Distribution of Antibiotic Resistance Genes in soils and crops. A field study in legume plants (*Vicia faba* L.) grown under different watering regimes.

Authors: Francisco Cerqueira¹, Víctor Matamoros¹, Josep Bayona¹, Goffe Elsinga², Luc M. Hornstra², Benjamin Piña¹

Filiations:

¹Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research, 08034, Barcelona, Spain

²KWR Waterycycle Research Institute, Groningenhaven 7, 3433 PE, Nieuwegein, the Netherlands

Abstract

Social concern has raised during the last years due to the development of antibiotic resistance hotspots in different environmental compartments, including the edible parts of crops. To assess the influence of the water quality used for watering, we collected samples from soil, roots, leaves and beans from the legume plant *Vicia faba* (broad beans) in three agricultural peri-urban plots (Barcelona, NE Spain), irrigated with either groundwater, river water, or reclaimed water. Antibiotic resistance genes (ARGs) *sulI*, *tetM*, *qnrS1*, *bla*_{CTX-M-32}, *bla*_{OXA-58}, *mecA*, and *bla*_{TEM} were quantified by real-time PCR, along with 16S rDNA and *intI1* sequences, as proxies for bacterial abundance and integron prevalence, respectively. Microbiome composition of all samples were analyzed by high-throughput DNA sequencing. Results show a gradient of bacterial species diversity and of ARG prevalence from highly diverse soil samples to microbially-poor beans and leaves, in which Rhizobiales essentially displaced all other groups, and that presented very small loads of ARGs and integron sequences. The data suggest that the microbiome and the associated resistome were likely influenced by agricultural practices and water quality, and that future irrigation water legal standards should consider the specific Physiology of the different crop plants.

Keywords Irrigation Water; Soil; Antibiotic Resistance Genes; Microbiomes; Rhizosphere; Endophytes.

Funding Sources

This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675530.

Highlights:

- ARGs presented different abundance profiles according to the sampling fields
- Rhizobiales dominate the microbiome of all legume plant parts
- *intI1* and ARGs abundances inversely correlate with Rhizobiales' predominance
- Rhizobiales roots' colonization influences the microbiome of surrounding soil

1. Introduction

Antibiotic resistance hotspots occur in essentially all environmental compartments subjected to anthropogenic pressure, such as wastewater treatment plants (WWTPs), pharmaceutical manufacturing plant effluents, and aquaculture facilities (Michael et al., 2013). While the reuse of polluted water from these sites represents a number of benefits, and that major advances have been made with respect to producing safe treated effluents, there are a number of issues that have to be solved before using reclaimed water in a general basis. Currently operating WWTPs fail to completely remove antibiotics, antibiotic resistant bacteria (ARB), and antibiotic resistance genes (ARGs) (Le-Minh et al., 2010; Manaia et al., 2018; Michael et al., 2013). Therefore, the use of reclaimed water in agriculture may favor the dissemination of antibiotics, ARB, and ARGs in the environment (Ben et al., 2017).

The proliferation of ARB and ARGs is frequently associated with mobile genetic elements (MGEs), being a relevant example gene cassettes present in integrons (Domingues et al., 2012). Class 1 integrons are widely reported in pathogens from clinical settings, carrying gene

cassettes associated with antibiotic resistance, and can be found in plasmids or associated with transposons, allowing their dissemination through horizontal gene transfer (HGT) (Gillings et al., 2008; Labuschagne et al., 2008; Li et al., 2006; Lind et al., 2006; Mendes et al., 2007). Integrons have been recently detected in microbial species from many different environments, including wastewater (Tennstedt et al., 2003), river water (Mukherjee and Chakraborty, 2006), and soil (Agerso and Sandvang, 2005). This suggests that the dissemination of integrons may be a general problem of environmental pollution and public health (Stokes et al., 2006).

HGT frequency is determined by two complementary main barriers: Phylogeny, and Ecology. Considering antibiotic exposure as a source of selection, ecology provides a physical barrier, as donor and recipient bacteria have to be in the same niche for HGT to occur. Once this physical barrier has been broken, phylogeny determines the HGT network of ARGs and other transposable elements, as intra-taxon genetic exchange occurs more often than inter-taxon (Hu et al., 2016). Microbial communities in natural environments harbor ARB and ARGs, even in unpolluted and/or isolated areas (Allen et al., 2010; Bhullar et al., 2012). Selective pressure on these environments, as it may occur by the incorporation of antibiotics from different farming practices, may increase the abundance of ARB in soil microbiomes (Chen et al., 2016; Han et al., 2016; Hartmann et al., 2015; Heuer et al., 2011; Hu et al., 2016). This effect may in turn lead to the subsequent increase of antibiotic resistance levels on soils and in harvested vegetables (Han et al., 2016; Marti et al., 2013; Westhoff et al.), which may ultimately result in an impact in human health (Becerra-Castro et al., 2015). The estimate number of deaths caused annually by ARB in the United States of America and in the European Union, calculated as 23,000 and 25,000, respectively (<https://www.cdc.gov/drugresistance/>, ECDC/EMEA, 2009), but the relevance of the presence of ARB and ARGs in food in these figures is currently unknown.

In this work we surveyed seven ARGs of environmental and clinical relevance (*sulI*, *bla_{TEM}*, *bla_{CTX-M-32}*, *mecA*, *qnrS1*, *tetM*, and *bla_{OXA-58}*) together with the integrase *intI1*, diagnostic for potentially HGT activity and marker of anthropogenic pollution (Berendonk et al., 2015; Bertini et al., 2007; Gillings et al., 2015; Hembach et al., 2017; Tamang et al., 2008; Wielders et al., 2002). Both integrons and ARGs have been found associated to MGEs in crop fields (Berendonk et al., 2015), and in edible vegetables (Hu et al., 2008; Nonaka et al., 2012; Wang et al., 2015). In a previous study, we found a gradient of microbial species and ARB-

related genetic elements along the soil-plant-fruit continuum in commercial tomato crops grown under different water regimes, and found a correlation between the presence of ARGs in fruits and the soil and plant microbiome composition, which are, in turn, under the influence of agricultural practices (Cerqueira et al., 2019). Here we present a comparable study using a very popular legume crop, the broad bean *Vicia faba*, which, contrary to tomatoes, is consumed generally (but not always) cooked, once removed from the pod. The study was carried out in three farm fields, two of them indirectly irrigated with treated wastewater (TWW), and a rural organic farming plot irrigated with groundwater, all located at less than 30 km from Barcelona (Spain). Our main goal is generating a predictive model allowing the classification of the sampling zones as control or polluted areas, incorporating to it as many different types of crops and different water qualities as possible, always within the current EU regulation for crop irrigation and soil amendment procedures.

2. Methods

2.1. Site description

Broad beans crops were collected in agricultural fields in the delta of the Llobregat River (Catalonia, Spain, Fig. 1). The reference zone (Z1) was located in a field near Cabrera de Mar, northeast of the Barcelona metropolitan area, away from most human impact, irrigated with groundwater, and fertilized with local pigeon manure. In contrast, zone Z2, in the Llobregat low valley, receives water from the Infanta Channel, which is mainly fed by TWW from 10 different WWTP effluents that discharge into the Rubí Creek, representing about 92% of the total water flow. Z3 was irrigated with river water also impacted by WWTPs (see map as Supplementary Fig. 1), but in this case TWW represented less than 18% of the water flow (WTTé et al., 2012). Z2 and Z3 applied chemical fertilizers when needed. Broad beans plants were collected in April of 2017, when they reached their commercial size. The samples included bulk soil, roots, leaves, and beans, removed from the pod. Plants were collected in a systematic random sampling scheme, in which each collection site was divided into five equal sections. One plant (roots, leaves, and broad beans) and a soil sample were collected per each sampling site (Z1, Z2, Z3) and section (five sections per site), 60 samples in total. Samples were collected

into sterile plastic containers, and, within few hours, transported on ice to the laboratory, where they were processed and total DNA was extracted.

