the plant journal

METABOLITE PROFILING OF NON-STERILE RHIZOSPHERE SOIL

Journal:	*The Plant Journal*
Manuscript ID	TPJ-00076-2017.R2
Manuscript Type:	Technical Advance
Date Submitted by the Author:	11-Jul-2017
Complete List of Authors:	Pétriacq, Pierre; University of Sheffield, Animal and Plant Sciences Williams, Alex; University of Sheffield, Animal and Plant Sciences Cotton, T. E. Anne; University of Sheffield, Animal and Plant Sciences McFarlane, Alexander; University of Sheffield, Animal and Plant Sciences Rolfe, Stephen; University of Sheffield, Animal and Plant Sciences Ton, Jurriaan; University of Sheffield, Animal and Plant Sciences;
Key Words:	rhizosphere chemistry, rhizosphere microbiome, root exudates, metabolomics, Arabidopsis thaliana, soil, maize, benzoxazinoids

METABOLITE PROFILING OF NON-STERILE RHIZOSPHERE SOIL.

Pierre Pétriacq[1, 2, 3] *, Alex Williams[3], T. E. Anne Cotton[3], Alexander E. McFarlane[3], Stephen A. Rolfe[1,3], & Jurriaan Ton[1, 3] *

[1] Plant Production and Protection (P^3) Institute for Translational Plant & Soil Biology, Department of Animal and Plant Sciences, The University of Sheffield, Sheffield, S10 2TN, UK

[2] *bi*OMICS Facility, Department of Animal and Plant Sciences, The University of Sheffield, Sheffield, S10 2TN, UK

[3] Department of Animal and Plant Sciences, The University of Sheffield, Sheffield, S10 2TN, UK

* Corresponding authors: p.petriacq@sheffield.ac.uk and j.ton@sheffield.ac.uk

Running title: rhizosphere metabolomics.

Keywords: rhizosphere chemistry, rhizosphere microbiome, root exudates, metabolomics, Arabidopsis, maize, soil, benzoxazinoids.

Significance statement: Multitrophic interactions in the rhizosphere are critical for plant growth and health, and are influenced by root exudates and their microbial breakdown products. In this study, we describe a straightforward method for metabolic profiling of non-sterile rhizosphere soil, which represents a powerful

technique to identify novel semiochemicals that shape the microbial community structure and activity of the rhizosphere.

Word Count: Core (8955), Complete (14730), Summary (220), Significance Statement (58), Introduction (987), Results (3404), Discussion (1941), Experimental Procedures (977), Acknowledgements (89), Short Legends for Supporting Information (115), References (2110), Figure Legends (1337), Supplemental Figure Legends (758), Supplemental Methods (2734).

SUMMARY

Rhizosphere chemistry is the sum of root exudation chemicals, their breakdown products and microbial products of soil-derived chemicals. To date, most studies about root exudation chemistry are based on sterile cultivation systems, which limits the discovery of microbial breakdown products that act as semiochemicals and shape microbial rhizosphere communities. Here, we present a method for untargeted metabolic profiling of non-sterile rhizosphere soil. We have developed an experimental growth system that enables collection and analysis of rhizosphere chemicals from different plant species. High-throughput sequencing of 16S rRNA genes demonstrated that plants in the growth system support a microbial rhizosphere effect. To collect a range of (a)polar chemicals from the system, we developed extraction methods that do not cause detectable damage to root cells or soil-inhabiting microbes, thus preventing contamination with cellular metabolites. Untargeted metabolite profiling by UPLC-Q-TOF mass spectrometry, followed by uni- and multivariate statistical analyses identified a wide range of secondary metabolites that are enriched in plant-containing soil compared to control soil without roots. We show that the method is suitable for profiling rhizosphere chemistry of maize in agricultural soil, demonstrating applicability to different plant-soil combinations. Our study provides a robust method for comprehensive metabolite profiling of non-sterile rhizosphere soil, which represents a technical advance towards the establishment of causal relationships between the chemistry and microbial composition of the rhizosphere.

INTRODUCTION

Plant roots convert their associated soil into complex mesotrophic environments which support a highly diverse microbial community (Dessaux *et al.*, 2016). This so-called rhizosphere effect is mediated by exudation of plant metabolites from roots (van Dam and Bouwmeester, 2016; Oburger and Schmidt, 2016; Badri and Vivanco, 2009). The chemical composition of these root exudates and their microbial breakdown products plays a crucial role in rhizosphere interactions between plants and beneficial soil microbes (Oburger and Schmidt, 2016). While developments in sequencing technology have revolutionised our ability to characterise rhizosphere microbial communities (van Dam and Bouwmeester, 2016; Oburger and Schmidt, 2016), the chemical diversity of the rhizosphere remains largely unexplored. This knowledge gap is mostly due to a lack of suitable methods to collect and comprehensively analyse metabolites from non-sterile rhizosphere soil.

It has been estimated that plants exude up to 21% of their carbon through their roots, where it is metabolised by the microbial community in the rhizosphere (Badri and Vivanco, 2009; Neumann *et al.*, 2009; Hinsinger *et al.*, 2006). Hence, plant roots drive multitrophic interactions in the rhizosphere via root exudation chemistry. Apart from serving as a primary carbon source for rhizosphere microbes, root exudates can influence rhizosphere interactions via selective biocidal and/or signalling activity (Berendsen *et al.*, 2012). Both polar and apolar compounds have been reported to influence rhizosphere interactions. In addition to polar primary metabolites, such as organic and amino acids (Rudrappa *et al.*, 2008; van Dam and Bouwmeester, 2016; Ziegler *et al.*, 2015), more complex apolar secondary

metabolites, like flavonoids, coumarins and benzoxazinoids (Hassan and Mathesius, 2012; Neal *et al.*, 2012; Szoboszlay *et al.*, 2016; Ziegler *et al.*, 2015), have been reported to play an important role in influencing rhizosphere microbes. For instance, the benzoxazinoid DIMBOA, which is exuded by roots of maize seedlings, has chemotactic properties on *Pseudomonas putida* KT2440 (Neal *et al.*, 2012), a rhizobacterial strain that primes host defences against herbivores (Neal and Ton, 2013). Likewise, the release of malic acid from *Arabidopsis thaliana* (Arabidopsis) roots attracts the Gram-positive rhizobacteria *Bacillus subtilis*, which in turn induces disease resistance against *Pseudomonas syringae* pv. *tomato* (Rudrappa *et al.*, 2008). Furthermore, it was shown recently that plant-derived flavonoids have profound impacts on the structure of soil bacterial communities (Szoboszlay *et al.*, 2016). Although these studies illustrate the importance of specific classes of root-derived chemicals in rhizosphere interactions, untargeted metabolome studies of root exudation products remain scarce, thereby limiting scope for discoveries of important rhizosphere signals (Lakshmanan *et al.*, 2012; Neal *et al.*, 2012).

In addition to plant genotype and nutrition, various other factors can influence root exudation chemistry, such as plant developmental stage, temperature, humidity, and physiochemical soil properties (Zhang *et al.*, 2016; Boyes *et al.*, 2001; Badri and Vivanco, 2009; Uren, 2007). Environmental effects of root exudation chemistry has been studied mostly in (semi)sterile hydroponic systems (Song *et al.*, 2012; Vranova *et al.*, 2013; da Silva Lima *et al.*, 2014). An important justification for the use of such soil-free growth conditions is that they allow for tight maintenance of environmental variables (Bowsher *et al.*, 2016; Ziegler *et al.*, 2015). In addition, hydroponic growth systems prevent sorption of metabolites to soil particles and microbial degradation. A recent study made a compelling case for the use of sterile root systems for studying

root exudation chemistry by demonstrating that root exudates collected from non-sterile systems underestimated the quantity and diversity of carbon-containing metabolites due to microbial breakdown (Kuijken *et al.*, 2014). Using hydroponically grown roots under sterile conditions, Strehmel *et al.* (2014) reported wide-ranging chemical diversity in root exudates of Arabidopsis, including mostly secondary metabolites such as (deoxy)nucleosides, anabolites and catabolites of glucosinolates, derivatives of phytohormones (*e.g.* SA, JA, oxylipins) and phenylpropanoids (*e.g.* coumarins, hydroxynammic acids). Nonetheless, there are disadvantages to hydroponically grown, sterile root systems. Hydroponically cultivated roots often develop root morphologies that differ from those of soil-grown roots, which likely reflects an underlying difference in physiology that may impact exudation chemistry (Sgherri *et al.*, 2010; Tavakkoli *et al.*, 2010). Furthermore, microbial degradation products of root exudates, rather than the root-exuded plant metabolites themselves, might act as potent rhizosphere signals. For instance, benzoxazinoids exuded from cereal roots can be converted into stable 2-aminophenoxazin-3-one, which has strong antimicrobial and allelopathic activities (Atwal *et al.*, 1992; Macías *et al.*, 2005). In addition, it is plausible that certain root exudation products stimulate the production of signalling and/or biocidal compounds by rhizosphere microbes (Cameron *et al.*, 2013). Therefore, ignoring the rhizosphere microbiome by studying sterile root systems limits the identification of novel semiochemicals that can shape microbial communities and their activities in the rhizosphere (Prithiviraj *et al.*, 2007).

To date, various methods have been described to collect root exudates from non-sterile rhizosphere soil. These methods have been used mostly to determine total organic carbon and/or nitrogen content (Yin *et al.*, 2014; Phillips *et al.*, 2008), or

to assay for biological response activity (Khan *et al.*, 2002). Some of these studies revealed biological activities by amino acids, organic acids, and other extractable elements (Oburger *et al.*, 2013; Shi *et al.*, 2011; Haase *et al.*, 2008; Bravin *et al.*, 2010; Chaignon *et al.*, 2009). However, the lack of comprehensive metabolic analyses of non-sterile rhizosphere soil limits our ability to establish relationships between microbial community structure and rhizosphere chemistry. Here, we describe a method for untargeted metabolite profiling from non-sterile rhizosphere soil with high microbial diversity. We have developed methods for extraction of polar and apolar metabolites that do not cause detectable levels of damage to root cells, nor affect viability of soil- and rhizosphere-inhabiting microbes. Using UPLC-Q-TOF mass spectrometry followed by uni- and multivariate statistical analyses, we demonstrate quantitative and qualitative differences in metabolite profiles between soil without plants and soil with plants, and putatively identify the rhizosphere metabolites that are enriched in extracts from Arabidopsis and maize soil. We discuss the potential of this technique for discovering semiochemicals that shape microbial community structure and activity in the rhizosphere.

RESULTS

Development of a plant cultivation system for extraction of rhizosphere chemicals.

We used the model plant species *Arabidopsis thaliana* (Arabidopsis) to develop a plant cultivation system that is suitable for extraction of rhizosphere chemicals. Individual plants were grown for 5 weeks in 30-mL plastic tubes with drainage holes

in the bottom (Figure 1). Since Arabidopsis naturally grows in sandy soils (Lev-Yadun and Berleth, 2009), the tubes contained a homogenous 1:9 (v/v) mixture of fresh M3 compost and sand. Control tubes without plants were included for extraction of chemicals from control soil. All tubes were placed in individual trays, in order to prevent cross contamination of microbes and chemicals (Figure 1). Each tube was watered once per week (5 mL) from the base with a final watering three days before sampling (relative water content after sampling of 88 ± 4.5% per g). This watering regime provided reproducible levels of relative water content at the time of sampling. Under these conditions, flushing the tubes with 5 mL of water or extraction solution (see below) consistently yielded 4 - 4.5 mL collected volume after 1 min of incubation.

Microbial diversity of roots and rhizosphere soil and rhizosphere effect.

