the earthworm community for families with >100 sequences in either the host or habitat. Taxa are ordered by magnitude of difference between soil and host and indicates that large shifts can be attributed to family level changes.

Diversity and richness is summarised in Figure 3a (detailed in Supplementary Figure 4). 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*, $P<0.05$, Supplementary Figure 4a). Chao1 richness was significantly lowered in all but one site (t-test, $P<0.05$) and Observed Species was significantly reduced in 5 of 7 (Supplementary Figures 4b & 4c respectively). To assess the soil-host community differences from 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 (Supplementary Figure 5) but also highlights that a large amount of diversity is yet to be captured.

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 (Supplementary Figure 3b). 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, which were not observed in soils. Seven ‘core’ OTUs were detected at all sites in at

least one individual, and these OTUs contributed to 5.4% of all earthworm-derived reads (Supplementary Figure 3). Of these core OTUs, six 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.

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 3b, Supplementary Figure 6). 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, Supplementary Figure 4). 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 demonstrates a consistent microbial population being present in earthworms from the same site (Figure 3B) 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 sporadic presence/absence (5 sites <0.7 mg kg^{-1} Cd | 2 sites >7mg kg^{-1} Cd) may over-represent the impact.

OTUs which drive the observed variance are identified in Figure 4. 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 two clusters. (ANOVA $P<0.05$ = association, $P>0.05$ = shared (FDR correction)). Site 5 samples were omitted from OTU association calculations due to

individuals from this site being outliers. 11 of the 48 abundant OTUs 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.

Discussion

We have described how the earthworm microbiome is distinct from the surrounding soil microbial community. Notably, 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 5 visually summarises the co-occurrence of OTUs across the dataset, demonstrating that while the majority of species are shared between all samples (host and soil), there is higher shared OTU incidence between worm individuals and their site of origin. Further notable is the number of OTUs which occur solely in the earthworms and remain absent from the soil, representing host-associated species not found in abundance in the soil. 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 they concur with a later study that found the same major taxonomic groups at but at different proportions (Nechitaylo *et al.*, 2010).

We demonstrate that 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 mico-habitats where the symbiont is 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 act in nitrogen cycling (Ventura *et al.*, 2007) and *Frankia* function as facultative nitrogen-fixing symbionts in plants (Normand *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, we have found that it is highly sensitive to environmental arsenic contamination. 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 apparent 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 are more similar between sites than individual earthworms and their site-specific soil. The combinatorial effect of iron with arsenic may relate to Fe-As complexes affecting arsenic speciation promoting the oxidation of arsenic to the As(V) species (Bednar *et al.*, 2005). It has been shown that leaching of arsenic from soils by the action of microbiota is increased in the presence of a carbon source (Turpeinen *et al.*, 1999) which 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 marginally higher pH and higher copper concentration than the other most contaminated

sites although the multifactorial environmental characteristics which were assessed have not discerned the cause of this anomalous site.

We identified 18 abundant OTUs 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). 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 *Serratia* 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 research.

There were 9 contaminant-sensitive OTUs identified, including *Bacillus*, *Clostridia*, *Rhizobiales,* and the *Verminephrobacter* symbiont. All of these were strongly 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. the *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 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.

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. Therefore, 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. Moreover, the demonstrated impact of anthropogenic contaminants on the microbial community of a representative member of an ecologically-important taxon raises concerns for both host health and causal effects on the global environment.

Supplementary information is available at the Environmental Microbiology website.

Experimental Procedures

Site description and soil chemistry

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) (Supplementary Figure 1). 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 previously performed (described in Kille *et al.* 2013) where triplicate samples were taken from the epigeic level (surface 10cm), dried at 80^{o}C and analysed via aqua regia digestion for total concentrations of various metals (Supplementary Figure 1). pH varies within small boundaries and is independent of the arsenic gradient. Five sites were identified within the mine in addition to two 'clean' reference sites. The first was located at a site adjacent to the contaminated area, which displays relatively increased arsenic level (On-Site Control) and another 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 (described below). Individuals were immediately washed with distilled water, frozen in liquid nitrogen, ground using a pestle and mortar and stored at -80^{o}C until required. Soil samples were collected from the epigeic surface layer (10 cm; *L. rubellus* habitat) in a one metre square 'W' formation and hand mixed in a sterile bag before being divided into three replicates, chilled and DNA extracted within 24 hours.

DNA extraction

Total DNA was extracted from 5 randomly selected earthworm samples and the three soil replicates from each site. Earthworm extraction was performed to manufacturer specifications using the Qiagen blood and tissue extraction kit (Qiagen Inc., Crawley, UK) with the substitution of proteinase K digestion for a bead-beating step. ~0.5 g 0.1 mm glass beads and ~20 1.0 mm zirconia/silica beads (Biospec products Inc (Bartlesville, Oklahoma, USA)) were placed into 2 ml screw-cap tubes and homogenised using an MPBio FastPrep-24 tissue and cell homogeniser (Solon, Ohio,

USA). The resultant supernatant was utilised in the downstream extraction with the Blood and Tissue kit. DNA was quantified using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE) prior to PCR. Soil extraction was performed to specification using the Soil PowerBio kit (MO BIO Laboratories, CA, USA).