2.2. DNA extraction

Preliminary procedures for DNA extraction were performed as described (Bodenhausen et al., 2013; Qvit-Raz et al., 2008) with slight modifications. Broad beans were washed in sterilized distilled water. Roots and leaves samples in addition had their surfaces sterilized by washing in 70% ethanol for 1 min, rinsing with sterilized distilled water for 1 min, washed again in ethanol 70%, and finally rinsed with sterilized distilled water twice. Samples from the water used to the last rinse were plated in LB agar media for three days in an incubator at 28 °C to evidence lack of contaminating bacteria. Leaves, roots, and broad beans seeds were macerated with mortar and pestle, adding 20 mL of PBS in the case of leaves, and transferred to 50 ml tubes, through a 100 µm nylon Cell strainer (Corning® Cell Strainer). The samples were centrifuged in an Eppendorff Centrifuge 5804 R (Eppendorf, AG) at 4500 rpm for 15 min. The supernatants were discarded, and the pellets (about 250 mg) processed for DNA extraction using the Qiagen DNeasy PowerSoil kit (Qiagen, Inc., Valencia, CA), to a final elution volume of 100µL. The concentration and the quality of the DNA were tested using a NanoDrop Spectrophotometer 8000 (ThermoFisher Scientific, Inc, Waltham, MA). Extracted DNA samples were stored at -20 °C until further use.

2.3. Detection and quantification of ARGs copy numbers

Primers for *int11*, *sul1*, *tetM*, *qnrS1*, *mecA*, *bla_{TEM}*, *bla_{CTX-M-32}*, *bla_{OXA-58}*, and bacterial 16S rDNA sequences are listed in Supplementary Table ST1 (Bradford, 2001; Chen et al., 2014; Doenhoefer et al., 2012; Frank et al., 2007; Mugnaioli et al., 2006; Tamang et al., 2008). Dynamo ColorFlash SYBR Green (ThermoFisher Scientific, Inc.) chemistry was used for *mecA*, and *bla_{OXA-58}* qPCR quantification; all the other ARGs were quantified with SYBR green (A F. Hoffmann–La Roche AG, Inc, Switzerland) in a Lightcycler 480 (A F.Hoffmann–La Roche AG, Inc). Thermal cycling conditions differed for both technologies: 95°C for 10 min, 45 cycles at 95°C for 15 s and Ta for 1 min for the first one, and 95°C for 7 min, 40 cycles at 95°C for 10 s and Ta for 30 s for the second one. All qPCR reactions were performed in a Lightcycler 480

II (A F. Hoffmann–LaRoche AG, Inc). Melting curves were obtained to confirm amplification specificity. Optimized primer concentrations were 0.15 μ M for *tetM*, *mecA*, and *bla*_{OXA-58} primers, and 0.3 μ M for *sulI*, *intI1*, *bla*_{TEM}, *bla*_{CTX-M-32}, *qnrS1*, and 16S rDNA primers; qPCR reactions were performed by duplicates, using a fixed dilution of raw DNA extract. The quality criteria used was an $R^2 > 0.99$ and an accepted efficiency of reactions from 90% to 110%.

2.4. Standards and normalization methods

Four plasmid vectors were used as standards for absolute quantification of qPCR amplicons. For 16S rDNA gene, *bla*_{TEM}, *bla*_{CTX-M-32}, *sulI*, *qnrS1*, *intI1*, purified DNA samples from pNORM1 plasmid (Gat et al., 2017) were used as quantification standards (Supplementary Table ST1). For *tetM*, *bla*_{OXA-58} and *mecA*, purified DNA samples from three different pUC19 plasmids were used as standards (Laht et al., 2014; Szczepanowski et al., 2009; Tamminen et al., 2011) (Supplementary Table ST1). Standard curves (Ct per log copy number) for quantification of all genes were obtained for each run, using 10-fold serial dilutions of the corresponding standards. Gene copy number of standards was determined via the plasmid/genomic DNA concentration using the NanoDrop Spectrophotometer 8000, applying the DNA copy number and the Dilution Calculator algorithm (<https://www.thermofisher.com/us>). Three technical replicates were run for the standard curves; complete reaction mixes plus nuclease free water, but no template, were used as negative control qPCR reactions. ARG concentration in samples was calculated using the standard curve equation and the measured Ct value. The quality control of raw Ct values for standard curve and unknown samples was performed before the analysis. The limit of quantification (LOQ) was defined as the lowest point on the linear part of the standard curve: 100 gene copies per reaction for ARGs and *intI1* and 1000 gene copies per reaction for 16S rDNA. Copy numbers of all ARGs and *intI1* were normalized to gram of sample (gene copies g⁻¹).

2.5. Data analysis for ARGs and *intI1*

The analysis of the ARGs profile and abundance was performed in the R environment (version 3.4.0; <http://www.r-project.org/>). The ARGs quantified by qPCR were normalized by gram of sample and log transformed. Principal Component Analysis (PCA) was performed with the

function *principal*, from the package *psych* (<https://cran.rproject.org/web/packages/psych/index.html>). One-way and two-way ANOVA tests, Tukey'B post-hoc analysis, Pearson's correlations, and *fdr* (false discovery ratio) correction for multiple tests were also performed with the *psych* package. Due to the amplification of chloroplastidial and mitochondrial DNA, bacterial 16S rDNA copy numbers were calculated by assessing the proportion of reads attributed to chloroplasts and mitochondria after sequencing the corresponding 16s rDNA amplicons using *dada2* (Callahan et al., 2016) and *phyloseq* packages in R (McMurdie and Holmes, 2013).

2.6. 16S rDNA library preparation and sequencing

We prepared of a pooled library of 16S rDNA amplicons from 60 soil and broad beans samples' DNA, using primers 331F-TCCTACGGGAGGCAGCAGT and 518R-ATTACCGCGGCTGCTGG (Supplementary Table ST1). Illumina Nextera XT Index Kit v2 - Index 1 (i7) and Nextera XT Index Kit v2 - Index 2 (i5) adapters were used to tag the samples. A two-step PCR protocol was used to prepare samples. The first step added overhangs to the 16s rDNA amplicons, using the following thermal cycling conditions: 98°C for 1 min, 8 cycles at 95°C for 30 s, 60°C for 1 min, and 72°C for 1 min. The second PCR step added the Illumina Nextera XT adapters under the following thermal cycling conditions: 95°C for 3 min, 8 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s. The composed, tagged sample was mass sequenced at the KWR Watercycle Research Institute (Nieuwegein, Netherlands) (<https://www.kwrwater.nl/>), with high-throughput sequencing, using an Illumina MiSeq. Quality control of the sequences was done using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Illumina Nextera adapters were removed and the sequences were quality trimmed for a minimum Phred quality score ($Q \geq 30$) with Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). R environment was used for all the downstream analysis of the sequences, specifically, the *dada2* package. Sequences shorter than 160 bp were discarded. The sequences were dereplicated and PCR chimeras were checked and filtered out. The taxonomic assignment of Operational Taxonomic Units (OTUs) was done with SILVA database v128 for *dada2* package. Sequences were assigned to OTUs based on a 97% similarity. The sequences classified as chloroplasts and

mitochondria were removed. Raw sequences were deposited at the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under the accession number SRP151778.