Root-derived chemicals mediate the rhizosphere effect (Bakker *et al.*, 2013; Jones *et al.*, 2009). To verify whether plants in our cultivation system showed a rhizosphere effect, we extracted DNA from control soil (without plants) and Arabidopsis roots plus adhering rhizosphere soil. Thus, the 'root plus rhizosphere' samples capture microbial diversity of the rhizosphere, the rhizoplane, and the root cortex. Paired-end 250 bp MiSeq Illumina sequencing of amplified partial 16S rRNA genes was used to profile microbial communities. A total of 2,280,754 raw sequences were obtained with an average of 285,094 per sample. Of these, 1,693,274 reads passed quality controls, chimera removal and singleton removal. Operational Taxonomic Units (OTUs) were generated by clustering at 97% similarity and cross-referenced against the Greengenes 13.8 database (DeSantis *et al.*, 2006), yielding a total of 3,863 OTUs. Rarefaction analysis (Supplemental Figure S1) indicated sufficient

sequencing depth to capture the majority of OTUs. Dominant bacterial taxa at the phylum level were *Actinobacteria* (10.0% across all samples), and *Proteobacteria* (87.8%) comprised mostly of *α*-, *β*- and *γ*-*Proteobacteria* (17.1%, 44.8% and 25.3%, respectively), while at the family level we detected *Burkholderiaceae* (16.6% across all samples), *Oxalobacteraceae* (16.4%), *Pseudomonadaceae* (14.6%) and *Xanthomonadaceae* (10.3%; Figure S2). In addition, we detected ten families of the *Rhizobiales* (9.1%) including *Bradyrhizobiaceae* (3.4%) and *Rhizobiaceae* (1.6%). Many of these phyla and families have previously been reported to be associated with plant roots (Lundberg *et al.*, 2012; Bulgarelli *et al.*, 2015), illustrating that the soil substrate of our cultivation system harbours a microbiome that is typical for microbe-rich soil. To investigate whether the growth system produced a rhizosphere effect by plant roots, we analysed samples for statistically significant differences in OTUs between 'root plus rhizosphere' samples and soil samples. To minimize confounding effects from low-abundance OTUs, data were filtered to include only sequences that appeared *i*) > 5 times across 30% of the samples, and *ii*) 20 times or more across all samples, resulting in a final selection of 662 OTUs. Principal Coordinate Analysis (PCoA) using Unifrac distances revealed a difference in phylogenetic similarity (Figure 2a) between the 'root plus rhizosphere' samples and control soil samples, which was confirmed by PERMANOVA analysis ($F_{1,6,}$ P = 0.023). A total of 178 OTUs were found to differ significantly in relative abundance between 'root plus rhizosphere' and control soil samples, including an increased abundance of 17 *Rhizobiales* OTUs in root samples (*e.g. Rhizobiaceae*, *Methylobacteriaceae*, *Hyphomicrobiaceae*, *Phyllobacteriaceae* and *Bradyrhizobiaceae*; Figure 2b). While the mean Shannon diversity index did not differ between soil and 'root plus rhizosphere' samples (3.58; SD = 0.001 and 3.22; SD = 0.001, respectively;

Student's *t*-test t(3) = 0.92, *P* = 0.39), mean OTU richness of 'root plus rhizosphere' samples (717, SD = 2.1) was significantly lower than that of control soil samples (1177, SD = 2.3; Student's *t*-test t(3) = 3.51, *P* = 0.04; Figure S1), showing an influence of roots on the microbial communities. Hence, the presence of plant roots in our experimental system produces a statistically significant rhizosphere effect.

Selection of extraction solutions that do not cause detectable damage to root and microbial cells.

Plant-derived metabolites range from polar/hydrophilic (*e.g.* organic and amino acids, nucleotides) to apolar/hydrophobic (*e.g.* lipids, phenylpropanoids). Consequently, comprehensive metabolic profiling of rhizosphere soil requires extraction solutions of different polarities. However, the extraction solution should not damage cells from roots or soil microbes, which could contaminate the extract with cellular metabolites (see Figure S3 for a conceptual model). Although water-based solutions without organic solvents are unlikely to cause cellular damage, they are unsuitable for extracting apolar (hydrophobic) metabolites. Conversely, solutions containing organic solvents extract apolar compounds, but risk cell damage by destabilization of membrane lipids (Patra *et al.*, 2006). With a polarity index of 5.1, methanol (MeOH) is capable of extracting polar and apolar metabolites (Figure S3). Accordingly, we selected MeOH as the organic solvent in our extraction solutions.

To test whether exposure to the MeOH-containing extraction solutions has a damaging effect on plant roots, we incubated intact roots of Arabidopsis for 1 min in acidified extraction solutions with different MeOH concentrations (0, 50 and 95% (v/v) MeOH + 0.05 % (v/v) formic acid). As a negative control, tissues were incubated for 1 min in water. To minimize root damage prior to treatment, roots were

collected from agar-grown plants. As a positive control for cell damage, tissues were wounded before incubation. After incubation, tissues were transferred to sterile water for quantification of electrolytes leakage, which is a sensitive method to quantify cell damage in Arabidopsis (Pétriacq *et al.*, 2016a). As shown in Figure 3a, none of the extraction solutions increased the level of electrolytes leakage in comparison to water-incubated roots (Figure 3a). Hence, 1-min exposure to the MeOH-containing solutions does not induce ions leakage from root cells of Arabidopsis. To investigate further potentially damaging effects of the MeOH-containing solutions on root cell integrity, we carried out microscopy studies. Based on the assumption that cell damage by MeOH would permeabilise root cells and cause denaturation of cytoplasmic proteins, we used fluorescence of a C-terminal fusion between the cytoplasmic aspartyl-tRNA synthetase IBI1 and YFP as a marker for root cell integrity (Luna *et al.*, 2014). Roots of two-week-old *35S::IBI1:YFP* plants were carefully removed from MS agar medium, incubated for 1 min in extraction solutions or water (negative control), and analysed for YFP fluorescence (Figure S4). As a positive control for cell damage, *35S::IBI1:YFP* roots were incubated for 15 min in 100% MeOH. YFP fluorescence in roots incubated in acidified 0% MeOH and 50% MeOH solutions was similar to roots incubated for 1 min in water (negative control). Some roots incubated in acidified 95% MeOH showed a weaker YFP signal, although this reduction was less severe than the near complete loss of YFP fluorescence in roots after incubation for 15 min in 100% MeOH (positive control). Thus, 1-min exposures to the 0% and 50% MeOH solutions does not have detectable effects on root cell integrity, which is in line with our conductivity measurements (Figure 3).

To investigate whether the extraction solutions affect soil microbes, control and Arabidopsis soils were drenched for 1 min with the extraction solutions, and microbial viability was tested by dilution plating onto (non-)selective LB agar plates. Viability of culturable soil bacteria was quantified by colony counting on non-selective plates. To test impacts on specific rhizosphere-colonizing bacterial strains, the Gram-negative *Pseudomonas* (*P.*) *simiae* WCS417r (formally known as *P. fluorescens* WCS417r; Berendsen *et al.*, 2012) and the Gram-positive *Bacillus* (*B.*) *subtilis* 168 (Yi *et al.*, 2016) were introduced into separate tubes two days prior to extraction solution treatment, and plated onto selective agar plates after application of extraction solution. Colony forming units (CFU) from solution-treated soils were compared to water-treated soils (1 min; negative control), as well as soils that had been treated for 45 min with 95% MeOH (positive control for microbial cell damage). While the 45-min incubation with 95% MeOH reduced bacterial counts by 10- to 100-fold, none of the acidified MeOH solutions had a statistically significant effect on CFU counts from either soil type in comparison to water-treated soil (Figures 3b and 3c).

In summary, our control experiments for cell damage show that 1-min extraction with the 0% and 50% MeOH solutions does not have detectable impacts on root cell integrity and viability of soil bacteria. However, direct exposure of roots to acidified 95% MeOH solution does have a minor effect on root cell integrity, as evidenced by the faint loss of YFP fluorescence (Figure S4). Accordingly, we cannot exclude the possibility that metabolic profiles obtained with the 95 % MeOH solution are contaminated with cellular metabolites from damaged root cells.

Untargeted metabolic profiling of control and Arabidopsis soil by UPLC-Q-TOF mass spectrometry.

Soil samples were extracted with the three acidified solutions (0.05% formic acid, v/v), containing increasing MeOH concentrations (0, 50 and 95% MeOH). Chemical profiles were obtained by untargeted UPLC-Q-TOF mass spectrometry (MS), using MS^E profiling technology (see Supplemental Methods), which enables simultaneous acquisition of both intact parent ions and fragmented daughter ions (Glauser *et al.,* 2013; Gamir *et al.*, 2014a, 2014b; Planchamp *et al.*, 2014; Pétriacq *et al.*, 2016a, 2016b). Prior to statistical analysis, chemical profiles of ion intensity were aligned and integrated using XCMS software (Smith *et al.*, 2006; Pétriacq *et al.*, 2016a, 2016b). Similarities and differences in ion intensities from both positive (ESI^+, 17,518 cations) and negative ionization modes (ESI^-, 19,488 anions) were first examined by multivariate data analysis, using MetaboAnalyst (v. 3.0) software (Xia *et al.*, 2015). Unsupervised three-dimensional principal component analysis (3D-PCA) separated samples from both soil types that had been extracted with the same solution (Figure 4a), indicating global metabolic differences between control and Arabidopsis soil. These differences were reproducible between three independent experiments (Figure S5). Extractions with the 95% MeOH solution resulted in higher levels of variation than extractions with the 50% and 0% MeOH solutions (Figure 4a and Figure S5). Cluster analysis (Pearson's correlation) revealed complete segregation between control soil samples and Arabidopsis soil samples analysed in positive ionization mode (ESI^+), while samples analysed in negative ionization mode (ESI^-) showed partial segregation between both these soil types. Although samples from the same extraction solution clustered relatively closely within the dendrogram, extracts from the 95% MeOH solution showed more variation than the other solutions (Figure 4b). Finally, we used supervised partial least square discriminant analysis (PLS-DA) to compare metabolite profiles between samples from control soil

and Arabidopsis soil (Figure 4c). Comprehensive analysis of all samples revealed clear separation between all different soil/solution combinations, in both the ESI^+ and ESI^- data. The corresponding PLS-DA models displayed high levels of correlation (R^2 ESI^+ = 0.998; R^2 ESI^- = 0.951) and predictability (Q^2 ESI^+ = 0.619; Q^2 ESI^- = 0.657). Binary comparisons between control and Arabidopsis soil for each extraction solution confirmed these differences, each with high levels of correlation ($R^2 > 0.94$) and predictability ($Q^2 > 0.59$) of the PLS-DA models (Figure S6). However, as was also clear from 3D-PCA and Pearson's correlation analyses, samples extracted with the 95% MeOH solution were more variable than extracts obtained with the 0% and 50% MeOH solutions (Figures 4a-c). The enhanced variation between samples extracted with the 95% MeOH solution is consistent with our finding that direct exposure of roots to 95% MeOH solution causes minor cell damage (Figure S4). Together, our results show consistent differences in polar and apolar metabolite composition between control soil and Arabidopsis soil, indicating a global influence of roots on the chemical composition of the soil in our cultivation system.

Quantitative differences in metabolites between extractions from rhizosphere and control soil.