All samples were analysed using Denaturing Gradient Gel Electrophoresis (DGGE) to as an initial assessment of bacterial diversity and community structure following the method described in Webster *et al.* (2006) (Data not shown).

Bar-code Amplification

PCRs were performed in 50µl reactions in an aseptic UV cabinet with sterile plasticware and nuclease-free molecular-grade H_2O as follows: 1x reaction buffer, 1.5 mM $MgCl_2$, 0.4 pmol μL^{-1} each primer, 0.25 mM each dNTP, 1.25 U Taq polymerase plus 1 µl concentration-normalised template. PCR mixture for soil samples contained an additional 10 mg bovine serum albumin (BSA; Promega Corporation, Madison, WI).

Earthworm species confirmation was achieved via sequencing of the COI barcode gene (Primers: LCO-1490 (5'-GGTCAACAAATCATAAAGATATTGG-3'), HCO-2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3'). 16S rRNA community sequencing used universal bacterial primers (357f (5'-CCTACGGGAGGCAGCAG-3') and 907r (5'-CCGTCAATTCMTTTGAGTTT-3')) with 12bp barcode and 454 sequencing adaptors (Roche, CT, USA).

PCR conditions were: initial denaturation of 95^{O}C for 5 mins, 35 amplification cycles of 95^{o}C for 30 seconds, 54^{o}C for 40 seconds, 72^{o}C for 1 Min, and a final single extension cycle of 72^{o}C for 1 minute. In all cases triplicate PCRs were performed and pooled in an equimolar mix prior to sequencing.

Next Generation Sequencing and Bioinformatic Analysis

A total of ~1,200,000 sequence reads were obtained from Research and Testing laboratories (Lubbrock, USA). This dataset was primarily composed of 530,320 454 GS FLX+ reads and expanded with an additional 681,891 454 FLX Titanium reads. Reads were screened at >25 average quality, within 3 standard deviations from mean length and truncated to 650 bp prior to denoising using acacia (Bragg *et al.*, 2012),

incorporating the Quince model (Quince *et al.*, 2009). 726,884 corrected reads were filtered further utilising the QIIME pipeline (Caporaso *et al.*, 2010) to restrict length (350<X<600 bp); remove homopolymers >6; and reject mismatched primers. 579,526 reads were filtered to remove contaminating *L. rubellus* host sequence (22,454) and *Monocystis agilis* (6,893); a known eukaryotic parasite. The remaining 550,179 reads were demultiplexed by sample and randomly subsampled to the lowest sample size whilst still retaining at least three replicates (2,811) which resulted in removal of three *L. rubellus* individuals from analysis. ~148,983 reads were utilised for processing and analysis using the QIIME pipeline (Caporaso *et al.*, 2010) (For detailed processing see Supplementary Figure 2). OTUs were generated at 0.97, 0.94 and 0.88 where appropriate using UCLUST (Edgar, 2010). Taxonomy identification was performed using BLAST with the greengenes reference dataset (McDonald *et al.*, 2012).

Statistical analysis was performed using R (R Core Team & R Development Core Team, 2013) including the Vegan (Oksanen *et al.*, 2013) and ggplot2 (Wickham, 2009) packages. To visually examine the relationship between the earthworm associated microbiomes across the different sites Non Metric Multidimensional Scaling (NMDS) from unifrac distances (Lozupone & Knight, 2005) was performed. To describe and compare community structure Shannon diversity, chao1 richness and observed species metrics' were calculated with QIIME.

To represent association of major OTUs to site conditions, network analysis was performed with QIIME and analysed with Cytoscape (Shannon *et al.*, 2003). OTUs (>200 abundance per sample (7%)) were labelled to most accurate taxonomic level available and coloured by association to site origin conditions (ANOVA P<0.05 = association, P>0.05 = shared (FDR correction)). Site 5 samples were omitted from OTU association calculations due to individuals from this site having distinct geochemical properties (discussed below).

All work was done on the Bio-Linux operating system (Field *et al.*, 2006) and performed on a local compute cluster.

Acknowledgements

We thank Tim Booth for continued support with Bio-Linux. This work was supported by the UK Natural Environmental Research Council Environmental Bioinformatics Centre (NEBC). Processing was performed on a compute cluster funded by The Royal Society.

Bibliography

Barr CJ, Bunce RGH, Smart S, Whittaker HA. (1978). Countryside Survey 1978 vegetation plot data, NERC Environmental Information Data Centre.

Bednar AJ, Garbarino JR, Ranville JF, Wildeman TR. (2005). Effects of iron on arsenic speciation and redox chemistry in acid mine water. *J Geochemical Explor* **85**:55–62.

Bernard L, Chapuis-Lardy L, Razafimbelo T, Razafindrakoto M, Pablo A-L, Legname E, *et al.* (2012). Endogeic earthworms shape bacterial functional communities and affect organic matter mineralization in a tropical soil. *ISME J* **6**:213–22.

Bragg L, Stone G, Imelfort M, Hugenholtz P, Tyson GW. (2012). Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat Methods* **9**:425–426.

Brown GG, 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. *Eur J Soil Biol* **36**:177−198.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**:335–6.