2.7. Analysis of Microbiomes

The analysis of microbial communities was performed with R environment (version 3.4.0), with the packages *phyloseq*, *vegan* (<https://cran.rproject.org/package=vegan>), *VennDiagram* (<https://cran.rproject.org/package=VennDiagram>), *caret* (<https://cran.r-project.org/package=caret>), and *mixOmics* (<https://cran.r-project.org/package=mixOmics>). OTUs abundances were log transformed for β -diversity analysis.

Microbiomes variation between samples were visualized by a Non-metric Multidimensional Scaling (NMDS) using Weighted-UniFrac dissimilarity (Lozupone et al., 2007), with the functions *ordinate* of the *Phyloseq* package. The function *adonis* from *Vegan* package to measure effect size and significances on β -diversity. The *betadisper* function was utilized to test multivariate homogeneity of group variances. Taxa with a prevalence lower than 3 % were excluded from the β -diversity analyses, in order to study the “core” microbiomes.

Differences in OTU distribution between reference and TWW-impacted sampling zones were explored using the supervised machine learning classification technique Random Forest (Breiman, 2001). The data was partitioned in train and test set (70% and 30%, respectively) randomly splits it according the grouping classes. Absolute abundance values were centered and scaled. Accuracy was the metric chosen for model evaluation. For re-sampling a 5-fold cross validation repeated 3 times was chosen. The model was fitted over different tuning parameters, using the function *train*. Variable importance was measured by the Median Accuracy Decrease. The model created with the training data was utilized to predict the sampling fields on the test data according the OTUs abundances. A confusion matrix was created to evaluate the predictions.

3. Results

3.1. ARG abundance and distribution

Bacterial 16S rDNA and *bla*_{TEM} DNA sequences were detected in all samples analyzed. Whereas *int11*, a marker for anthropogenic pollution, was found in 43% of the samples (Table 1). Bacterial 16S sequences, considered as a proxy for bacterial abundance, were found at counts from 140 to more than 5×10^7 copies g⁻¹ of sample. The integron *int11* sequence was found at approximately 70 times lower levels even at its maximum amounts, at around 7.5×10^5 copies g⁻¹ of sample (Table 1). Among ARGs, abundance levels varied in several orders of magnitude, being *bla*_{TEM} the most abundant one by far, with counts ranging from 30 to 4×10^5 copies g⁻¹ of sample (Table 1). All other studied ARGs were found at much lower levels or below the limit of detection (Table 1); *mecA* levels were below the limit of quantification in all samples (not shown). Despite having a prevalence of only 43%, the maximum counts of *int11* surpasses the levels of the rest of ARGs in samples where it was present, except for *bla*_{TEM}. Thus, the data suggest that at least *bla*_{TEM} is widespread in the bacterial population, and not necessarily linked to MGEs or, at least, not to *int11* (Table 1).

The distribution of genetic element abundances in the different sample types and zones was further investigated by two-way ANOVA ("Zone" times "Sample type", Table 1). The analysis shows a very strong influence of sample type (soil vs. root endophytes vs. leaves vs. beans) for all genetic elements, and for bacterial 16S abundance, whereas the factor "Zone" was only relevant for the most prevalent ARGs (*bla*_{TEM}, *sul1*, *qnrS1*) and *int11* (Table 1). The interaction between the two factors was non-significant or very weak in all cases (Table 1). These different distribution patterns are shown in Fig. 1, which includes results from a one-way ANOVA/Tukey's B post hoc test across the different zones. Note the decrease in bacterial abundances from soils (10^6 - 10^8 copies/g of sample) to roots and leaves (10^4 - 10^6 copies/g of sample) and to beans (10^4 - 10^5 copies/g of sample), as well as the essentially lack of *int11* and ARG sequences in beans and leaves (except for some Z1 samples). In fact, *bla*_{TEM} was the only ARG present in beans and leaves at least moderate levels (some 100 copies g⁻¹ of sample). This genetic element showed, in addition, a defined distribution by zones, being more prevalent in samples from Z3 than in those from Z1 or Z2. This differential distribution was observed for all sample types, from soils to beans, suggesting that it the soil determined the final loads in *bla*_{TEM} in the plant (Fig. 1). A similar pattern was observed for *qnrS1*, although the levels in bean samples in this case were too low for any meaningful statistical analysis (Fig. 1). Finally, *int11* was more abundant in soil and root samples from Z1 than from Z2 or Z3, and too scarce

in leaves and beans for statistical analyses (Fig. 1). ARGs *tetM* and *bla*_{OXA-58} were only found in soil samples and at low abundances in all cases (100 copies g⁻¹ of sample or less, Fig. 1)

3.2. Microbiome Analysis

A total of 839 OTUs were annotated from a total of 3,114,367 readings, in 60 samples with a median sequencing depth of 57,617 (range 1,883 to 89,098) reads per sample after quality control. The presence of chloroplastidial and mitochondrial sequences was only marginal in most soil samples, but became predominant in roots, leaves and particularly beans samples, in which the presence of actual bacterial sequences was only marginal, and even absent in some leaves' samples and in most beans' samples (Supplementary Fig. 2). In fact, none of the Zone 3 beans' samples showed significant numbers of bacterial reads (not shown). The distribution of OTUs among the different samples, taxonomically grouped at the Order level, is shown in the graph in Fig. 2. The graph shows a rich bacterial diversity in soil samples, whereas plant samples showed a clear prevalence of *Rizobiales*' sequences, which started at root endophyte microbiomes to become clearly predominant in leaves' and beans' microbiomes (red sectors in Fig. 2, note that there was no data for Z3 beans). Therefore, we concluded that *Rhizobiales* essentially displaced most of the other bacterial groups in the endophytic microbiome, occasionally mixed with other groups (*Micrococcales* in leaves from Z2, *Enterobacteriales* in leaves from Z3, and several minor groups in beans from Zone 1, Fig. 2).

Variations among the different soil and endophytic microbiomes were analyzed by β -diversity and Random Forest classification analyses. The NDMS plot in Fig. 3A (stress=0.06), based on Weighted UniFrac dissimilarity index, revealed a much higher dissimilarity for the OTUs from soil and roots (brown and orange symbols in Fig. 3A) than for those from leaves and beans (light and dark green symbols). The contribution of the sampling zone on data dispersion was tested by the adonis algorithm, which detected significant differences in soil microbiome composition between sampling zones, but not for roots' or leaves' microbiomes (Table 2). We interpret these results as the demonstration that both sampling zones and sample types determine the microbiome composition, and that the geographical effect was only significant for soils. Fig. 3B shows the different distributions of eight representative Orders among soil samples from the different sampling zones; the complete results from the ANOVA+Tukey's B analyses are shown in the Supplementary Table ST2.

To explore the effects of agricultural practices (including irrigation water) on the microbiome we build a Random Forest classification model in which all samples from Z1 (irrigated with groundwater) were confronted with all samples from Z2 and Z3 (both irrigated with TWW-fed effluents) combined. The Random Forest classifier validated the model, correctly grouping the different samples with an accuracy of 88.9 % (Supplementary Table ST3). The random forest model also assigned a variable importance score to each OTU by estimating the Mean Decrease in Accuracy. The highest scores (i.e., the highest importance for classification) corresponded to OTUs classified in the orders Sphingomonadales, followed by Rhizobiales, Propionibacteriales, Pseudomonadales, Caulobacteriales and Enterobacteriales (See Supplementary Fig. 3). We thus concluded that the prevalence of these orders in the microbiomes was significantly different in Z1 than in Z2 and Z3, although the exact driver or drivers behind these differences could not be determined at present.