Quantification of the total number of detected ions (*m/z* values) yielded marginally higher numbers from samples of control soil compared to that of Arabidopsis soil (Figure S7a). A substantial fraction could be detected in both soil types (66.9%, 64.1% and 49.4% for the 0%, 50% and 95% MeOH solutions, respectively; Figure S7a), indicating a large number of metabolites that were present in both rhizosphere and control soil. Ions that were uniquely present in one or more sample from Arabidopsis soil were most abundant in extractions with the 95% MeOH solution

(6,448), followed by the 50% MeOH solution (4,362) and the 0% MeOH solution (3,991; Figure S7a). To select for ions that were statistically over- or under-represented in Arabidopsis soil, we constructed volcano plots that expressed statistical significance of each ion (*m/z* value) against fold-change between both soil types (Figure 5a). Using a statistical threshold of $P < 0.01$ (Welch's *t*-test) and a cut-off value of > 2 fold-change (Log2 > 1), numbers of ions enriched in control soil were generally higher than those enriched in Arabidopsis soil (Figure 5a). Furthermore, there was relatively little overlap in differentially abundant ions between extraction solutions ($P < 0.01$, Welch's *t*-test, Figures 5b and S7). This pattern was equally clear for ions that were specifically enriched in either soil type ($P < 0.01$, Welch's *t*-test, >2 fold-change; Figures 5b and S7b, middle and right), illustrating the fact that the acidified solutions extracted different classes of metabolites. The 50% MeOH solution yielded the highest number of rhizosphere-enriched ions (178), followed by the 0% MeOH solution (115) and 95% MeOH solution (81). Since the 50% MeOH solution also yielded relatively low levels of variability between replicate samples (Figure 4 and Figure S5), our results suggest that this solution is most suitable for extraction of rhizosphere-enriched metabolites.

Composition of rhizosphere- and control soil-enriched metabolites.

To study which metabolite classes drive the global differences between rhizosphere and control soil (Figure 4 and Figure 5), we pooled the top 20-ranking ions from each volcano plot which were ranked by fold-change and statistically significant difference between control and Arabidopsis soil, resulting in a total of 120 metabolic markers for each soil type. To enhance statistical stringency, ions were subsequently filtered by statistical significance between all soil/solution combinations (ANOVA; $P < 0.01$),

using a Benjamini-Hochberg correction for false-discovery rate (FDR). The final selection yielded a total of 76 rhizosphere-enriched ions and 75 control soil-enriched ions. MarVis software (Kaever *et al.*, 2012) was used to correct for adducts and/or C isotopes (tolerance: *m/z* = 0.1 Da and RT = 10 s), after which the predicted masses were used for putative identification (Table S1), using METLIN, PubChem, MassBank, Lipid Bank, ChemSpider, Kegg, AraCyc and MetaCyc databases (Kaever *et al.*, 2009; Gamir *et al.*, 2014a, 2014b; Pastor *et al.*, 2014; Kaever *et al.*, 2012; Pétriacq *et al.*, 2016a, 2016b). To obtain a global profile of soil- and rhizosphere-enriched chemistry, putative compounds were assigned to different metabolite classes (Figure 6). Putative chemicals that unlikely accumulate as natural products in (rhizosphere) soil, such as synthetic drugs or mammalian hormones, were excluded from these profiles (Table S1). In comparison to control soil, Arabidopsis soil was enriched with ions that putatively annotate to flavonoids (8 *vs* 2%), lipids (33 *vs* 6%) and alkaloids (5% in Arabidopsis soil only; Figures 6 and S8; Table S1), which supports the notion that rhizosphere soil is enriched with plant-derived metabolites. The global composition of control soil showed a higher fraction of metabolites that could not be annotated (Figures 6 and S8; Table S1), likely due to an under-representation of soil metabolites in publically available databases.

Applicability of the method to maize in agricultural soil.

Having established that our method is suitable for detecting rhizosphere-enriched metabolites from Arabidopsis, we investigated whether the method could be applied to profile rhizosphere metabolites from a crop species (maize; *Zea mays*) in agricultural soil. To this end, the cultivation system was up-scaled to 50-mL tubes that were filled with a mixture of agricultural soil from arable farmland (Spen farm,

Leeds, UK) and perlite (75:25, v/v). The perlite was added to improve drainage of the soil, which improved plant growth and ensured that sufficient solution was collected from the base of the tubes within 1 min of extraction solution application. Maize plants were grown for 17 days, and rhizosphere chemistry was extracted using the 50% MeOH solution (+ formic acid 0.05%, v/v). Further validation experiments showed that 1-min exposure of maize roots to this solution did not lead to increased electrolytes leakage (Figure 7a). Comparative analysis of metabolites by UPLC-Q-TOF identified a total of 6,071 cations (ESI^+) and 9,006 anions (ESI^-). 3D-PCA showed complete separation between samples from control (red) and maize (green) soil (Figure 7b). Quantitative differences were determined by volcano plots (Welch's *t*-test, $P < 0.01$: fold-change > 2), revealing 287 cations (ESI^+) and 197 anions (ESI^-) that were statistically enriched in maize soil (Figure 7c). Cross-referencing the 100 most significant ions (top 50 anions + top 50 cations) against public databases indicated higher levels of chemical diversity in maize soil samples compared to control soil samples. Most metabolic markers could be putatively identified (Table S2) and annotated to different metabolite classes (Figure 7d). As described for the profiling of the Arabidopsis rhizosphere (Figure 6), these final profiles did not include putative compounds that unlikely accumulate in (rhizosphere) soil, such as synthetic drugs (Table S2). Strikingly, a relatively large fraction of maize rhizosphere-enriched ions could be annotated to flavonoids (28%) and benzoxazinoids (21%), which mediate below-ground interactions (Neal and Ton, 2013; Neal *et al.*, 2012; Robert *et al.*, 2012) For instance, HBOA, DIBOA and HMBOA, displayed strong rhizosphere enrichment in maize soil samples (Figure S8), and are known to be produced by maize roots (Marti *et al.*, 2013). Thus, our profiling method is sufficiently robust and

sensitive to profile plant-derived rhizosphere chemicals from a crop species in agricultural soil.

Profiling chemistry in distal rhizosphere fractions.

The rhizosphere was defined by Lorenz Hiltner in 1904 as '*the soil compartment influenced by the root*' (Smalla *et al.*, 2006). However, many rhizosphere studies focus exclusively on soil that is closely associated with plant roots (after removal of loosely associated soil), which may not encompass the total rhizosphere as more distal and loosely associated soil could still be influenced by root-derived chemistry. To investigate whether our profiling method detects chemical influences beyond soil that is closely associated with roots, we used an alternative growth system that separated roots from distal soil (Figure S9). Maize plants were grown in small, fine mesh bags within larger 150-mL tubes containing soil (see Supplemental Methods), which prevented outward root growth, yet allowed for passage of root-derived chemicals and microbes into the distal soil. Similar plant-free tubes were constructed as no plant controls. After 24 days of growth, mesh bags were carefully removed, after which metabolites were extracted from the remaining distal soil that surrounded the mesh bags, using the 50% MeOH extraction solution. As a control for whole soil fractions, metabolites from empty and maize-containing tubes were extracted before removing the mesh bag from the tube, as described earlier. Thus, the experimental design allowed comparison between four soil fractions: *1*) distal soil surrounding mesh bags without roots, *2*) distal soil surrounding mesh bags with maize roots, *3*) whole soil from tubes with mesh bags without roots and *4*) whole soil from tubes with mesh bags with maize roots. Extracts were analysed by UPLC-Q-TOF in ESI^- (26,011 anions) and subjected to unsupervised PCA (Figure S9b). Comparison of

whole soil fractions confirmed a clear separation between plant-free and maize soil samples, illustrating the chemical rhizosphere effect of maize. Although less pronounced than the whole soil fractions, PCA of the distal soil fractions still revealed separate clustering between plant-free and maize soil (Figure S9b), indicating that the chemical influence of the rhizosphere extended beyond soil closely associated with roots. To verify this distant rhizosphere effect, we quantified levels of DIMBOA, which acts as a relatively stable rhizosphere semiochemical influencing behaviour of both rhizobacteria and arthropods (Neal *et al.*, 2012; Robert *et al.*, 2012). In comparison to both plant-free soil fractions, statistically higher quantities of DIMBOA were detected in both whole maize soil and distal maize soil (Figure S9c). Hence, DIMBOA acts as a mobile long-range rhizosphere signal that extends beyond soil that is closely associated with roots. Considering that maize roots contain high quantities of DIMBOA (Robert *et al.*, 2012), and that the distal soil was separated from the roots prior to chemical extraction with the 50% MeOH solution, this result also confirms that the 50% MeOH extraction solution does not have a damaging effect on maize roots, as exemplified by similar DIMBOA levels in whole maize soil and distal maize soil (Figure S9c).

DISCUSSION

Rhizosphere chemistry is a complex mixture of root exudation chemicals, their microbial breakdown products, and microbial breakdown products of soil-specific chemicals. While it is known that microbial diversity in the rhizosphere can influence plant growth and health (Berendsen *et al.*, 2012), the chemical signals mediating

these interactions remain poorly understood. The majority of root exudation studies are based on hydroponic and/or sterile growth systems (Kuijken *et al.*, 2014; Khorassani *et al.*, 2011; Bowsher *et al.*, 2016). Although sterile growth systems are appropriate for exact quantification of root-exuded plant chemicals (Kuijken *et al.*, 2014), these systems do not consider the importance of rhizosphere signals that are of microbial origin, such as microbial breakdown products of root exudates, or metabolites that are specifically produced by rhizosphere-inhabiting microbes. Consequently, linking rhizosphere chemistry to microbial communities and/or activities remains problematic when the biochemical diversity of the non-sterile rhizosphere is not considered (Oburger and Schmidt, 2016). Furthermore, although root exudation studies are increasingly relying on sensitive analytical methods (van Dam and Bouwmeester, 2016; Khorassani *et al.*, 2011; Ziegler *et al.*, 2015), the majority of these studies employs targeted analyses of specific compounds (*e.g.* organic and amino acids, coumarins), which do not address the biochemical diversity of rhizosphere soil. Recent advances in liquid chromatography, mass spectrometry, and uni- and multivariate data analysis have made it possible to conduct untargeted metabolic profiling of complex metabolite mixtures, such as root exudates and soil extracts (van Dam and Bouwmeester, 2016; Khorassani *et al.*, 2011; Swenson *et al.*, 2015; Ziegler *et al.*, 2015; Strehmel *et al.*, 2014). In this study, we employed untargeted UPLC-Q-TOF analysis of soil extracts, followed by uni- and multivariate data reduction to separate rhizosphere-specific chemistry from common soil chemistry. We show that this method is suitable to profile *in situ* rhizosphere chemistry from different plant species and soil types.

The microbial rhizosphere effect is driven by root exudation chemistry (Jones *et al.*, 2009). Accordingly, we verified whether our cultivation system supported the

generation of a difference in microbial communities between control soil samples (without plant roots) and root samples plus adhering rhizosphere soil, using 16S rRNA gene sequencing. This analysis identified a total number of 3,863 OTUs, which by rarefaction analysis appeared to be sufficient to cover the majority of dominant OTUs (Figure S1). Many of the taxa detected in our samples (*e.g.* *Oxalobacteraceae*, *Pseudomonadaceae*, *Xanthomonadaceae* and the *Rhizobiaceae*) are commonly associated with soil and/or plant roots (Lundberg *et al.*, 2012). Comparative analysis identified a range of OTUs with differential relative abundance between control soil and 'root plus rhizosphere' samples (Figure 2), which provided evidence for a rhizosphere effect in our experimental growth system. Many of the corresponding taxa have been linked to rhizosphere effects, such as an enhanced relative abundance of *Oxalobacteraceae* (Figure 2b, Figure S2; Lundberg *et al.*, 2012; Bulgarelli *et al.*, 2015), as well as the *Rhizobiales*, which are commonly associated with plant roots (Hao *et al.*, 2016).