Daniel O, Anderson JM. (1992). Microbial biomass and activity in contrasting soil materials after passage through the gut of the earthworm *Lumbricus rubellus* Hoffmeister. **24**:465–470.

Davidson SK, Powell R, James S. (2013). A global survey of the bacteria within earthworm nephridia. *Mol Phylogenet Evol* **67**:188–200.

Davidson SK, Stahl DA. (2008). Selective recruitment of bacteria during embryogenesis of an earthworm. *ISME J* **2**:510–8.

Depkat-Jakob PS, Brown GG, Tsai SM, Horn MA, Drake HL. (2013). Emission of nitrous oxide and dinitrogen by diverse earthworm families from Brazil and resolution of associated denitrifying and nitrate-dissimilating taxa. *FEMS Microbiol Ecol* **83**:375–91.

Depkat-Jakob PS, Hunger S, Schulz K, Brown GG, Tsai SM, Drake HL. (2012). Emission of methane by *Eudrilus eugeniae* and other earthworms from Brazil. *Appl Environ Microbiol* **78**:3014–9.

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*, Vol. 6, Springer-Verlag, pp. 65–87.

Drake HL, Horn MA. (2007). As the worm turns: the earthworm gut as a transient habitat for soil microbial biomes. *Annu Rev Microbiol* **61**:169–89.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**:2460–1.

Edwards CA. (2004). Earthworm Ecology. Taylor & Francis.

Egert M, Marhan S, Wagner B, Scheu S, Friedrich MW. (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 Microbiol Ecol* **48**:187–97.

Farmer JJ, Davis BR, Hickman-Brenner FW, McWhorter A, Huntley-Carter GP, Asbury MA, *et al.* (1985). Biochemical identification of new species and biogroups of Enterobacteriaceae isolated from clinical specimens. *J Clin Microbiol* **21**:46–76.

Field D, Tiwari B, Booth T, Houten S, Swan D, Bertrand N, *et al.* (2006). Open software for biologists: from famine to feast. *Nat Biotechnol* **24**:801–3.

Goffredi SK, Johnson SB, Vrijenhoek RC. (2007). Genetic diversity and potential function of microbial symbionts associated with newly discovered species of Osedax polychaete worms. *Appl Environ Microbiol* **73**: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. *PLoS One* **6**:e24786.

Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley AS. (2011). The bacterial biogeography of British soils. *Environ Microbiol* **13**:1642–54.

Hackstein JH, Stumm CK. (1994). Methane production in terrestrial arthropods. *Proc Natl Acad Sci U S A* **91**:5441–5.

Ihssen J, Horn M, Matthies C. (2003). N2O-producing microorganisms in the gut of the earthworm *Aporrectodea caliginosa* are indicative of ingested soil bacteria. *Appl …* **69**:1655–1661.

Kille P, Andre J, Anderson C, Ang HN, Bruford MW, Bundy JG, *et al.* (2013). DNA sequence variation and methylation in an arsenic tolerant earthworm population. *Soil Biol Biochem* **57**:524–532.

Klinck BA, Palumbo B, Cave M, Wragg J. (2005). Arsenic dispersal and bioaccessibility in mine contaminated soils: a case study from an abandoned arsenic min in Devon, UK.

Langdon CJ, Morgan AJ, Charnock JM, Semple KT, Lowe CN. (2009). As-resistance in laboratory-reared F1, F2 and F3 generation offspring of the earthworm *Lumbricus rubellus* inhabiting an As-contaminated mine soil. *Environ Pollut* **157**:3114–9.

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

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. *Environ Pollut* **158**:3186–93.

Lavelle P, Decaëns T, Aubert M, Barot S, Blouin M, Bureau F, *et al.* (2006). Soil invertebrates and ecosystem services. *Eur J Soil Biol* **42**:S3–S15.

Ley RER, Lozupone CA, Hamady M, Knight R, Gordon JI. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* **6**:776–788.

Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H, *et al.* (2008). Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci U S A* **105**:2117–22.

Li X, Fisk MC, Fahey TJ, Bohlen PJ. (2002). Influence of earthworm invasion on soil microbial biomass and activity in a northern hardwood forest. *Soil Biol Biochem* **34**:1929–1937.

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

Lubbers IM, van Groenigen KJ, Fonte SJ, Six J, Brussaard L, van Groenigen JW. (2013). Greenhouse-gas emissions from soils increased by earthworms. *Nat Clim Chang* **3**:187–194.

576 Lund MB, Davidson SK, Holmstrup M, James S, Kjeldsen KU, Stahl DA, *et al.* (2009).
577 Diversity and host specificity of the *Verminephrobacter*-earthworm symbiosis. *Environ*
578 *Microbiol* **12**:2142–2151.

579 Lund MB, Holmstrup M, Lomstein BA, Damgaard C, Schramm A. (2010). Beneficial
580 effect of *Verminephrobacter* nephridial symbionts on the fitness of the earthworm
581 *Aporrectodea tuberculata*. *Appl Environ Microbiol* **76**:4738–43.

582 McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, *et al.* (2012).
583 An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary
584 analyses of bacteria and archaea. *ISME J* **6**:610–8.

585 Morgan AJ, Kille P, Stürzenbaum SR. (2007). Microevolution and Ecotoxicology of
586 Metals in Invertebrates. *Environ Sci Technol* **41**:1085–1096.