3.3. Correlations between microbiome composition and ARG abundance

The correlations between OTU abundance (relative abundances at the Order level) and genetic element prevalence was analyzed for all studied samples (Fig. 4). Only 16 Orders showed significant correlation (Pearson's correlation, fdr correction) with at least one genetic element. Conversely, only *intl1*, *bla*_{TEM}, *bla*_{CTX-M-32} and *sul1* prevalences correlated with the relative abundance of one or more bacterial Orders, plus the total amount of bacterial 16S gene (Fig. 4). All observed correlations were positive, except for the order Rhizobiales, whose relative abundance (which can amount to 100% in some samples, see Fig. 3) showed negative correlations with all five genetic elements (Fig. 4). Complete results from the analysis is shown in Supplementary Table ST4; Fig. 4 shows some of the most significant correlations, with added information about sampling zone and sample type.

4. Discussion

The study of plant-associated microbiomes is gaining a lot of interest in the scientific community for two apparently disparate reasons. First, recent studies using of culture-independent quantitative methods demonstrated the amazing complexity of plant microbiomes, as well as their complex interaction with many physiological functions of the plant (Andreote

et al., 2014; Berg et al., 2014; Hardoim et al., 2008; Reinhold-Hurek and Hurek, 2011; Ruppel et al., 2006; Ryan et al., 2008). Secondly, there is the suspicion that plant may become contaminated with toxicants or potentially pathogenic bacteria or bacterial genetic elements and thus become a risk for consumers, particularly linked to the reuse of waste water for irrigations (Becerra-Castro et al., 2015; Ben et al., 2017; Han et al., 2016; Marti et al., 2013; Piña et al., 2019; van Overbeek et al., 2014; Westhoff et al., 2017; Zhu et al., 2017).

Our results showed that the bacterial abundance in both roots and leaves (and, presumably, in most parts of *Vicia faba* plants) is around 1-10% of the levels found in soils, a figure that is consistent with previous estimations using comparable methodologies for *Brassica oleacea* (Ruppel et al., 2006). These levels drop another 10-fold, at least, in beans. In addition, and with the exception of *bla*_{TEM}, most of the studied genetic elements (*intI1*, ARGs) were only found in significant amounts in soil and root endophytes. These low values in *intI1* and ARGs appear to correlate with the predominance of Rhizobiales in all plant compartments, likely a specific feature of legume plants. In fact, we have data from non-leguminous cultivars showing a radically different distribution of both endophytes and antibiotic resistance-related genetic elements (Cerqueira et al., 2019).

We found a higher prevalence of *intI1* in soil and roots than in the aerial parts of plant, with significant differences between sampling zones. For example, *intI1* was completely absent from beans' samples and showed very low representation in leaves, which nevertheless contained low, but significant levels of *bla*_{TEM} or, in the latter, *qnrS1* sequences. These results are in line with the fact that the presence of integrase genes, like *intI1*, have been associated to a higher potential for dissemination of antibiotic resistance in both clinical samples and different environmental compartments (Domingues et al., 2012). Therefore, our results may suggest that the consumption of beans represents a low human health risk in terms of exposition to antibiotic resistance gens. Furthermore, *intI1* was found at 5-10 fold higher levels in the roots and soils from the field irrigated with groundwater (Z1) compared to those irrigated with river water impacted by TWW (Z2, Z3). While this may reflect the influence of agricultural practices (manure fertilization in the field irrigated with groundwater) in the abundance of ARGs and related genetic elements, as previously suggested (Cerqueira et al., 2019; Piña et al., 2019). It is interesting to remark that Z1 was irrigated with groundwater, so the results indicated that other factors, both geographic and operational (manure application), may be at least equally important for the interpretation of *intI1* abundance.

In addition to *intl1*, only *bla*_{TEM} and *qnrS1* abundances showed clear differences between sampling zones. In the case of *qnrS1*, this variability reflects its virtual absence in Z1 samples, whereas *bla*_{TEM} showed a very consistent distribution pattern for all sample parts, in which the highest prevalence corresponded to Z3 samples. The abundance of *bla*_{TEM} was similar for all plant parts, around 100-1000 copies per g of sample, and clearly higher than the levels on *intl1* in most cases. Then, the results suggest that the endophytic *bla*_{TEM} appears to be independent from the presence of integrons, at least to *intl1*. They also are at variance with the commonly assumed link between the presence of *intl1* and the presence of anthropogenic-driven abundance of ARGs. The high abundance of *bla*_{TEM} in Z3 is consistent with the observation that this ARG is prevalent in sediments impacted by WWTPs (O'Flaherty and Cummins, 2017). Z3 is the only sampling site 92% irrigated with TWW, in comparison with the 18% of TWW for Z2 and essentially nil for Z1.

Different recent studies showed the influence of using TWW, or water with strong TWW contribution, in soil and roots' microbiome from non-legume plants, like *Solanum lycopersicum* or *Lactuca sativa* (Cerqueira et al., 2019; Zolti et al., 2018). Our results showing that only a small part of the OTUs present appeared as responsible for differentiating between fields with different irrigation regimes, may be due to the characteristics of the plant host, as previously suggested (Lundberg et al., 2012; Ofek et al., 2014; Ofek-Lalzar et al., 2014). In our case, it is particularly remarkable the prevalence of Rhizobiales species (mainly, *Rhizobium spp.*) in plant parts' microbiome, up to 100% of bacterial sequences from many leaves' and beans' samples. The presence of Rhizobiales in roots from legume plants is expected, as *Rhizobium* is known to colonize legume roots as a symbiont, where it fixes nitrogen in specific structures called nodules (Ferguson et al., 2018). However, as nitrogen fixation in legumes does not occur outside roots' nodules, their presence (and even predominance) in aerial parts is somewhat surprising. A previous study (Zgadza et al., 2016) found that the enrichment of symbiosis-linked root bacteria imposes distinctive rhizosphere, root, and nodule bacterial communities, affects the soil microbiome. Our results are consistent with this study, suggesting that this influence extends to the microbiome of surrounding soil, with relatively high levels of Rhizobiales, and that it propagates to the rest of the plant, outside the nodule-root system, and eventually becoming the predominant, if not unique, endobacterial group. The role of this presence, that may be assimilated to comensalism, in plant physiology it is unclear. Despite nitrogen fixation cannot occur outside the root nodules, as a rule, bacteria of *Rhizobium* genus

may have other functions that benefit the host plants, such as phosphate solubilization or phytohormone synthesis from abscisic acid to auxins, among others (Alori et al., 2017; Ferguson and Mathesius, 2014). Our data suggests that the prevalence of Rhizobacteria associates to a very low level both of other bacterial types and of ARGs, which may have relevant meaning for both plant (as it may well allow the plant to fend off potentially pathogenic bacteria) and for consumer health. In fact, the extremely low levels of ARGs and of non-Rhizobial bacterial groups in beans, in contrast with the corresponding high levels in soils, suggests that the requirements for water quality may be lower for legume crops than for other cultivars in terms of microbial food safety. TWW irrigation is currently limited for certain crops in different countries, like lettuces and other leafy food products in Cyprus (Kalavrouziotis et al., 2015). Whereas the main rationale for these limitations is the way they are consumed (i.e., raw *versus* cooked), our results suggest that the Physiology of the crop plant should be considered to set water standard qualities for irrigation of different edible crops.

Acknowledgments:

The authors want to acknowledge the staff from the Institute of Hydrobiology of the Technische Universität Dresden, Germany for providing the pUC19 plasmids, and to Mr. Roberto Marano and Dr. Eddie Cytrin, from Agriculture Research Organisation of Israel - The Volcani Center, Israel for providing the pNORM1 plasmids.

Funding:

This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675530.

Disclaimer for ANSWER MSCA No 675530: The content of this article reflects only the authors' views and the Research Executive Agency is not responsible for any use that may be made of the information it contains.

References

Agerso, Y., and Sandvang, D. (2005). Class 1 integrons and tetracycline resistance genes in *alcaligenes*, *arthrobacter*, and *Pseudomonas* spp. isolated from pigsties and manured soil. *Appl Environ Microbiol* 71, 7941-7947.

Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J., and Handelsman, J. (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nature Reviews Microbiology* 8, 251-259.

Alori, E.T., Glick, B.R., and Babalola, O.O. (2017). Microbial Phosphorus Solubilization and Its Potential for Use in Sustainable Agriculture. *Frontiers in Microbiology* 8, 971.

Andreote, F.D., Gumiore, T., and Durrer, A. (2014). Exploring interactions of plant microbiomes. *Scientia Agricola* 71, 528-539.

Becerra-Castro, C., Lopes, A.R., Vaz-Moreira, I., Silva, E.F., Manaia, C.M., and Nunes, O.C. (2015). Wastewater reuse in irrigation: A microbiological perspective on implications in soil fertility and human and environmental health. *Environment International* 75, 117-135.

Ben, W.W., Wang, J., Pan, X., and Qiang, Z.M. (2017). Dissemination of antibiotic resistance genes and their potential removal by on-farm treatment processes in nine swine feedlots in Shandong Province, China. *Chemosphere* 167, 262-268.

Berendonk, T.U., Manaia, C.M., Merlin, C., Fatta-Kassinos, D., Cytryn, E., Walsh, F., Buergermann, H., Sorum, H., Norstrom, M., Pons, M.-N., *et al.* (2015). Tackling antibiotic resistance: the environmental framework. *Nature Reviews Microbiology* 13, 310-317.

Berg, G., Grube, M., Schlöter, M., and Smalla, K. (2014). Unraveling the plant microbiome: looking back and future perspectives. *Frontiers in Microbiology* 5.

Bertini, A., Poirol, L., Bernabeu, S., Fortini, D., Villa, L., Nordmann, P., and Carattoli, A. (2007). Multicopy *bla*(OXA-58) gene as a source of high-level resistance to carbapenems in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy* 51, 2324-2328.

Bhullar, K., Waglechner, N., Pawlowski, A., Koteva, K., Banks, E.D., Johnston, M.D., Barton, H.A., and Wright, G.D. (2012). Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS One* 7, e34953.

Bodenhausen, N., Horton, M.W., and Bergelson, J. (2013). Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS One* 8, e56329.

Bradford, P.A. (2001). Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. *Clin Microbiol Rev* 14, 933-951, table of contents.

Breiman, L. (2001). Random forests. *Machine Learning* 45, 5-32.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13, 581-+.

Cerqueira, F., Matamoros, V., Tauler, R., Bayona, J.M., and Piña, B. (2019). Antibiotic resistance genes distribution in microbiomes from the soil-plant-fruit continuum in commercial *Lycopersicon esculentum* fields under different agricultural practices. *Science of the Total Environment* 665, 660-670.

Chen, F.-J., Wang, C.-H., Chen, C.-Y., Hsu, Y.-C., and Wang, K.-T. (2014). Role of the *mecA* Gene in Oxacillin Resistance in a *Staphylococcus aureus* Clinical Strain with a *pvl*-Positive ST59 Genetic Background. *Antimicrobial Agents and Chemotherapy* 58, 1047-1054.

Chen, Q.L., An, X.L., Li, H., Su, J.Q., Ma, Y.B., and Zhu, Y.G. (2016). Long-term field application of sewage sludge increases the abundance of antibiotic resistance genes in soil. *Environment International* 92-93, 1-10.

Doenhoefer, A., Franckenberg, S., Wickles, S., Berninghausen, O., Beckmann, R., and Wilson, D.N. (2012). Structural basis for TetM-mediated tetracycline resistance. *Proceedings of the National Academy of Sciences of the United States of America* 109, 16900-16905.

Domingues, S., da Silva, G.J., and Nielsen, K.M. (2012). Integrons: Vehicles and pathways for horizontal dissemination in bacteria. *Mob Genet Elements* 2, 211-223.

ECDC/EMEA Joint Working Group (2009). The Bacterial Challenge: Time to React (Stockholm: European Centre for Disease Prevention and Control,).

Ferguson, B.J., and Mathesius, U. (2014). Phytohormone Regulation of Legume-Rhizobia Interactions. *Journal of Chemical Ecology* 40, 770-790.

Ferguson, B.J., Mens, C., Hastwell, A.H., Zhang, M., Su, H., Jones, C.H., Chu, X., and Gresshoff, P.M. (2018). Legume nodulation: The host controls the party. *Plant, cell & environment*.

Frank, T., Gautier, V., Talarmin, A., Bercion, R., and Arlet, G. (2007). Characterization of sulphonamide resistance genes and class 1 integron gene cassettes in *Enterobacteriaceae*, Central African Republic (CAR). *Journal of Antimicrobial Chemotherapy* 59, 742-745.

Gat, D., Mazar, Y., Cytryn, E., and Rudich, Y. (2017). Origin-Dependent Variations in the Atmospheric Microbiome Community in Eastern Mediterranean Dust Storms. *Environmental Science & Technology* 51, 6709-6718.

Gillings, M.R., Gaze, W.H., Pruden, A., Smalla, K., Tiedje, J.M., and Zhu, Y.-G. (2015). Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *Isme Journal* 9, 1269-1279.

Gillings, M.R., Krishnan, S., Worden, P.J., and Hardwick, S.A. (2008). Recovery of diverse genes for class 1 integron-integrases from environmental DNA samples. *FEMS Microbiol Lett* 287, 56-62.

Han, X.M., Hu, H.W., Shi, X.Z., Wang, J.T., Han, L.L., Chen, D.L., and He, J.Z. (2016). Impacts of reclaimed water irrigation on soil antibiotic resistome in urban parks of Victoria, Australia. *Environmental Pollution* 211, 48-57.

Hardoim, P.R., van Overbeek, L.S., and van Elsas, J.D. (2008). Properties of bacterial endophytes and their proposed role in plant growth. *Trends in Microbiology* 16, 463-471.

Hartmann, M., Frey, B., Mayer, J., Mader, P., and Widmer, F. (2015). Distinct soil microbial diversity under long-term organic and conventional farming. *The ISME journal* 9, 1177-1194.

Hembach, N., Schmid, F., Alexander, J., Hiller, C., Rogall, E.T., and Schwartz, T. (2017). Occurrence of the mcr-1 Colistin Resistance Gene and other Clinically Relevant Antibiotic Resistance Genes in Microbial Populations at Different Municipal Wastewater Treatment Plants in Germany. *Frontiers in Microbiology* 8.

Heuer, H., Schmitt, H., and Smalla, K. (2011). Antibiotic resistance gene spread due to manure application on agricultural fields. *Curr Opin Microbiol* 14, 236-243.

Hu, J.Y., Shi, J.C., Chang, H., Li, D., Yang, M., and Kamagata, Y.C. (2008). Phenotyping and genotyping of antihiotic-resistant *Escherichia coli* isolated from a natural river basin. *Environmental Science & Technology* 42, 3415-3420.

Hu, Y.F., Yang, X., Li, J., Lv, N., Liu, F., Wu, J., Lin, I.Y.C., Wu, N., Weimer, B.C., Gao, G.F., *et al.* (2016). The Bacterial Mobile Resistome Transfer Network Connecting the Animal and Human Microbiomes. *Applied and Environmental Microbiology* 82, 6672-6681.

Kalavrouziotis, I.K., Kokkinos, P., Oron, G., Fatone, F., Bolzonella, D., Vatyliotou, M., Fatta-Kassinos, D., Koukoulakis, P.H., and Varnavas, S.P. (2015). Current status in wastewater treatment, reuse and research in some mediterranean countries. *Desalination and Water Treatment* 53, 2015-2030.