Our cultivation system was designed for *in situ* extraction of chemicals from biologically complex non-sterile rhizosphere soils. The soil matrix for the Arabidopsis experiments consisted of 9:1 (v/v) mixture of sand and compost, which is comparable to the sandy soil types of naturally occurring Arabidopsis accessions (Lev-Yadun and Berleth, 2009). This matrix also allowed relatively short collection times of the extracts (1 min), which was sufficient to recover 90% of the volume applied and prevent root damage due to extended exposure to MeOH in the extraction solution. The soil matrix for the maize experiments contained agricultural soil from an arable farm field, which was supplemented with 25% (v/v) autoclaved perlite to prevent compaction and allowed sufficient elution of metabolites over the 1-min extraction period. Using this system, we detected quantitative and qualitative

differences in chemistry between extracts from control and maize soil (Figure 7), demonstrating that the method was applicable for profiling of rhizosphere chemistry from a crop species in agricultural soil.

A major challenge for *in situ* profiling of rhizosphere chemistry is to prevent damage of root cells and microbes during the extraction procedure that could otherwise contaminate the extract with metabolites that are not exuded from intact roots. While water-based extraction solutions are unlikely to cause cellular damage, they are less suitable for extraction of apolar metabolites. Conversely, solutions containing organic solvents extract apolar metabolites, but can damage cell membranes. Due to limited understanding of root exudation chemistry in natural soil types, it remains difficult to distinguish between naturally exuded metabolites and metabolites leaking from damaged root tissues or lysed microbial cells. Therefore, we carried out a range of experiments to investigate whether the MeOH-containing extraction solutions caused cell damage: (*i*) quantification of root electrolytes leakage (Figure 3a), (*ii*) epi-fluorescence microscopy to assess root cell integrity (Figure S4), (*iii*) dilution plating to assess viability of soil- and rhizosphere-colonising bacteria after incubation of the soil in extraction solutions (Figures 3b and 3c), and *iv*) detection of plant-derived chemicals in root-free soil fractions (Figure S9). Firstly, exposure of both Arabidopsis and maize roots to the MeOH-containing solutions did not increase electrolytes leakage for the duration of the extraction procedure (1 min; Figures 3 and 7). Secondly, microscopic analysis of root cells from YFP-expressing Arabidopsis roots did not reveal loss of cell integrity after 1-min exposure to 0% and 50% MeOH-containing solutions (Figure S4). However, this assay did reveal a weak impact by the 95% MeOH solution, indicating that extraction of rhizosphere chemistry with this solution could affect root cell integrity. Thirdly, extraction of

control and Arabidopsis soil with the MeOH-containing extraction solutions did not reduce viability of culturable soil microbes, nor did it affect viability of the Gram-negative rhizobacterial strain *P. simiae* WCS417r and the Gram-positive rhizobacterial strain *B. subtilis* 168 (Figures 3b and 3c). Finally, using the 50% MeOH extraction solution and a compartmentalised growth system that separated maize roots from peripheral rhizosphere soil, we showed that extraction of the peripheral soil after removal of maize roots yielded similar DIMBOA quantities as extraction of soil containing maize roots (Figure S9c). Since maize roots accumulate high quantities of DIMBOA (Robert *et al.*, 2012), this result further confirms that the 50% MeOH extraction solution does not damage maize roots in the soil. Accordingly, we conclude that 1-min exposure to the 0% or 50% MeOH extraction solution does not cause detectable levels of cell damage to roots and soil microbes that could contaminate the chemical profiles from the soils with intracellular metabolites.

Multivariate data analysis and clustering revealed that the variability between replicate extractions was lower for the 0% and 50% MeOH extraction solutions compared to the 95% MeOH solution (Figure 4). This is consistent with our finding that direct exposure to this solution sometimes reduced YFP fluorescence in transgenic Arabidopsis roots (Figure S4). Data projection in volcano plots showed that extraction with the 50% MeOH solution yielded the highest number of rhizosphere-enriched ions in comparison to other extraction solutions (Figure 5a). Hence, the 50% MeOH extraction solution performs best in terms of variability between extractions and total numbers of differentially detected ions. Quantitative analysis of MS profiles revealed slightly lower numbers of rhizosphere-enriched ions than control soil-enriched ions, which was apparent for both Arabidopsis (Figure 5 and Figure S7) and maize (Figure 7). It is possible that this difference is due to the

rhizosphere effect, which reduces bacterial richness (Figure S1 and Figure 2), resulting in lower biochemical diversity in the rhizosphere (Prithiviraj *et al.*, 2007).

The sets of ions enriched in control and plant-containing soil differed substantially in composition (Figures 6 and 7). Interestingly, the number of ions annotated to putative metabolites from publicly available databases was higher for rhizosphere selection (Tables S1 and S2). We attribute this difference to the fact that plant-containing soil is enriched with plant-derived metabolites, which are better represented in publicly available databases than soil-specific metabolites (Strehmel *et al.*, 2014; Swenson *et al.*, 2015). Indeed, the selection of putative rhizosphere metabolites from Arabidopsis contained a relatively high fraction of flavonoids, lipids, and other amino acid-derived secondary metabolites, such as alkaloids and phenylpropanoids (Figure 6, Table S1), whereas the set of putative rhizosphere metabolites from maize included relatively large fractions of flavonoids and benzoxazinoids (Figure 7, Table S2). It should be noted, however, that the analytical method used in this study is limited by the putative identification of single ions. Unless the identity of a single metabolite is confirmed by subsequent targeted analyses, such as specific chromatographic retention time, fragmentation or NMR patterns, its annotation remains putative (*i.e.* inconclusive). However, the novelty of our method does not come from the applied mass spectrometry detection method, but the combined use of the experimental design, extraction methods, mass spectrometry profiling and statistical techniques to deconstruct rhizosphere chemistry. Once a wider profile of rhizosphere chemistry has been established, targeted techniques can be used to confirm metabolite identities. Furthermore, where multiple putative metabolites annotate to the same metabolite class, a more reliable conclusion can be drawn about the involvement of this metabolite class. In

our case, multiple rhizosphere ions could be annotated to the same plant-metabolic pathways, suggesting that the overall rhizosphere profile is influenced by these plant metabolite classes. In support of this, previous studies have reported the presence of the same secondary compounds in plant root exudates (Hassan and Mathesius, 2012; Oburger *et al.*, 2013; Szoboszlay *et al.*, 2016; Oburger and Schmidt, 2016). Moreover, benzoxazinoids, such as DIMBOA, have previously been implicated to act as belowground semiochemicals during maize-biotic interactions (Neal *et al.*, 2012; Robert *et al.*, 2012; Marti *et al.*, 2013). Hence, our method provides a new tool to explore rhizosphere semiochemicals for different plant species and soils.

Relatively few rhizosphere-enriched ions could be annotated to primary plant metabolites, such as proteinogenic amino acids or organic acids (Figures 6 and 7). Although these compounds are exuded in high quantities by roots (Rudrappa *et al.*, 2008; van Dam and Bouwmeester, 2016; Ziegler *et al.*, 2015), the microbial activity in the rhizosphere will quickly metabolize them, and the C18-UPLC separation is not optimal for separation of (often very polar) primary metabolites. Above all, we stress that our method is not suitable for quantitative analysis of primary and secondary root exudates, for which sterile root cultivation systems are more appropriate (Kuijken *et al.*, 2014; Strehmel *et al.*, 2014). Our method should only be used for profiling, identification and/or quantification of rhizosphere chemicals. These compounds can be microbial breakdown products of secondary metabolites in root exudates, but could equally well be synthesised *de novo* by rhizosphere-specific bacterial and fungal microbes. Using the experimental pipeline detailed in this paper, stable isotope labelling of plant root exudates via leaf exposure to $^{13}CO_2$ can potentially differentiate between these classes of rhizosphere metabolites, where plant-derived breakdown products will likely retain higher levels of ^{13}C than *de novo*

synthesised microbial products. Furthermore, as is illustrated by our study, the method allows for simultaneous assessment of rhizosphere chemistry and microbial composition, which can be used for genetic strategies that aim to establish a causal relationship between plant genotype, rhizosphere chemistry and microbial composition (Oburger and Schmidt, 2016). Such an approach would also advance studies on the effects of above-ground stimuli (such as light, atmospheric CO_2 and above-ground (a)biotic stresses) on below-ground plant-microbe interactions.

In summary, our study presents a straightforward method to obtain profiles of rhizosphere chemistry in non-sterile rhizosphere soil. The method is applicable to both model systems and soil-grown crops in agricultural soil. Considering that the microbial interactions in the rhizosphere can have both beneficial and detrimental effects on plant performance (Berendsen *et al.*, 2012; Cameron *et al.*, 2013), our method entails a powerful tool to advance rhizosphere biology and to decipher the chemistry driving plant-microbe interaction in complex non-sterile soils.

EXPERIMENTAL PROCEDURES

Chemicals and reagents.

All chemicals and solvents used for metabolomics were of mass spectrometry grade (Sigma-Aldrich, Germany). Other solvents were of analytical grade.

Experimental set-up of growth system.

Collection tubes for the Arabidopsis experiments were constructed by melting 7 mm holes in the base of 30-mL plastic tubes (Sterilin 128A, ThermoScientific, UK), using a soldering iron (see Figure 1). The drainage hole was covered with 4-cm^2 pieces of Millipore miracloth (pore size of 22-25 µm, https://uk.vwr.com) to avoid loss of soil and to prevent outgrowth by roots. Tubes were filled with ~ 45 g soil matrix, consisting of a homogenous 9:1 (v/v) mixture of sand (silica CH52) and dry compost (Levington M3), which is comparable to sandy soil types of naturally occurring *Arabidopsis thaliana* (Arabidopsis) accessions (Lev-Yadun and Berleth, 2009). To prevent cross contamination of rhizosphere microbes and chemicals between samples, each collection tube was placed onto an individual petri-dish (Nunclon™ Delta, 8.8 cm^2 ThermoScientific, UK) (Figure 1). Collection tubes were wrapped in aluminium foil to limit algal growth in the soil matrix. Seeds of Arabidopsis accession Columbia (Col-0) were stratified for two days in the dark in autoclaved water at 4 °C. Three to four seeds were pipetted onto individual tubes and placed into a growth cabinet (Fitotron, SANYO, UK) with the following growth conditions: 8.5/15.5 h light/dark at 21/19 °C with an average of 120 µmol m^{-2} s^{-1} photons at the top of the collection tubes and a relative humidity of 70%. Four days later, seedlings were removed to leave one seedling per pot, which was grown for 5 weeks until sampling.

All pots were watered twice per week by applying 5 mL of autoclaved distilled water to the petri-dishes, using a 5-mL pipette (Starlab, UK). The final watering date was set at three days before sampling, which resulted in consistent soil water contents at the time of sampling. Relative water content (RWC) was determined by the ratio of soil weight (W) minus soil dry weight (DW), divided by water-saturated soil weight (SW) minus soil dry weight:

$$RWC = \frac{W-DW}{SW-DW}$$

The applied watering regime provided reproducible RWC values at the time of sampling (88 ± 4.5%). Although the RWC during the cultivation of plants was frequently lower, the relatively high RWC value at the time of sampling allowed for constant and relatively high recovery volumes (4 - 4.5 mL) from the soil matrix.