587 Nahmani J, Hodson ME, Black S. (2007). A review of studies performed to assess metal
588 uptake by earthworms. *Environ Pollut* **145**:402–24.

589 Nechitaylo TY, Yakimov MM, Godinho M, Timmis KN, Belogolova E, Byzov BA, *et al.*
590 (2010). Effect of the earthworms *Lumbricus terrestris* and *Aporrectodea caliginosa* on
591 bacterial diversity in soil. *Microb Ecol* **59**:574–87.

592 Normand P, Lapierre P, Tisa LS, Gogarten JP, Alloisio N, Bagnarol E, *et al.* (2007).
593 Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range
594 and host plant biogeography. *Genome Res* **17**:7–15.

595 Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, *et al.* (2013).
596 vegan: Community Ecology Package.

597 Pandazis G. (1931). Zur Frage der Bakteriensymbiose bei Oligochaeten. *Zentralbl*
598 *Bakteriol* **120**:440–453.

599 Pinel N, Davidson SK, Stahl DA. (2008). *Verminephrobacter eiseniae* gen. nov., sp.
600 nov., a nephridial symbiont of the earthworm Eisenia foetida (Savigny). *Int J Syst Evol*
601 *Microbiol* **58**:2147–57.

602 Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, Head IM, *et al.* (2009). Accurate
603 determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* **6**:639–
604 41.

605 R Core Team, R Development Core Team. (2013). R: A Language and Environment for
606 Statistical Computing.

607 Sabri A, Leroy P, Haubruge E, Hance T, Frère I, Destain J, *et al.* (2011). Isolation, pure
608 culture and characterization of *Serratia symbiotica* sp. nov., the R-type of secondary
609 endosymbiont of the black bean aphid Aphis fabae. *Int J Syst Evol Microbiol* **61**:2081–8.

610 Schramm A, Davidson SK, Dodsworth J a., Drake HL, Stahl DA, Dubilier N. (2003).
611 *Acidovorax*-like symbionts in the nephridia of earthworms. *Environ Microbiol* **5**:804–809.

612 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al.* (2003).
613 Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction
614 Networks. 2498–2504.

615 Sizmur T, Watts MJ, Brown GD, Palumbo-Roe B, Hodson ME. (2011). Impact of gut
616 passage and mucus secretion by the earthworm *Lumbricus terrestris* on mobility and
617 speciation of arsenic in contaminated soil. *J Hazard Mater* **197**:169–75.

618 Teixeira L, Ferreira A, Ashburner M. (2008). The bacterial symbiont *Wolbachia* induces
619 resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biol* **6**:e2.

620 Thakuria D, Schmidt O, Finan D, Egan D, Doohan FM. (2010). Gut wall bacteria of
621 earthworms: a natural selection process. *ISME J* **4**:357–66.

622 Turpeinen R, Pantsar-Kallio M, Häggblom M, Kairesalo T. (1999). Influence of microbes
623 on the mobilization, toxicity and biomethylation of arsenic in soil. *Sci Total Environ*
624 **236**:173–80.

625 Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, *et al.* (2007).
626 Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum.
627 *Microbiol Mol Biol Rev* **71**:495–548.

628 Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, *et al.*
629 (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding
630 higher termite. *Nature* **450**:560–565.

631 Webster G, Parkes RJ, Cragg B a, Newberry CJ, Weightman AJ, Fry JC. (2006).
632 Prokaryotic community composition and biogeochemical processes in deep subseafloor
633 sediments from the Peru Margin. *FEMS Microbiol Ecol* **58**:65–85.

634 Weiss BL, Maltz M, Aksoy S. (2012). Obligate symbionts activate immune system
635 development in the Tsetse Fly. *J Immunol*.

636 Wickham H. (2009). ggplot2: elegant graphics for data analysis. Springer New York.

637 Wüst PK, Horn MA, Drake HL. (2011). Clostridiaceae and Enterobacteriaceae as active
638 fermenters in earthworm gut content. *ISME J* **5**:92–106.

639 Wust PK, Horn MA, Drake HL, Wüst PK. (2009). In Situ Hydrogen and Nitrous Oxide as
640 Indicators of Concomitant Fermentation and Denitrification in the Alimentary Canal of
641 the Earthworm *Lumbricus terrestris*. *Appl Environ Microbiol* **75**:1852–9.

642

643

Titles and legends to figures

Figure 1. Contrasting the *Lumbricus rubellus* and soil microbiomes.

Figure demonstrating separation between soil (squares) and *L. rubellus* (circles) showing change in community structure from soil to host. (A) PCoA of unifrac distances with distinct separation on the primary axis. Each point represents an individual microbiome sample. (B) Bacterial families with significant difference between host and soil. If Family level annotation was not possible Order was given denoted by (o). Additive presence for all sites ordered by magnitude and plotted with standard deviation error bars. Families with >3.5% host or soil reads and significant change displayed. Right box describes family, or next identifiable taxa. t-test denotes significance of change in family abundance between soil and host (* $p<0.05$, ** $p<0.01$).

Figure 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. 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. Proteobacteria has been displayed at class level as the largest Phyla. Full taxonomic analysis is in main text body.