Labuschagne, C.D.J., Weldhagen, G.F., Ehlers, M.M., and Dove, M.G. (2008). Emergence of class 1 integron-associated GES-5 and GES-5-like extended-spectrum beta-lactamases in clinical isolates of *Pseudomonas aeruginosa* in South Africa. *Int J Antimicrob Agents* 31, 527-530.

Laht, M., Karkman, A., Voolaid, V., Ritz, C., Tenson, T., Virta, M., and Kisand, V. (2014). Abundances of Tetracycline, Sulphonamide and Beta-Lactam Antibiotic Resistance Genes in Conventional Wastewater Treatment Plants (WWTPs) with Different Waste Load. *Plos One* 9.

Le-Minh, N., Khan, S.J., Drewes, J.E., and Stuetz, R.M. (2010). Fate of antibiotics during municipal water recycling treatment processes. *Water Research* 44, 4295-4323.

Li, X., Shi, L., Yang, W., Li, L., and Yamasaki, S. (2006). New array of aacA4-catB3-dfrA1 gene cassettes and a noncoding cassette from a class-1-integron-positive clinical strain of *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 50, 2278-2279.

Lind, K., Stahlberg, A., Zoric, N., and Kubista, M. (2006). Combining sequence-specific probes and DNA binding dyes in real-time PCR for specific nucleic acid quantification and melting curve analysis. *Biotechniques* 40, 315-319.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 73, 1576-1585.

Lundberg, D.S., Lebeis, S.L., Paredes, S.H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., del Rio, T.G., *et al.* (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488, 86-+.

Manaia, C.M., Rocha, J., Scaccia, N., Marano, R., Radu, E., Biancullo, F., Cerqueira, F., Fortunato, G., Iakovides, I.C., Zammit, I., *et al.* (2018). Antibiotic resistance in wastewater treatment plants: Tackling the black box. *Environ Int* 115, 312-324.

Marcé, R., Honey-Rosés, J., Manzano, A., Moragas, L., Catllar, B., and Sabater, S. (2012). The Llobregat River Basin: A Paradigm of Impaired Rivers Under Climate Change Threats. In *Handbook of Environmental Chemistry*, pp. 1-26.

Marti, R., Scott, A., Tien, Y.C., Murray, R., Sabourin, L., Zhang, Y., and Topp, E. (2013). Impact of Manure Fertilization on the Abundance of Antibiotic-Resistant Bacteria and Frequency of Detection of Antibiotic Resistance Genes in Soil and on Vegetables at Harvest. *Applied and Environmental Microbiology* 79, 5701-5709.

McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *Plos One* 8.

Mendes, R.E., Castanheira, M., Toleman, M.A., Sader, H.S., Jones, R.N., and Walsh, T.R. (2007). Characterization of an integron carrying bla_{MP-1} and a new aminoglycoside resistance gene, aac(6')-31, and its dissemination among genetically unrelated clinical isolates in a Brazilian hospital. *Antimicrob Agents Chemother* 51, 2611-2614.

Michael, I., Rizzo, L., McArdell, C.S., Manaia, C.M., Merlin, C., Schwartz, T., Dagot, C., and Fatta-Kassinos, D. (2013). Urban wastewater treatment plants as hotspots for the release of antibiotics in the environment: A review. *Water Research* 47, 957-995.

Mugnaioli, C., Luzzaro, F., De Luca, F., Brigante, G., Perilli, M., Amicosante, G., Stefani, S., Toniolo, A., and Rossolini, G.M. (2006). CTX-M-type extended-spectrum beta-lactamases in Italy: molecular epidemiology of an emerging countrywide problem. *Antimicrob Agents Chemother* 50, 2700-2706.

Mukherjee, S., and Chakraborty, R. (2006). Incidence of class 1 integrons in multiple antibiotic-resistant Gram-negative copiotrophic bacteria from the River Torsa in India. *Research in microbiology* 157, 220-226.

Nonaka, L., Maruyama, F., Miyamoto, M., Miyakoshi, M., Kurokawa, K., and Masuda, M. (2012). Novel Conjugative Transferable Multiple Drug Resistance Plasmid pAQU1 from *Photobacterium damsela* subsp. *damsela* Isolated from Marine Aquaculture Environment. *Microbes and Environments* 27, 263-272.

O'Flaherty, E., and Cummins, E. (2017). Antibiotic resistance in surface water ecosystems: Presence in the aquatic environment, prevention strategies, and risk assessment. *Hum Ecol Risk Assess* 23, 299-322.

Ofek, M., Voronov-Goldman, M., Hadar, Y., and Minz, D. (2014). Host signature effect on plant root-associated microbiomes revealed through analyses of resident vs. active communities. *Environ Microbiol* 16, 2157-2167.

Ofek-Lalzar, M., Sela, N., Goldman-Voronov, M., Green, S.J., Hadar, Y., and Minz, D. (2014). Niche and host-associated functional signatures of the root surface microbiome. *Nature communications* 5, 4950.

Piña, B., Bayona, J.M., Christou, A., Fatta-Kassinos, D., Guillon, E., Lambropoulou, D., Michael, C., Polesel, F., and Sayen, S. (2019). On the contribution of reclaimed wastewater irrigation to the potential exposure of humans to antibiotics, antibiotic resistant bacteria and antibiotic resistance genes - NEREUS COST Action ES1403 position paper. *Journal of Environmental Chemical Engineering in press*.

Qvit-Raz, N., Jurkevitch, E., and Belkin, S. (2008). Drop-size soda lakes: Transient microbial habitats on a salt-secreting desert tree. *Genetics* 178, 1615-1622.

Reinhold-Hurek, B., and Hurek, T. (2011). Living inside plants: bacterial endophytes. *Curr Opin Plant Biol* 14, 435-443.

Ruppel, S., Ruhlmann, J., and Merbach, W. (2006). Quantification and localization of bacteria in plant tissues using quantitative real-time PCR and online emission fingerprinting. *Plant Soil* 286, 21-35.

Ryan, R.P., Germaine, K., Franks, A., Ryan, D.J., and Dowling, D.N. (2008). Bacterial endophytes: recent developments and applications. *Fems Microbiology Letters* 278, 1-9.

Stokes, H.W., Nesbo, C.L., Holley, M., Bahl, M.I., Gillings, M.R., and Boucher, Y. (2006). Class 1 integrons potentially predating the association with tn402-like transposition genes are present in a sediment microbial community. *J Bacteriol* 188, 5722-5730.

Szczepanowski, R., Linke, B., Krahn, I., Gartemann, K.-H., Guetzkow, T., Eichler, W., Puehler, A., and Schlueter, A. (2009). Detection of 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to selected antibiotics. *Microbiology-(UK)* 155, 2306-2319.

Tamang, M.D., Seol, S.Y., Oh, J.Y., Kang, H.Y., Lee, J.C., Lee, Y.C., Cho, D.T., and Kim, J. (2008). Plasmid-Mediated Quinolone Resistance Determinants qnrA, qnrB, and qnrS among Clinical Isolates of Enterobacteriaceae in a Korean Hospital. *Antimicrobial Agents and Chemotherapy* 52, 4159-4162.

Tamminen, M., Karkman, A., Lohmus, A., Muziasari, W.I., Takasu, H., Wada, S., Suzuki, S., and Virta, M. (2011). Tetracycline Resistance Genes Persist at Aquaculture Farms in the Absence of Selection Pressure. *Environmental Science & Technology* 45, 386-391.

Tennstedt, T., Szczepanowski, R., Braun, S., Puhler, A., and Schluter, A. (2003). Occurrence of integron-associated resistance gene cassettes located on antibiotic resistance plasmids isolated from a wastewater treatment plant. *FEMS microbiology ecology* 45, 239-252.

van Overbeek, L.S., van Doorn, J., Wichers, J.H., van Amerongen, A., van Roermund, H.J.W., and Willemsen, P.T.J. (2014). The arable ecosystem as battleground for emergence of new human pathogens. *Frontiers in Microbiology* 5.