Collection tubes for the maize experiments were constructed by melting 7 mm holes in the base of 50-mL plastic tubes. Tubes were fitted with miracloth at the bottom and filled with a water-saturated mixture of agricultural soil:autoclaved perlite (75:25; v/v), in order to allow for sufficient collection volume 1 min after application of extraction solutions (see below). Soil was collected from an arable field (Spen farm; Leeds, UK), air-dried, sieved to a maximum particle size of 4.75 mm, and homogenised using a mixer. Maize seeds (*Zea mays* variety W22) were surface sterilised for 3 h by placing them in petri-dishes in an airtight container with 100 mL of bleach, to which 5 mL of concentrated HCl had been added. Seeds were imbibed overnight in autoclaved, sterile water before placing on petri-dishes containing sterile, damp filter paper in the dark at 23 °C for two days. Germinated seeds were planted in filled collection tubes, 1.5 cm from the soil surface. Collection tubes were wrapped in foil, covered with black plastic beads, and placed in a growth chamber with the following conditions: 12/12 h light/dark at 25/20 °C. The additional maize

experiment to profile distal rhizosphere chemistry is described in the Supplemental Methods.

Profiling of root associated microbial communities.

Details about DNA extraction, 16S rRNA gene sequencing and analysis of root-associated prokaryotic operational taxonomic units (OTUs) are presented in the Supplemental Methods.

Metabolite extraction from control and Arabidopsis/maize soil.

Plant soil samples were collected from tubes containing one 5-week-old Arabidopsis plant, or one 17-day-old maize plant. Plant soil chemistry was analysed from five replicated samples, whereas control soil chemistry was analysed from three replicated samples. All samples were all collected at the same time. For the Arabidopsis system, cold extraction solution (5 mL) containing 0%, 50% or 95% methanol (v/v) with 0.05% formic acid (v/v) was applied to the top of the tubes. After 1 min, 4 - 4.5 mL were collected from the drainage hole in 5-mL centrifuge tubes (Starlab, UK). For the maize system, 15 mL of the 50% methanol solution (0.05 % formic acid, v/v) was applied and flushed through the soil by applying pressure to the top of the pot, using a modified lid containing a syringe. After 1 min, 10 mL were collected in centrifuge tubes. For both cultivation systems, extracts were centrifuged to pellet soil residues (5 min, 3,500 *g*), after which 4 mL of supernatant were transferred into a new centrifuge tube and flash-frozen in liquid nitrogen, freeze-dried for 48 hours until complete dryness (Modulyo benchtop freeze dryer, Edwards, UK), and stored at - 80 °C. Dried aliquots were re-suspended in 100 µL of methanol: water: formic acid (50: 49.9: 0.1, v/v), sonicated at 4 °C for 20 min, vortexed and

centrifuged (15 min, 14,000 *g*, 4 °C) to remove potential particles that could block the UPLC column. Final supernatants (80 µL) were transferred into glass vials containing a glass insert prior to UPLC-Q-TOF analysis.

Assessment of cell damage by extraction solutions

Impacts of acidified extraction solutions on integrity of root cells were determined by conductivity measurement from electrolytes leakage and epi-fluorescence microscopy of transgenic YFP-expressing roots, as detailed in the Supplemental Methods. Impacts of extraction solutions on culturable soil bacteria and introduced soil- and rhizosphere-colonising bacteria were determined by dilution plating, as described in the Supplemental Methods.

UPLC-Q-TOF analysis of soil chemistry

Details of the UPLC-Q-TOF analysis, including targeted detection of DIMBOA, and uni- and multivariate data analyses to deconstruct rhizosphere chemistry are presented in Supplemental Methods.

ACKNOWLEDGEMENTS

The authors would like to thank Bob Turner for advice and feedback on assessing bacterial viability, and Choong-Min Ryu for providing the *B. subtilis* 168 strain. The research was supported by a consolidator grant from the European Research Council (ERC; no. 309944 "*Prime-A-Plant*"), a Research Leadership Award from the Leverhulme Trust (no. RL-2012-042), and an ERA-CAPS BBSRC grant

(BB/L027925/1, "*BENZEX*") to Jurriaan Ton and Stephen Rolfe. The Plant Production and Protection (P3) centre of the University of Sheffield supported Pierre Pétriacq's work. The authors declare no conflict of interest.

SHORT LEGENDS FOR SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Supplemental Figure S1. Rarefaction curves of 16S rRNA Operational Taxonomic Units (OTUs).

Supplemental Figure S2. Relative abundance of bacterial taxa.

Supplemental Figure S3. Solvent polarity and extraction of rhizosphere chemistry.

Supplemental Figure S4. Epi-fluorescence microscopy analysis of Arabidopsis root cell damage.

Supplemental Figure S5. Reproducibility of metabolite profiles between experiments.

Supplemental Figure S6. Binary PLS-DA analysis of metabolite profiles.

Supplemental Figure S7. Details of quantitative differences in metabolites.

Supplemental Figure S8. Relative quantities of benzoxazinoids.

Supplemental Figure S9. Profiling distal rhizosphere chemistry.

Supplemental Table S1. Putative identification of Arabidopsis metabolic markers.

Supplemental Table S2. Putative identification of maize metabolic markers.

Supplemental Methods.

REFERENCES

Ahmad, S., Veyrat, N., Gordon-Weeks, R., et al. (2011) Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. *Plant Physiol.*, **157**, 317–327.

Atwal, A.S., Teather, R.M., Liss, S.N. and Collins, F.W. (1992) Antimicrobial activity of 2-aminophenoxazin-3-one under anaerobic conditions. *Can J Microbiol*, **38**, 1084–1088.

Badri, D. V. and Vivanco, J.M. (2009) Regulation and function of root exudates. *Plant Cell Environ.*, **32**, 666–681.

Bakker, P.A.H.M., Berendsen, R.L., Doornbos, R.F., Wintermans, P.C.A. and Pieterse, C.M.J. (2013) The rhizosphere revisited: root microbiomics. *Front. Plant Sci.*, **4**, 165.

Berendsen, R.L., Pieterse, C.M.J. and Bakker, P. a H.M. (2012) The rhizosphere microbiome and plant health. *Trends Plant Sci.*, **17**, 478–86.

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.

Bowsher, A.W., Ali, R., Harding, S.A., Tsai, C.-J. and Donovan, L.A. (2016) Evolutionary Divergences in Root Exudate Composition among Ecologically-Contrasting Helianthus Species. *PloS One*, **11**, e0148280.

Boyes, D.C., Zayed, a M., Ascenzi, R., McCaskill, a J., Hoffman, N.E., Davis, K.R. and Görlach, J. (2001) Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell*, **13**, 1499–510.

Bravin, M.N., Michaud, A.M., Larabi, B. and Hinsinger, P. (2010) RHIZOtest: a plant-based biotest to account for rhizosphere processes when assessing copper bioavailability. *Environ. Pollut. Barking Essex 1987*, **158**, 3330–3337.

Bulgarelli, D., Garrido-Oter, R., Münch, P.C., Weiman, A., Dröge, J., Pan, Y., McHardy, A.C. and Schulze-Lefert, P. (2015) Structure and Function of the Bacterial Root Microbiota in Wild and Domesticated Barley. *Cell Host Microbe*, **17**, 392–403.

Cameron, D.D., Neal, A.L., Wees, S.C.M. van and Ton, J. (2013) Mycorrhiza-induced resistance: more than the sum of its parts? *Trends Plant Sci.*, **18**, 539–45.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., et al. (2010a) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, **7**, 335–336.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L. and Knight, R. (2010b) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinforma. Oxf. Engl.*, **26**, 266–267.

Chaignon, V., Quesnoit, M. and Hinsinger, P. (2009) Copper availability and bioavailability are controlled by rhizosphere pH in rape grown in an acidic Cu-contaminated soil. *Environ. Pollut. Barking Essex 1987*, **157**, 3363–3369.

Chelius, M.K. and Triplett, E.W. (2001) The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. *Microb. Ecol.*, **41**, 252–263.

Dam, N.M. van and Bouwmeester, H.J. (2016) Metabolomics in the Rhizosphere: Tapping into Belowground Chemical Communication. *Trends Plant Sci.*, **21**, 256–65.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, **72**, 5069–5072.

Dessaux, Y., Grandclément, C. and Faure, D. (2016) Engineering the Rhizosphere. *Trends Plant Sci.*, **21**, 566–78.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinforma. Oxf. Engl.*, **26**, 2460–2461.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinforma. Oxf. Engl.*, **27**, 2194–2200.

Gamir, J., Cerezo, M. and Flors, V. (2014) The plasticity of priming phenomenon activates not only common metabolomic fingerprint but also specific responses against P. cucumerina. *Plant Signal. Behav.*, **9**, e28916.

Gamir, J., Pastor, V., Kaever, A., Cerezo, M. and Flors, V. (2014) Targeting novel chemical and constitutive primed metabolites against Plectosphaerella cucumerina. *Plant J.*, **78**, 227–240.

Glauser, G., Veyrat, N., Rochat, B., Wolfender, J.-L. and Turlings, T.C.J. (2013) Ultra-high pressure liquid chromatography-mass spectrometry for plant metabolomics: a systematic comparison of high-resolution quadrupole-time-of-flight and single stage Orbitrap mass spectrometers. *J. Chromatogr. A*, **1292**, 151–159.

Haase, S., Rothe, A., Kania, A., Wasaki, J., Römheld, V., Engels, C., Kandeler, E. and Neumann, G. (2008) Responses to iron limitation in Hordeum vulgare L. as affected by the atmospheric CO2 concentration. *J. Environ. Qual.*, **37**, 1254–1262.

Hao, D.C., Song, S.M., Mu, J., Hu, W.L. and Xiao, P.G. (2016) Unearthing microbial diversity of Taxus rhizosphere via MiSeq high-throughput amplicon sequencing and isolate characterization. *Sci. Rep.*, **6**, 22006.

Hassan, S. and Mathesius, U. (2012) The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions. *J. Exp. Bot.*, **63**, 3429–3444.

Hinsinger, P., Plassard, C. and Jaillard, B. (2006) Rhizosphere: A new frontier for soil biogeochemistry. *J. Geochem. Explor.*, **88**, 210–213.

Hochberg, Y. and Benjamini, Y. (1990) More powerful procedures for multiple significance testing. *Stat. Med.*, **9**, 811–818.

Jones, D.L., Nguyen, C. and Finlay, R.D. (2009) Carbon flow in the rhizosphere: carbon trading at the soil–root interface. *Plant Soil*, **321**, 5–33.

Kaever, A., Landesfeind, M., Possienke, M., Feussner, K., Feussner, I. and Meinicke, P. (2012) MarVis-Filter: ranking, filtering, adduct and isotope correction of mass spectrometry data. *J. Biomed. Biotechnol.*, **2012**, 263910.

Kaever, A., Lingner, T., Feussner, K., Göbel, C., Feussner, I. and Meinicke, P. (2009) MarVis: a tool for clustering and visualization of metabolic biomarkers. *BMC Bioinformatics*, **10**, 92.

Khan, Z.R., Hassanali, A., Overholt, W., Khamis, T.M., Hooper, A.M., Pickett, J.A., Wadhams, L.J. and Woodcock, C.M. (2002) Control of witchweed Striga hermonthica by intercropping with Desmodium spp., and the mechanism defined as allelopathic. *J. Chem. Ecol.*, **28**, 1871–1885.

Khorassani, R., Hettwer, U., Ratzinger, A., Steingrobe, B., Karlovsky, P. and Claassen, N. (2011) Citramalic acid and salicylic acid in sugar beet root exudates solubilize soil phosphorus. *BMC Plant Biol.*, **11**, 121.