Figure 3. The effect of anthropogenic stress on community structure.

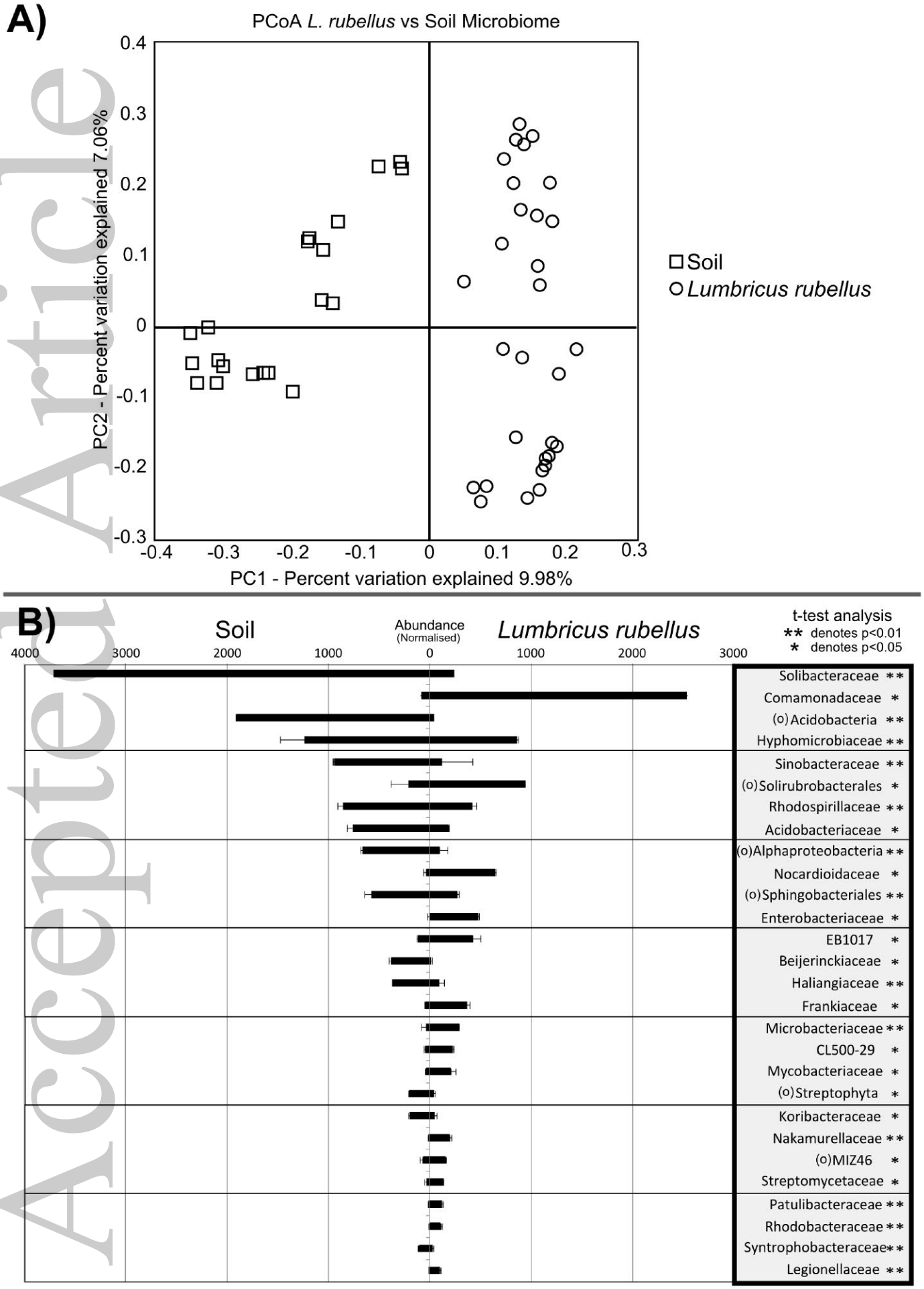
(A) Overview of Diversity and Richness (Shannon and Chao1 respectively) for all soil (Squares) and *Lumbricus rubellus* (circles) microbiomes as coloured by site origin. Lower right box displays magnified area for clarity. Also see Supplementary Figure 3.
(B) Non-parametric Multi-Dimensional Scaling (NMDS) plot representing divergence of *L. rubellus* microbiota profile and site similarity in conjunction with environmental factors. 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 due to iron affecting. Site-specific grouping is observed, as is the effect of increasing stress on the microbiome community structure.