Wang, F.-H., Qiao, M., Chen, Z., Su, J.-Q., and Zhu, Y.-G. (2015). Antibiotic resistance genes in manure-amended soil and vegetables at harvest. *Journal of Hazardous Materials* 299, 215-221.

- Westhoff, S., van Leeuwe, T.M., Qachach, O., Zhang, Z., van Wezel, G.P., and Rozen, D.E. (2017). The evolution of no-cost resistance at sub-MIC concentrations of streptomycin in *Streptomyces coelicolor*. *Isme Journal* *11*, 1168-1178.
- Wielders, C.L.C., Fluit, A.C., Brisse, S., Verhoef, J., and Schmitz, F.J. (2002). *mecA* gene is widely disseminated in *Staphylococcus aureus* population. *Journal of Clinical Microbiology* *40*, 3970-3975.
- Zgadaj, R., Garrido-Oter, R., Jensen, D.B., Koprivova, A., Schulze-Lefert, P., and Radutoiu, S. (2016). Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* *113*, E7996-E8005.
- Zhu, B.K., Chen, Q.L., Chen, S.C., and Zhu, Y.G. (2017). Does organically produced lettuce harbor higher abundance of antibiotic resistance genes than conventionally produced? *Environment International* *98*, 152-159.
- Zolti, A., Green, S.J., Ben Mordechay, E., Hadar, Y., and Minz, D. (2018). Root microbiome response to treated wastewater irrigation. *Sci Total Environ* *655*, 899-907.

Figure Legends

Figure 1: Absolute abundances of the different genetic elements in all samples, separated by sample types (Beans, Leaves, Roots, and Soil). Data are expressed as copies of each sequence per g of tissue, and code colored by zones as in Fig. 1: cyan, yellow and gray correspond to Zones 1, 2, and 3, respectively. Lowercase letters a to c, when corresponding) indicate statistically different groups of data ($p < 0.05$, ANOVA+Tukey HSD test); absence of letters indicates no statistical differences between zones. Sample points not fitting the normal distribution (outliers) are represented by empty dots.

Figure 2: Hierarchical clustering of samples grouped by their abundance in OTUs, grouped at taxonomical order level. The graph shows the log-transformed absolute number of reads for each Order and sample. Only OTUs representing more than 0.1% of the total abundance were included in the analysis. The left vertical side bar indicates the different sample types, as brown, orange, light green and dark green represent soil, root endophytes, leaves and beans, respectively. The left-most bar graph shows the total number of DNA reads included on the analysis (note that plastid sequences were excluded), in logarithmic annotation. Colors indicate sampling zones, coded as in Fig. 1. Cell colors follow the classical heatmap scheme, in which red indicates the lowest values (in this case, absence of the sequences, transformed to 0.1 to allow log transformation) and white the highest ones.

Figure 3. Analysis of OTU distribution, grouped at the taxonomic order level. **A)**

NMDS plot for all samples, labelled by sample type as in Fig. 2, and by sampling zone (triangles, diamonds and circles represent samples from Z1, Z2 and Z3, respectively).

B) Taxon distribution in soil samples from the different sampling zones. Graphs present absolute abundances (read counts) for different taxonomical Orders in soil samples from Z1, Z2, and Z3 (labelled as in Fig. 1, 2). Boxes indicate the first to 3rd quartile interval, whiskers correspond to the 95% distribution interval, and the thick line indicates the distribution mean. Low-case letters at the top of the graphs indicate statistically different groups of data ($p < 0.05$, ANOVA+Tukey HSD test).

Figure 4. Correlation between genetic element abundance (in copies per g of sample) and relative proportion of different taxonomical Order in all samples. **A)** Heatmap with hierarchical clustering. Red and blue cells indicate positive and negative correlations, respectively, asterisks indicate significant correlations (Pearson's + fdr correction, one, two or three asterisks represent p-values under 0.05, 0.01 and 0.001, respectively) only Orders with at least one significant correlation were included

B) Individual biplots showing different examples of significant correlations between genetic elements and relative abundances of some bacterial Orders. Samples are shown individually, with colors representing the different sample types, and symbols indicating sampling zones, as in Fig. 3

Figure 1

Figure 2

Figure 4

Supplementary Material

Distribution of Antibiotic Resistance Genes in legume plants along the soil-plant continuum (*Vicia faba* L.) grown under different watering regimes. A field study.

Authors: Francisco Cerqueira¹, Víctor Matamoros¹, Josep Bayona¹, Goffe Elsinga², Luc M. Hornstra², Benjamin Piña¹

¹Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research, 08034, Barcelona, Spain

²KWR Watercycle Research Institute, Groningenhaven 7, 3433 PE, Nieuwegein, the Netherlands

Supplementary Figure 1. Geographic location of sampling sites

Supplementary Figure 2. Results from the sequencing data, separated by sampling zone and sample type. The vertical bars correspond to the relative proportion of Mitochondrial (yellow), chloroplast (green) and bacterial (red) sequences for each group of samples (left axis); dots indicate the average amount of bacterial 16S, in copies per g of sample) corresponding to each group and calculated by qPCR combined with 16S amplicon sequencing (right axis, note the log scale). Note the lack of bacterial sequences in bean samples from zone Z3.

Supplementary Figure 3. Relative importance of the top 20 Orders for sample discrimination, as calculated by the random forest analysis.

Supplementary Table ST1. Primers utilized for detection and quantifications of antibiotic resistance genes in soil and crops

Target gene	Primers	Sequence (5'→ 3')	Amplicon size (bp)	Tm(°C)	Reference
intI1	intI1LC5	GATCGGTGCGAATGCGTGT	196	60	Barraud et al, 2010
	intI1LC1	GCCTTGATGTTACCCGAGAG			
16s rDNA	331F	TCCTACGGGAGGCAGCAGT	195	60	Bräuer et al., 2011
	518R	ATTACCGCGGCTGCTGG			
bla _{TEM}	blaTEM-F	TTCCTGTTTTTGCTCACCCAG	113	60	Di Cesare et al., 2016
	blaTEM-R	CTCAAGGATCTTACCCTGTTG			
bla _{ctx-m-32}	ctx-m-32-FW	CGTCACGCTGTTGTTAGGAA	156	60	Hembach et al., 2017
	ctx-m-32-R	CGCTCATCAGCACGATAAAG			
bla _{OXA-58}	OXA58F	GCAATTGCCTTTTAAACCTGA	152	63	Laht et al., 2014
	OXA58R	CTGCCTTTTCAACAAAACCC			
mecA	mecAF	AAAAAGATGGCAAAGATATTCAA	185	63	Szczepanowski et al., 2009
	mecAR	TTCTTCGTTACTCATGCCATACA			
qnrS	qnrSrtF11	GACGTGCTAACTTGC GTGAT	118	60	Marti & Balcázar, 2013
	qnrSrtR11	TGGCATTGTTGGAAACTTG			
sul1	sul1-FW	CGCACCGGAAACATCGTGCAC	162	60	Pei et al., 2006
	sul1-RV	TGAAGTTCGCCGCAAGGCTCG			
tetM	tetMF	GCAATTCTACTGATTCTGC	186	60	Tamminen & Karkman, 2011
	tetMR	CTGTTTGATTACAATTTCCGC			