Kuijken, R.C.P., Snel, J.F.H., Heddes, M.M., Bouwmeester, H.J. and Marcelis, L.F.M. (2014) The importance of a sterile rhizosphere when phenotyping for root exudation. *Plant Soil*, **387**, 131–142.

Lakshmanan, V., Kitto, S.L., Caplan, J.L., Hsueh, Y.-H., Kearns, D.B., Wu, Y.-S. and Bais, H.P. (2012) Microbe-associated molecular patterns-triggered root responses mediate beneficial rhizobacterial recruitment in Arabidopsis. *Plant Physiol.*, **160**, 1642–61.

Lev-Yadun, S. and Berleth, T. (2009) Expanding ecological and evolutionary insights from wild Arabidopsis thaliana accessions. *Plant Signal. Behav.*, **4**, 796–797.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

Luna, E., Hulten, M. van, Zhang, Y., et al. (2014) Plant perception of β-aminobutyric acid is mediated by an aspartyl-tRNA synthetase. *Nat. Chem. Biol.*, **10**, 450–6.

Lundberg, D.S., Lebeis, S.L., Paredes, S.H., et al. (2012) Defining the core Arabidopsis thaliana root microbiome. *Nature*, **488**, 86–90.

Macías, F.A., Marín, D., Oliveros-Bastidas, A., Castellano, D., Simonet, A.M. and Molinillo, J.M.G. (2005) Structure-Activity Relationships (SAR) studies of benzoxazinones, their degradation products and analogues. phytotoxicity on standard target species (STS). *J. Agric. Food Chem.*, **53**, 538–548.

Marti, G., Erb, M., Boccard, J., et al. (2013) Metabolomics reveals herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots. *Plant Cell Environ.*, **36**, 621–639.

McMurdie, P.J. and Holmes, S. (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data M. Watson, ed. *PLoS ONE*, **8**, e61217.

Neal, A.L., Ahmad, S., Gordon-Weeks, R. and Ton, J. (2012) Benzoxazinoids in root exudates of maize attract Pseudomonas putida to the rhizosphere. *PloS One*, **7**, e35498.

Neal, A.L. and Ton, J. (2013) Systemic defense priming by Pseudomonas putida KT2440 in maize depends on benzoxazinoid exudation from the roots. *Plant Signal. Behav.*, **8**, e22655.

Neumann, G., George, T.S. and Plassard, C. (2009) Strategies and methods for studying the rhizosphere—the plant science toolbox. *Plant Soil*, **321**, 431–456.

Oburger, E., Dell'mour, M., Hann, S., Wieshammer, G., Puschenreiter, M. and Wenzel, W.W. (2013) Evaluation of a novel tool for sampling root exudates from soil-grown plants compared to conventional techniques. *Environ. Exp. Bot.*, **87**, 235–247.

Oburger, E. and Schmidt, H. (2016) New Methods To Unravel Rhizosphere Processes. *Trends Plant Sci.*, **21**, 243–255.

Pastor, V., Gamir, J., Camañes, G., Cerezo, M., Sánchez-Bel, P. and Flors, V. (2014) Disruption of the ammonium transporter AMT1.1 alters basal defenses generating resistance against Pseudomonas syringae and Plectosphaerella cucumerina. *Front. Plant Sci.*, **5**, 231.

Patra, M., Salonen, E., Terama, E., Vattulainen, I., Faller, R., Lee, B.W., Holopainen, J. and Karttunen, M. (2006) Under the Influence of Alcohol: The Effect of Ethanol and Methanol on Lipid Bilayers. *Biophys. J.*, **90**, 1121–1135.

Pétriacq, P., Stassen, J.H. and Ton, J. (2016a) Spore density determines infection strategy by the plant-pathogenic fungus Plectosphaerella cucumerina. *Plant Physiol.*, **170**, 2325–2339.

Pétriacq, P., Ton, J., Patrit, O., Tcherkez, G. and Gakière, B. (2016b) NAD acts as an integral regulator of multiple defense layers. *Plant Physiol.*, **172**, 1465–1479.

Phillips, R.P., Erlitz, Y., Bier, R. and Bernhardt, E.S. (2008) New approach for capturing soluble root exudates in forest soils. *Funct. Ecol.*, **22**, 990–999.

Planchamp, C., Glauser, G. and Mauch-Mani, B. (2014) Root inoculation with Pseudomonas putida KT2440 induces transcriptional and metabolic changes and systemic resistance in maize plants. *Front. Plant Sci.*, **5**, 719.

Prithiviraj, B., Paschke, M.W. and Vivanco, J.M. (2007) Root communication: The role of root exudates. *Encycl. Plant Crop Sci.*

Robert, C.A.M., Veyrat, N., Glauser, G., et al. (2012) A specialist root herbivore exploits defensive metabolites to locate nutritious tissues: A root herbivore exploits plant defences. *Ecol. Lett.*, **15**, 55–64.

Rudrappa, T., Czymmek, K.J., Paré, P.W. and Bais, H.P. (2008) Root-secreted malic acid recruits beneficial soil bacteria. *Plant Physiol.*, **148**, 1547–56.

Sgherri, C., Cecconami, S., Pinzino, C., Navari-Izzo, F. and Izzo, R. (2010) Levels of antioxidants and nutraceuticals in basil grown in hydroponics and soil. *Food Chem.*, **123**, 416–422.

Shi, S., Richardson, A.E., O'Callaghan, M., DeAngelis, K.M., Jones, E.E., Stewart, A., Firestone, M.K. and Condron, L.M. (2011) Effects of selected root exudate components on soil bacterial communities. *FEMS Microbiol. Ecol.*, **77**, 600–610.

Silva Lima, L. da, Olivares, F.L., Rodrigues de Oliveira, R., Vega, M.R.G., Aguiar, N.O. and Canellas, L.P. (2014) Root exudate profiling of maize seedlings inoculated with Herbaspirillum seropedicae and humic acids. *Chem. Biol. Technol. Agric.*, **1**, 1–18.

Smalla, K., Sessitsch, A. and Hartmann, A. (2006) The Rhizosphere: "soil compartment influenced by the root." *FEMS Microbiol. Ecol.*, **56**, 165–165.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.*, **78**, 779–787.

Song, F., Han, X., Zhu, X. and Herbert, S.J. (2012) Response to water stress of soil enzymes and root exudates from drought and non-drought tolerant corn hybrids at different growth stages. *Can. J. Soil Sci.*, **92**, 501–507.

Strehmel, N., Böttcher, C., Schmidt, S. and Scheel, D. (2014) Profiling of secondary metabolites in root exudates of Arabidopsis thaliana. *Phytochemistry*, **108**, 35–46.

Swenson, T.L., Jenkins, S., Bowen, B.P. and Northen, T.R. (2015) Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biol. Biochem.*, **80**, 189–198.

Szoboszlay, M., White-Monsant, A. and Moe, L.A. (2016) The Effect of Root Exudate 7,4'-Dihydroxyflavone and Naringenin on Soil Bacterial Community Structure. *PloS One*, **11**, e0146555.

Tavakkoli, E., Rengasamy, P. and McDonald, G.K. (2010) The response of barley to salinity stress differs between hydroponic and soil systems. *Funct. Plant Biol.*, **37**, 621–633.

Uren, N. (2007) Types, Amounts, and Possible Functions of Compounds Released into the Rhizosphere by Soil-Grown Plants. In *In The Rhizosphere: Biochemistry and Organic Substances at the Soil- Plant Interface.Eds. R Pinton, Z Varanini and P Nannipieri. pp. 19–40. Marcel Dekker, Inc, New York.* pp. 1–21.

Vranova, V., Rejsek, K., Skene, K.R., Janous, D. and Formanek, P. (2013) Methods of collection of plant root exudates in relation to plant metabolism and purpose: A review. *J. Plant Nutr. Soil Sci.*, **176**, 175–199.

Xia, J., Sinelnikov, I.V., Han, B. and Wishart, D.S. (2015) MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res.*, **43**, 251–257.

Yi, H.-S., Ahn, Y.-R., Song, G.C., Ghim, S.-Y., Lee, S., Lee, G. and Ryu, C.-M. (2016) Impact of a Bacterial Volatile 2,3-Butanediol on Bacillus subtilis Rhizosphere Robustness. *Front. Microbiol.*, **7**, 993.

Yin, H., Wheeler, E. and Phillips, R.P. (2014) Root-induced changes in nutrient cycling in forests depend on exudation rates. *Soil Biol. Biochem.*, **78**, 213–221.

Zamioudis, C., Hanson, J. and Pieterse, M.J. (2014) b -Glucosidase BGLU42 is a MYB72-dependent key regulator of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in Arabidopsis roots. *New Phytol.*, **204**, 368–379.

Zhang, D., Zhang, C., Tang, X., Li, H., Zhang, F., Rengel, Z., Whalley, W.R., Davies, W.J. and Shen, J. (2016) Increased soil phosphorus availability induced by faba bean root exudation stimulates root growth and phosphorus uptake in neighbouring maize. *New Phytol.*, **209**, 1–9.

Ziegler, J., Schmidt, S., Chutia, R., Müller, J., Böttcher, C., Strehmel, N., Scheel, D. and Abel, S. (2015) Non-targeted profiling of semi-polar metabolites in Arabidopsis root exudates uncovers a role for coumarin secretion and lignification during the local response to phosphate limitation. *J. Exp. Bot.*, **67**, 1421–1432.

ACCESSION NUMBERS

The sequences used in this study can be found in the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under accession number PRJEB17782.

FIGURE LEGENDS

Figure 1. Experimental growth system and analytical approach for comprehensive chemical profiling of non-sterile rhizosphere soil.

(**a**) *1.* Collection tubes (30 mL) with bottom holes (7 mm) covered by miracloth were filled with a sand:compost mixture 9:1 (v/v) and wrapped in aluminium foil to prevent excess algal growth. Individual Arabidopsis plants (Col-0) were grown for 5 weeks in tubes. Additional tubes containing control soil without plants were maintained under similar conditions. *2.* After application of 5 mL of extraction solution, metabolite samples were collected for 1 min, centrifuged and freeze-dried. *3.* Concentrated samples were analysed by ultra-high-pressure liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (UPLC-Q-TOF). *4.* Multi- and univariate statistical methods were used to determine qualitative and quantitative differences between extracts from control soil and Arabidopsis soil. Selection of ions by statistical difference and fold-change between soil types enabled putative identification of metabolites that were enriched in non-sterile rhizosphere soil.

(**b**) Photographs of the experimental system. Top: tubes after 4.5 weeks of growth. Bottom: tubes after 3 weeks of growth taped onto petri-dishes to prevent cross contamination of metabolites and microbes.

Figure 2. Rhizosphere effect by Arabidopsis in the cultivation system based on 16S rRNA gene sequencing.

Shown are comparisons of bacterial communities between samples from control soil (without roots) and root samples plus adhering rhizosphere soil.

(**a**) Principal coordinate analysis of OTUs in root + rhizosphere samples (red) and control soil samples (green). Ordinations were performed using weighted Unifrac distances. PERMANOVA analysis showed that the root and control soil samples differed significantly (P = 0.023).

(**b**) OTUs that differ in relative abundance between root + rhizosphere samples and control soil samples. OTUs with positive fold changes are more abundant in the root plus rhizosphere samples than control samples. Results are plotted by family for OTUs that showed a significant difference in abundance as calculated using DESeq2, corrected for false discovery. Only OTUs which have a mean count ≥ 20 are shown for clarity. NA, taxonomy not available.

Figure 3. Effects of methanol (MeOH)-containing extraction solutions on electrolytes leakage from Arabidopsis roots (**a**) and viability of soil microbes (**b**, **c**).