Figure 4. Network Analysis of all *L. rubellus* samples with associated abundant OTUs

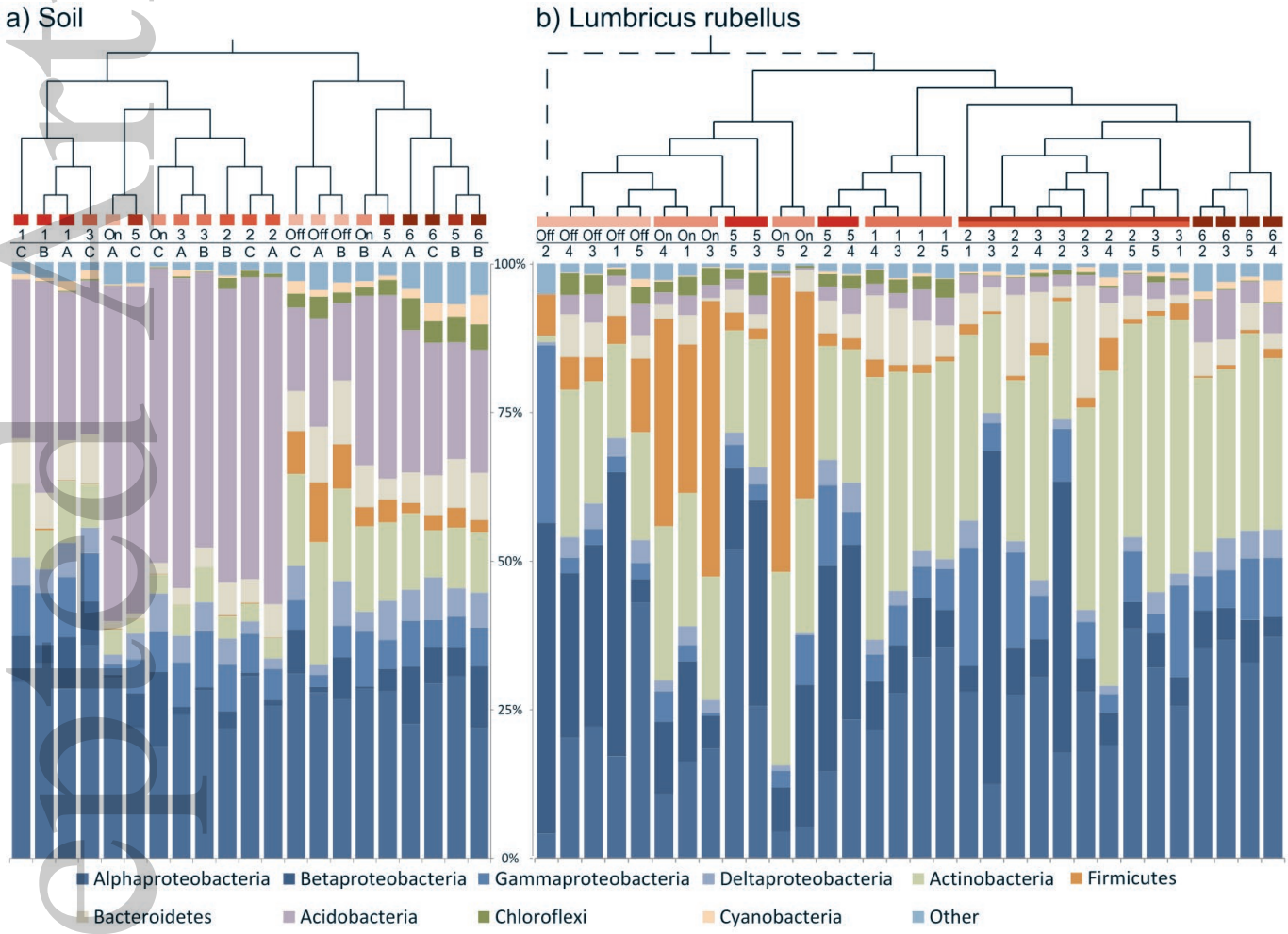
Significantly present OTUs (>7% abundance, diamonds) in network association with earthworm individuals (*L. rubellus*, blue circles). Coloured by association to site origin conditions when ANOVA testing associates OTU with condition (P<0.05 = association, FDR correction). All samples were incorporated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

Figure 5. Venn diagram summarising shared OTUs between soil and earthworm samples at High and Low contaminant sites.