References

- Barraud, O., Baclet, M. C., Denis, F., & Ploy, M. C. (2010). Quantitative multiplex real-time PCR for detecting class 1, 2 and 3 integrons. *Journal of Antimicrobial Chemotherapy*, 65(8), 1642–1645. <https://doi.org/10.1093/jac/dkq167>
- Bräuer, S. L., Adams, C., Kranzler, K., Murphy, D., Xu, M., Zuber, P., ... Tebo, B. M. (2011). Culturable Rhodobacter and Shewanella species are abundant in estuarine turbidity maxima of the Columbia River. *Environmental Microbiology*, 13(3), 589–603. <https://doi.org/10.1111/j.1462-2920.2010.02360.x>
- Di Cesare, A., Losasso, C., Barco, L., Eckert, E. M., Conficoni, D., Sarasini, G., ... Ricci, A. (2016). Diverse distribution of Toxin-Antitoxin II systems in Salmonella enterica serovars. *Scientific Reports*, 6, 1–9. <https://doi.org/10.1038/srep28759>
- Hembach, N., Schmid, F., Alexander, J., Hiller, C., Rogall, E. T., & Schwartz, T. (2017). Occurrence of the mcr-1 colistin resistance gene and other clinically relevant antibiotic resistance genes in microbial populations at different municipal wastewater treatment plants in Germany. *Frontiers in Microbiology*, 8(JUL), 1–11. <https://doi.org/10.3389/fmicb.2017.01282>
- Laht, M., Karkman, A., Voolaid, V., Ritz, C., Tenson, T., Virta, M., & Kisand, V. (2014). Abundances of tetracycline, sulphonamide and beta-lactam antibiotic resistance genes in conventional wastewater treatment plants (WWTPs) with different waste load. *PLoS ONE*, 9(8), 1–8. <https://doi.org/10.1371/journal.pone.0103705>
- Marti, E., & Balcázar, J. L. (2013). Real-time PCR assays for quantification of qnr genes in environmental water samples and chicken feces. *Applied and Environmental Microbiology*, 79(5), 1743–1745. <https://doi.org/10.1128/AEM.03409-12>
- Pei, R., Kim, S. C., Carlson, K. H., & Pruden, A. (2006). Effect of River Landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). *Water Research*, 40(12), 2427–2435. <https://doi.org/10.1016/j.watres.2006.04.017>
- Szczepanowski, R., Linke, B., Krahn, I., Gartemann, K. H., Gützkow, T., Eichler, W., ... Schlüter, A. (2009). Detection of 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to selected antibiotics. *Microbiology*, 155(7), 2306–2319. <https://doi.org/10.1099/mic.0.028233-0>
- Tamminen, M., & Karkman, A. (2011). Tetracycline Resistance Genes Persist at Aquaculture Farms in the Absence of Selection Pressure Tetracycline Resistance Genes Persist at Aquaculture Farms in the Absence of Selection Pressure, 45(November 2015), 386–391. <https://doi.org/10.1021/es102725n>

Supplementary Table ST2. OTU order distribution among soil samples from the different sampling zones (Z1-Z3)

Orders	Abundances ^a			Tukey's B test ^b			ANOVA
	Z1	Z2	Z3	Z1	Z2	Z3	p value
Methylophilales	-0.62	-0.59	-0.34	b	b	a	0.0025
Rhizobiales	1.85	0.57	-0.12	a	b	b	0.0027
Verrucomicrobiales	-0.63	-0.56	-0.34	b	ab	a	0.0027
Enterobacteriales	-0.66	-0.59	-0.35	b	ab	a	0.0029
Gemmatimonadales	1.36	1.38	-0.22	a	a	b	0.0060
Unknown_Order	-0.53	-0.55	-0.34	a	a	a	0.0093
Rhodospirillales	1.07	0.89	0.05	a	ab	b	0.0143
Pseudomonadales	-0.36	-0.58	-0.26	a	a	a	0.0147
Phycisphaerales	-0.48	-0.57	-0.34	ab	b	a	0.0153
Anaerolineales	-0.08	-0.45	-0.20	a	a	a	0.0175
Acidimicrobiales	-0.40	-0.55	-0.33	a	a	a	0.0207
Burkholderiales	-0.58	-0.53	-0.33	b	ab	a	0.0236
Sphingomonadales	1.01	3.21	2.50	a	a	a	0.0278
Rhodobacterales	-0.54	-0.31	-0.27	a	a	a	0.1358
Flavobacteriales	-0.10	-0.27	1.87	a	a	a	0.1784
Propionibacteriales	1.02	0.55	-0.04	a	a	a	0.2775
Cytophagales	-0.26	-0.19	-0.06	a	a	a	0.3337
Blastocatellales	-0.34	-0.01	-0.25	a	a	a	0.3833
Caulobacterales	-0.34	-0.42	-0.30	a	a	a	0.3986
Xanthomonadales	-0.39	-0.42	-0.34	a	a	a	0.5934

a) Log transformation of absolute abundance (counts) in soil samples (average values)

b) Letters indicate statistically different distribution groups

Supplementary Table ST3. Detailed results of the Random Forest Classifier model predictions, based on the OTUs abundances

Accuracy	0.8889
95% Confidence Interval	0.6529-0.9862
No information Rate	0.6667
p-value	0.032
Kappa	0.727
Mcnemar's Test p-value	0.47950
Precision	1
Recall	0.666
F1	0.8
Prevalence	0.333
Detection Rate	0.222
Detection Prevalence	0.222
Balanced Accuracy	0.833

Supplementary Table ST4. Pearson's correlations^a between OTUs order prevalences and abundance of different genetic elements (all samples, only orders with at least one significant correlation)

	bacterial 16S	<i>bla</i> _{TEM}	<i>intl1</i>	<i>sul1</i>	<i>bla</i> _{CTX-M-32}	<i>tetM</i>	<i>qnrS1</i>	<i>bla</i> _{OXA-58}
Acidimicrobiales	0.78 ***	0.79 ***	0.82 ***	0.69 **	0.64 *	0.36	0.23	0.23
Anaerolineales	0.78 ***	0.75 ***	0.88 ***	0.65 **	0.61 *	0.40	0.10	0.19
Blastocatellales	0.80 ***	0.79 ***	0.74 **	0.59 *	0.58 *	0.34	0.34	0.28
Cytophagales	0.72 **	0.77 ***	0.61 *	0.48	0.54 *	0.22	0.31	0.17
Flavobacteriales	0.59 *	0.62 *	0.60 *	0.39	0.44	0.19	0.00	-0.04
Gemmatimonadales	0.80 ***	0.74 **	0.88 ***	0.66 **	0.60 *	0.45	0.22	0.33
Methylophilales	0.39	0.36	0.58 *	0.30	0.29	0.22	-0.29	-0.13
Phycisphaerales	0.79 ***	0.73 **	0.90 ***	0.67 **	0.59 *	0.46	0.14	0.28
Propionibacteriales	0.70 **	0.57 *	0.80 ***	0.54 *	0.46	0.41	-0.02	0.20
Rhizobiales	-0.79 ***	-0.86 ***	-0.71 **	-0.56 *	-0.60 *	-0.28	-0.28	-0.20
Rhodobacterales	0.78 ***	0.81 ***	0.79 ***	0.56 *	0.60 *	0.37	0.24	0.26
Rhodospirillales	0.79 ***	0.76 ***	0.78 ***	0.62 *	0.56 *	0.40	0.29	0.31
Sphingomonadales	0.57 *	0.58 *	0.38	0.28	0.40	0.08	0.24	0.08
Unknown_Order	0.79 ***	0.73 **	0.90 ***	0.67 **	0.59 *	0.46	0.14	0.28
Verrucomicrobiales	0.46	0.55 *	0.50	0.26	0.33	0.32	0.06	0.18
Xanthomonadales	0.49	0.51	0.52 *	0.34	0.32	0.31	0.05	0.15

^a) Pearson Correlations. *, p≤0.05; **, p≤0.01, ***, p≤0.001; n=60