(**a**) Quantification of electrolytes leakage from Arabidopsis roots after incubation for 1 min in acidified extraction solutions containing 0%, 50% or 95% MeOH (v/v) and 0.05% formic acid (v/v). The negative control treatment (-ctrl) refers to intact roots that had not been exposed to any extraction solution. As a positive control treatment for cell damage, wounding was inflicted prior to incubation by cutting roots with a razor blade. Shown are average levels of conductivity (n = 4, ± SEM), relative to the maximum level of conductivity after tissue lysis (set at 100%). Statistically significant differences between treatments were determined by a Welch's F test for ranked data

(*P* values indicated in the upper left corner), followed by Games-Howell post-hoc tests ($P < 0.05$; different letters indicate statistically significant differences).

(**b-c**) Effects of MeOH-containing extraction solutions on viability of soil (**b**) and rhizosphere (**c**) microbes. Shown are average values of colony forming units (CFU) per g of soil for culturable soil bacteria, *Bacillus subtilis* 168 and *Pseudomonas simiae* WCS417r from extraction solution-treated soils ($n = 3$, ± SEM). Asterisks indicate statistically significant differences between negative control (water-flushed soil) and the corresponding treatment *(*$P < 0.05$, Student's *t*-test). In all cases, only positive controls (*i.e.* incubation in 95% MeOH for 45 min) showed statistically significant differences.

Figure 4. Global differences in metabolite profiles between extracts from control soil ('soil') and Arabidopsis soil ('plant').

Shown are multivariate and hierarchical cluster analyses of mass spectrometry data from extracts with different extraction solutions (indicated by % MeOH). Ions (*m/z* values) were obtained by UPLC-Q-TOF analysis in both positive (ESI^+) and negative (ESI^-) ionization mode. Prior to analysis, data were median-normalized, cube-root-transformed and Pareto-scaled.

(**a**) Unsupervised three-dimensional principal component analysis (3D-PCA). Shown in parentheses are the percentages of variation explained by each principal component (PC).

(**b**) Cluster analysis (Pearson's correlation).

(**c**) Supervised partial least square discriminant analysis (PLS-DA). R^2 and Q^2 values indicate correlation and predictability values of PLS-DA models, respectively.

Figure 5. Quantitative differences in metabolite abundance between extracts from control soil and Arabidopsis soil.

(**a**) Volcano plots expressing statistical enrichment of ions (Welch's *t*-test) as a function of fold-difference in control soil (red; 'soil') and Arabidopsis soil (green; 'rhizosphere'). Data shown represent positive (ESI^+) and negative (ESI^-) ions from extractions with different solutions (indicated by % MeOH). Cut-off values were set at $P < 0.01$ (-Log10 = 2) and fold-change > 2 (Log2 = 1).

(**b**) Venn diagrams showing overlap in ions (cations and anions combined) that are significantly different between control and Arabidopsis soil samples (left panel; $P < 0.01$, Welch's *t*-test; without fold-change threshold), enriched in extracts from Arabidopsis soil (middle panel; > 2-fold enrichment to soil at $P < 0.01$, Welch's *t*-test), and enriched in control soil (right panel; < 2-fold enrichment to rhizosphere at $P < 0.01$, Welch's *t*-test).

Figure 6. Composition of putative metabolites enriched in control soil (left) or Arabidopsis soil (right).

Differentially abundant ions were selected from the top 20-ranking ions of each volcano plot (Figure 5a) and filtered for statistical significance between all soil/extraction solution combinations (ANOVA with Benjamini-Hochberg FDR; $P < 0.01$). The resulting 76 rhizosphere-enriched ions and 75 control soil-enriched ions were corrected for adducts and/or C isotopes (tolerance: m/z = 0.1 Da and RT = 10 s), and cross-referenced against publicly available databases for putative identification. A comprehensive table of all rhizosphere- and soil-enriched markers is presented in Supplemental Table S1. Multiple ions putatively annotating to the same metabolite were counted additively towards the metabolite classes in the pie-charts.

Putative metabolites that unlikely accumulate as natural products in (rhizosphere) soil (*e.g.* synthetic drugs, mammalian hormones) were not included in the final selection presented. Miscellaneous: putative metabolites that do not belong to any of the other metabolite classes listed. Unknown: ion markers that could not be assigned to any known compound.

Figure 7. Applicability of the profiling method for maize in agricultural soil.

The experimental system for extracting soil chemistry was based on 50-mL collection tubes filled with a mixture of agricultural soil from arable farmland and perlite (75:25, v/v). Samples were extracted with the 50% MeOH (v/v) solution 17 days after planting.

(**a**) Quantification of maize root damage after direct exposure to the extraction solutions. Five day-old maize roots were incubated for 1 min in acidified extraction solutions containing 0%, 50% or 95% MeOH (v/v) and tested for electrolytes leakage by conductivity. For details, see legend to Figure 3a. Shown are average levels of conductivity (n = 4, ± SEM), relative to the maximum level of conductivity after tissue lysis (set at 100%). Statistically significant differences between treatments were determined by a Welch's F test for ranked data (P values indicated in the upper left corner of each panel), followed by Games-Howell post-hoc tests ($P < 0.05$; different letters indicate statistically significant differences).

(**b**) Unsupervised 3D-PCA, showing global differences in metabolic profiles between control soil (red) and maize soil (green). Shown are data from extracts with the 50% MeOH (v/v) extraction solution. For further details, see legend to Figure 4.

(**c**) Volcano plots expressing statistical enrichment of ions (Welch's t-test) as a function of fold-difference in control soil (red; 'soil') and maize soil (green;

'rhizosphere'). Cut-off values were set at $P < 0.01$ (-Log10 = 2) and fold-change > 2 (Log2 = 1).

(**d**) Relative composition of putative metabolite classes enriched in control soil (left) or maize soil (right). Differentially abundant metabolites were selected from the top 50-ranking ions of each volcano plot (ESI^+ and ESI^-; **c**), corrected for adducts and/or C isotopes (tolerance: *m/z* = 0.1 Da and RT = 10 s), and cross-referenced against publicly available databases for putative identification. A comprehensive table of all rhizosphere- and soil-enriched markers is presented in Supplemental Table S2. Putative metabolites that unlikely accumulate as natural products in (rhizosphere) soil (*e.g.* synthetic drugs, mammalian hormones) were not included in the final selection presented. For further details, see legend to Figure 6.

SUPPORTING INFORMATION

Supplemental Figure S1. Rarefaction curves of detected OTUs.

Shown are curves after removal of singletons for replicate root + rhizosphere samples (green) and control soil samples (red).

Supplemental Figure S2. Relative abundance (%) of selected families in control soil samples ('Soil'; red) and root + rhizosphere samples ('Root'; green) from the Arabidopsis growth system.

Shown are families containing OTUs with relative abundances > 2% in one or more samples. Each bar represents an individual biological replicate. NA, taxonomy not available.

Supplemental Figure S3. Model of expected impacts of solvent polarity on the extraction of soil metabolites.

(**a**) Examples of solvent polarities and their impact on the type of metabolites extracted. Polarity index of water, methanol (MeOH) and hexane are shown within parentheses.

(**b**) Hypothesized impact of solvent polarity on cell damage of plant roots and soil microbes.

Supplemental Figure S4. Epi-fluorescence microscopy analysis of cell damage in Arabidopsis roots after exposure to MeOH-containing extraction solutions.

Transgenic roots producing the cytoplasmic aspartyl-tRNA synthase IBI1 fused to YFP (*35S::IBI1:YFP*; Luna *et al.*, 2014) were incubated for 1 min in water or acidified extraction solutions with increasing MeOH concentration (0, 50 or 95% MeOH, v/v + 0.05% formic acid, v/v). After incubation, roots were then rinsed in sterile water, and analysed for YFP fluorescence. Photographs show representative examples from observations of at least 12 roots for each treatment. As a positive control for cell damage, roots were incubated in 100% MeOH for 15 min. The experiment was performed four time with similar results. Scale bars: 50 µm.

Supplemental Figure S5. Reproducibility of differences in metabolite profiles between control and Arabidopsis soil over three independent experiments.

Shown are unsupervised three-dimensional principal component analyses (3D-PCA) from extracts by the different solutions (indicated by % MeOH). Ions (*m/z* values) were obtained by UPLC-Q-TOF in positive (ESI^+, left panels) and negative (ESI^-,

right panel) ionization modes. Analysis was carried out with MetaboAnalyst (v. 3.0), after median normalization, cube-root transformation and Pareto scaling of data. In parentheses are shown the percentages of variation explained by each principal component.

Supplemental Figure S6. Binary PLS-DA analysis of metabolite profiles from control soil and Arabidopsis soil for different extraction solutions (indicated by % MeOH).

Ions (*m/z* values) were obtained by UPLC-Q-TOF analysis in both positive (ESI^+, left panels) and negative (ESI^-, right panel) ionization mode. Prior to analysis, data were median-normalized, cube-root-transformed and Pareto-scaled. All R^2 (correlation) and Q^2 (predictability) values of PLS-DA models were above 0.94 and 0.59, respectively.

Supplemental Figure S7. Quantitative differences in detected ions (UPLC-Q-TOF) between extracts from control and Arabidopsis soil.

(**a**) Total numbers of ions (top) detected in Arabidopsis soil and control soil after extraction with the different extraction solutions (indicated by % MeOH). Venn diagrams (bottom) show overlap in total ion numbers between extracts for each extraction solution.

(**b**) Venn diagrams showing overlap in cations (ESI^+) and anions (ESI^-) that are statistically different between control and Arabidopsis soil (left panel; $P < 0.01$, Welch's *t*-test), that are enriched in extracts from Arabidopsis soil (middle panel; > 2-fold enrichment to soil at $P < 0.01$, Welch's *t*-test), and that enriched are in extracts from control soil (right panel; < 2-fold enrichment to soil at $P < 0.01$, Welch's *t*-test).

Supplemental Figure S8. Relative quantities of selected benzoxazinoid ions in extracts from maize soil and corresponding control soil.

Selective ions (*m/z*) of HBOA (2-hydroxy-4H-1,4-benzoxazin-3-one), DIBOA (2,4-dihydroxy-1,4-benzoxazin-3-one) and 2-hydroxy-7-methoxy-2H-1,4-benzoxazin-3(4H)-one were detected on the basis of retention time and *m/z* value, using UPLC-Q-TOF (ESI^+, Δppm = 0). Charts indicate means of relative abundances (*n* = 5, ± SEM). Levels of statistical significance are indicated in red above the corresponding bars (Student's *t*-test).

Supplemental Figure S9. Profiling distal rhizosphere chemistry.

(**a**) Experimental growth system to profile chemistry of distal rhizosphere fractions. Maize was grown within nylon mesh bags inside 150-mL tubes, containing agricultural soil from arable farmland and perlite (75:25, v/v). Similar plant-free tubes were constructed as controls. After 24 days of growth, chemicals were extracted with the 50% MeOH solution from either the entire pot (whole soil), or the soil surrounding the root containing mesh bag after its careful removal (distal soil).

(**b**) Binary PCAs showing chemical rhizosphere effects in whole soil fractions (upper panel; short + long distance influence) and distal soil fractions (lower panel; long distance influence), illustrating that the rhizosphere extends beyond soil that is closely associated with roots.

(**c**) Targeted quantification of DIMBOA by UPLC-Q-TOF. Shown are average ion intensities (± SEM; *n* = 6), normalised by soil weight. Letters indicate statistically significant differences between soil types (Student's *t*-test, $P < 0.05$).

Supplemental Table S1. Putative identities of ions enriched in Arabidopsis soil and corresponding control soil.