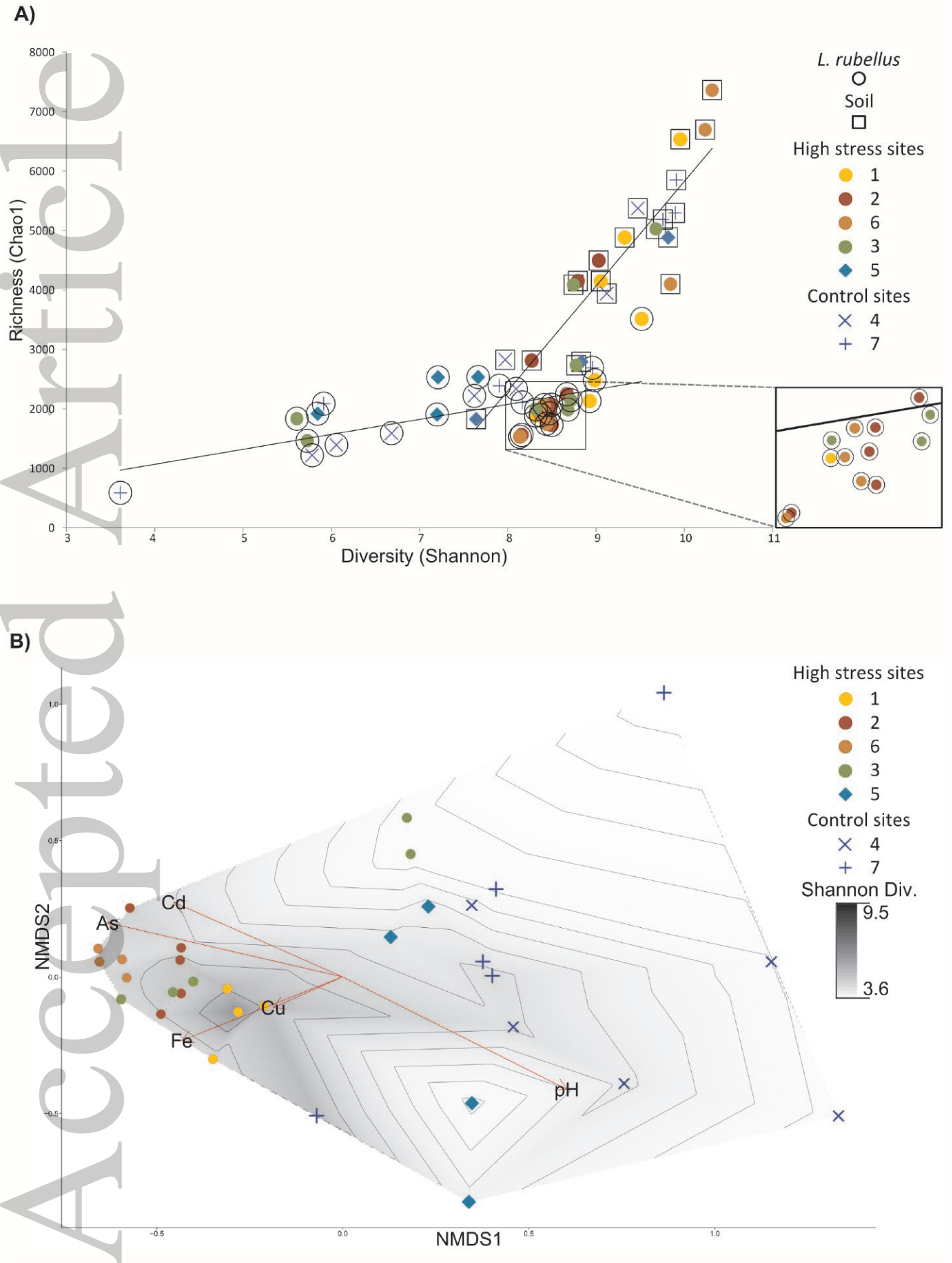
A high number of OTUs were observed in all situations correlating with the soil-derived microbiome hypothesis however, a smaller number of *L. rubellus*- OTUs were observed, implying presence of host-associated species. OTUs counted when derived from a non-singleton sequence.



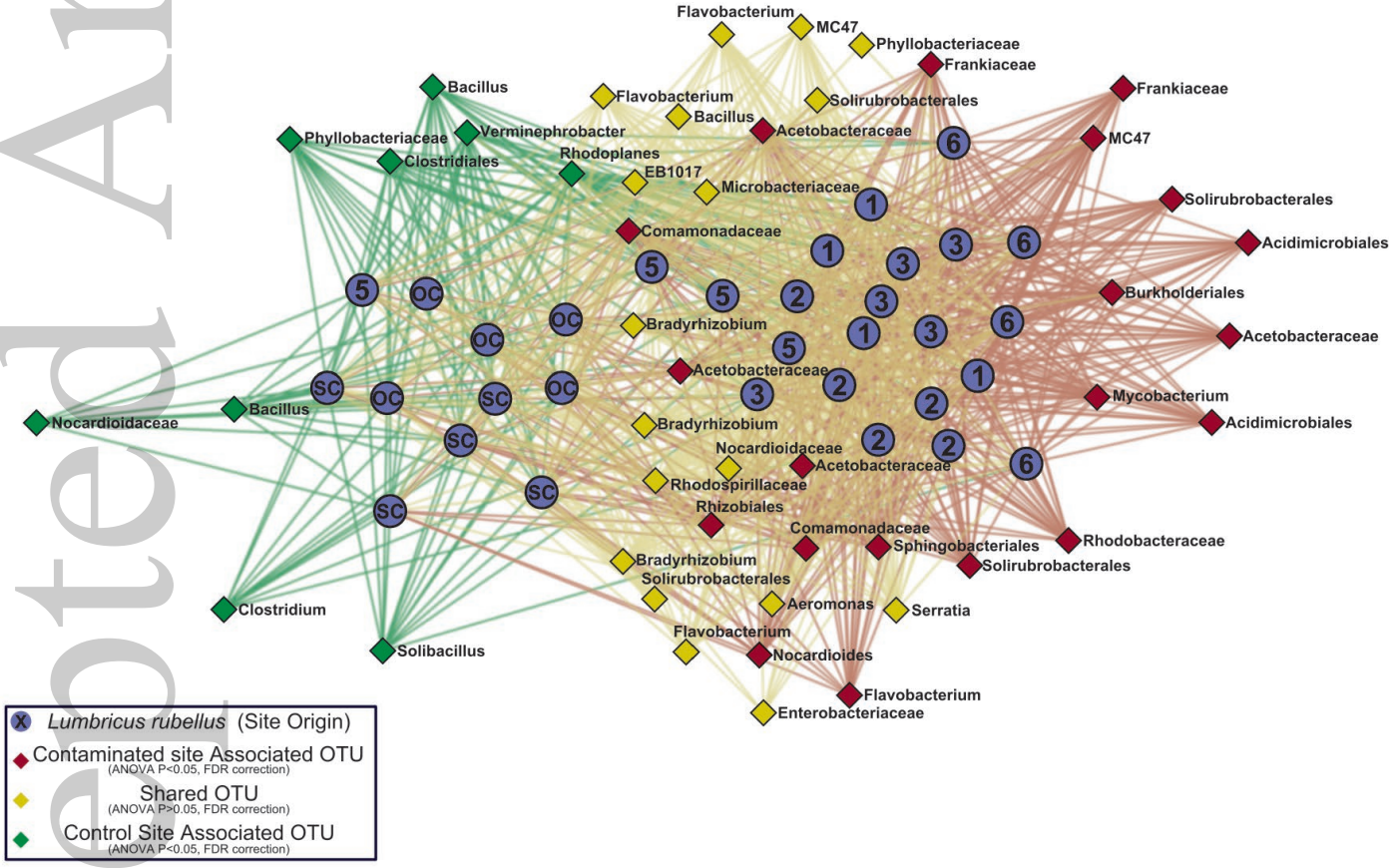
EMI_12712_F1.tiff



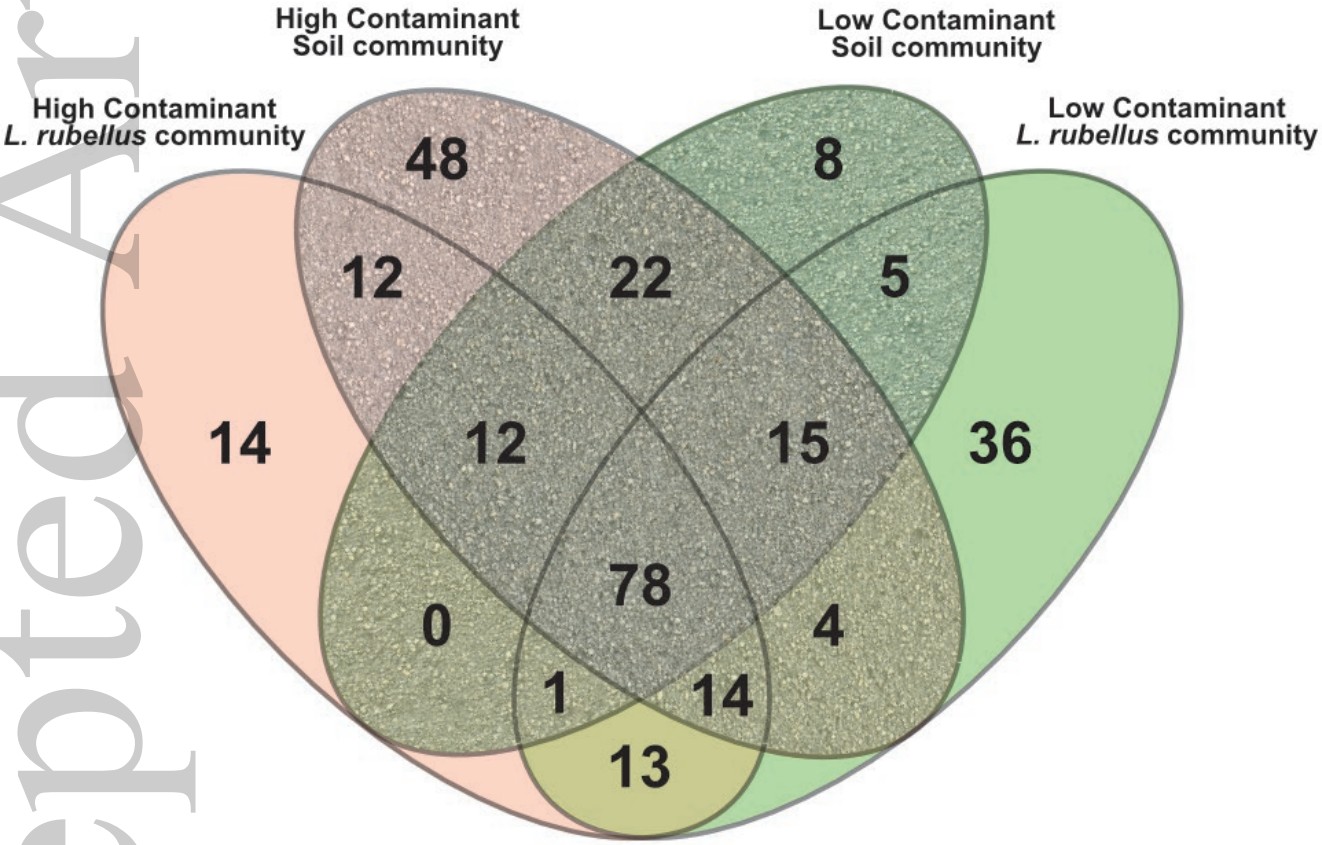
EMI_12712_F2.tiff



EMI_12712_F3.tiff



EMI_12712_F4.tiff



EMI_12712_F5.tiff