[1] Percentages indicate relative MeOH contents of the acidified extraction solutions.

[2] *P* values are derived from ANOVA followed by false discovery rate correction (Benjamini-Hochberg).

[3] Retention times (RT) and accurate *m/z* values, detected by UPLC-Q-TOF in negative (-) or positive (+) ion mode.

[4] Predicted parameters were derived from the METLIN database, using accurate *m/z* values.

[5] Putative metabolites and their corresponding pathways were validated by information from the PubMed chemical database.

[6] Putative metabolites that unlikely accumulate in (rhizosphere) soil.

Supplemental Table S2. Putative identities of ions enriched in maize soil and corresponding control soil.

[1] Fold-change between maize rhizosphere samples and control soil samples.

[2] *P* values are derived from Welch's *t*-test.

[3] Retention times (RT) and accurate *m/z* values, detected by UPLC-Q-TOF in negative (-) or positive (+) ion mode.

[4] Predicted parameters were derived from the METLIN database, using accurate *m/z* values.

[5] Putative metabolites and their corresponding pathways were validated by information from the PubMed chemical database.

[6] Putative metabolites that unlikely accumulate in (rhizosphere) soil.

SUPPLEMENTAL METHODS

DNA extraction, 16S rRNA gene sequencing and analysis.

For microbial profiling, eight additional growth tubes were set up, as described in the Experimental Procedures, but were not used for the collection of chemicals. Four of these tubes contained one Arabidopsis plant and four contained only growth substrate. After 5 weeks, plants were sampled by carefully loosening the soil around the edges of the growth tube, pulling up the roots and removing excess soil by shaking. Soil samples were also taken from the tubes without plants, using a sterile spatula and avoiding surface material. DNA was extracted from the resulting samples consisting of either roots covered in their closely adhering soil (root plus rhizosphere samples), or only soil (control soil), using a PowerSoil DNA extraction kit (MoBio Laboratories Inc., Carlsbad, CA, USA) according to the manufacturer's instructions. Partial prokaryotic 16S rRNA genes were amplified from this extract, using primers 799F and 1193R (Chelius and Triplett, 2001; Bodenhausen *et al.*, 2013), which were modified to include the Illumina overhang adapter nucleotide sequences (adapters shown in normal typeface, locus specific primers in bold letter font):

799F:

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**AACMGGATTAGATACCCKG**

1193R:

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**ACGTCATCCCCACCTTCC**.

PCRs were carried out, using 0.4 U of KAPA HiFi HotStart DNA polymerase (Kapa Biosystems Ltd, London, UK) on 2 µL of DNA extract in the presence of 2.5 mM $MgCl_2$, 1.2 mM deoxynucleoside triphosphates (dNTPs), 0.2 µM of each primer, and the manufacturer's reaction buffer in a total reaction volume of 20 µL (PCR conditions: 95 °C for 3 min; 25 cycles at 95 °C for 30 s, 58 °C for 30 s and 72 °C for 30 s; and 72 °C for 5 min). To reduce PCR bias, the PCR was performed in triplicate and amplicons were pooled. A sequencing library was constructed by cleaning up pooled PCR products, using AMPure XP beads (Beckman Coulter (UK) Ltd, High Wycomb, UK), followed by attachment of dual indices and Illumina sequencing adapters, using the Nextera XT Index Kit (Illumina Inc. Essex UK) and following the manufacturer's instructions. The indexed PCR products were cleaned using AMPure XP beads and sequencing was performed using a paired end 2 x 250 bp cycle kit v2 on a MiSeq machine running v2 chemistry (Illumina Inc, at The Genome Analysis Centre, Norwich, UK). Raw sequencing data were deposited in the European Nucleotide Archive (ENA) under accession number PRJEB17782. Sequences were analysed by USEARCH (Edgar, 2010) and Qiime pipelines (Caporaso *et al.*, 2010a). Sequences were filtered using USEARCH, retaining those with a maxEE value of 1 (equivalent to 1 in 1,000 errors) and 251 bp long. Chimeras were detected using UCHIME (Edgar *et al.*, 2011), using both reference based and *de novo* detection methods. After selection of OTUs by USEARCH (97% similarity), the representative sequences were aligned to the Greengenes 13_8 core reference alignment (DeSantis *et al.* 2006) using PyNAST (Caporaso *et al.*, 2010b). All other steps leading to the generation of OTU abundance tables were performed using Qiime. All statistical analyses of community data were performed using the R programming language (R Development Core Team, 2016; https://www.R-project.org/) and with

the packages phyloseq (McMurdie and Holmes, 2013), vegan (https://github.com/vegandevs/vegan) and DESeq2 (Love *et al.*, 2014).

Quantification of plant tissue damage by electrolytes leakage.

Tissue damage by the acidified extraction solutions was quantified by conductivity of cell electrolytes leakage, as described previously (Pétriacq *et al.*, 2016a, 2016b). For Arabidopsis, roots were collected from plants cultivated in half strength Murashige-Skoog, solidified with 0.8% Phytagel (Sigma-Aldrich, UK) and adjusted to pH 5.8. Root replicates consisted of one intact root system from 2-week-old plants, which was removed carefully from the agar medium. For maize, roots were collected from surface sterilised seeds, germinated and grown for five days on wetted filter paper in sealed petri-dishes. Tissues were incubated for 1 min in 10 mL of different acidified extraction solutions, containing 0.05% formic acid (v/v) and 0%, 50% or 95% methanol (v/v). As a negative control, tissues were incubated in double-distilled sterile water. As a positive control for cell damage, tissues were wounded prior to extraction solution incubation by cutting roots into 10 pieces with a razor blade. Directly after incubation, tissues were rinsed in double-distilled sterile water, then transferred into glass bottles containing 5 mL of double-distilled sterile water, and subsequently agitated at room temperature for 2 hours on an orbital shaker (200 rpm). Conductivity was then measured in the balanced solution, using a CMD 500 WPA conductivity meter. Subsequently, all samples were boiled for 30 min and re-measured for conductivity of lysed tissue. Cell damage was expressed as the average level of conductivity, relative to the maximum level of conductivity after tissue lysis (set at 100%). Each treatment was based on 4 replicated samples (n = 4). Data were analysed in IBM SPSS (v. 22), using a Welch's F test for ranked data,

followed by Games-Howell tests to assess individual differences ($P < 0.05$). The experiments were repeated three times with similar results.

Analysis of microscopic root cell damage by extraction solutions.

Transgenic Arabidopsis plants (Col-0) expressing the *35S:IBI1:YFP* construct, encoding the cytoplasmic aspartyl-tRNA synthatase IBI1 with a C-terminal fusion to Yellow Fluorescent Protein (Luna *et al.*, 2014; *35S::IBI1:YFP*), were cultivated for two weeks (8.5/15.5 h light/dark at 21/19 °C, 120 µmol m^{-2} s^{-1} photons, 70% relative humidity) on half strength Murashige-Skoog agar plates, solidified with 0.8% Phytagel (Sigma-Aldrich, UK) and adjusted to pH 5.8. Roots were extracted carefully form agar plates, and incubated for 1 min in the acidified MeOH-containing extraction solutions (0, 50, 95% MeOH with 0.05% formic acid, v/v). As negative and positive controls for cell damage, roots were incubated for 1 min in double-distilled sterile water, or for 15 min in 100% MeOH, respectively. After incubation, roots were rinsed in double-distilled sterile water prior to epi-fluorescence microscopy analysis. Fluorescence was observed using an epi-fluorescence microscope (Olympus BX51, excitation filter BP 470/40 nm, barrier filter BP 525/50 nm). For each treatment, root systems form 12 different plants were analysed and photos were taken of representative samples. The experiment was performed four times with similar results.

Analysis of impacts on soil and rhizosphere bacteria by extraction solutions.

Tubes (30 mL; $n = 3$) containing the sand:compost mixture (9:1 v/v) with or without 5-week-old Arabidopsis were left untreated, or were bacterized by syringe injection with 5 mL of 10 mM $MgSO_4$, containing either YFP-expressing *P. simiae* WCS417r

(Zamioudis *et al.*, 2014), or rifampicin-resistant *B. subtilis* 168 (Yi *et al.*, 2016), to a final density of 10^7 colony CFU g^{-1}. After 48 h, tubes were flushed with extraction solution (as detailed in Experimental procedures). Additional tubes were flushed with double-distilled sterile water (control), or 95% MeOH and left for 45 min (positive control for cell damage). Subsequently, 1 g of either control soil (without roots), or Arabidopsis roots plus adhering rhizosphere soil, was sampled from the tubes, suspended for 5 min into 50 mL of 10 mM $MgSO_4$, and centrifuged (5 min, 3,500 *g*). Pellets were re-suspended in 1 mL of 10 mM $MgSO_4$, and subjected to dilution plating onto Luria Broth (LB) agar medium supplemented with 5 µg mL^{-1} of the anti-fungal cycloheximide. For testing impacts on culturable soil bacteria, LB agar contained no further antibiotics; for testing impacts on *P. simiae* WCS417r and *B. subtilis* 168, plates were supplemented with 5 µg mL^{-1} tetracycline + 25 µg mL^{-1} rifampicin and 50 µg mL^{-1} rifampicin, respectively. Plates were kept for 24 - 48 h at 28 °C. Each biologically replicated sample was plated four times, after which the technical replicates were averaged to minimize confounding effects of heterogeneity in suspended pellets. Experiments were repeated twice with comparable results.

UPLC-Q-TOF mass spectrometry.

Untargeted metabolic profiling by UPLC-Q-TOF mass spectrometry (MS) was performed as described previously (Pétriacq *et al.*, 2016b) using an ACQUITY ultra-high-pressure liquid chromatography (UPLC) system coupled to a SYNAPT G2 Q-TOF mass spectrometer with an electrospray (ESI) ionization source (Waters, UK). The system was controlled by MassLynx v. 4.1 software (Waters). Chromatographic separation of samples was carried out at a flow rate of 0.4 mL min^{-1} using an ACQUITY UPLC BEH C18 column (2.1 × 50 mm, 1.7 µm, Waters) coupled to a C18

VanGuard pre-column (2.1 x 5 mm, 1.7 µm, Waters). The mobile phase consisted of solvent A (0.05 %, formic acid v/v, in water) and solvent B (0.05 % formic acid v/v in acetonitrile) with the following gradient: 0 – 3 min 5 – 35 % B, 3 – 6 min 35 – 100 % B, holding at 100 % B for 2 min, 8 – 10 min, 100 – 5 % B. The column was maintained at 45 °C and the injection volume was 10 µL. Between each condition, a blank was injected with 50% methanol (v/v) to clean the column. Sample runs in negative and positive ionization mode (ESI^- and ESI^+) were separated by two consecutive injections with 50% methanol (v/v) to allow stabilization of the ionization modes. An ACQUITY PDA detector (Waters) was used to monitor the UV trace (range 205 – 400 nm, sampling rate 40 points s^{-1}, resolution 1.2 nm). MS detection of ions was operated in sensitivity mode by SYNAPT G2 (50 - 1200 Da, scan time = 0.2 s) in both ESI^- and ESI^+, using a full MS scan (*i.e.* no collision energy) and applying the MS^E function with a ramp in the transfer cell in elevated energy mode (5 to 45 eV). The following conditions were applied for ESI^- (capillary voltage - 3 kV, sampling cone voltage - 25 V, extraction cone voltage -4.5 V, source temperature 120 °C, desolvation temperature 350 °C, desolvation gas flow 800 L h^{-1}, cone gas flow 60 L h^{-1}), and for ESI^+ (capillary voltage 3.5 kV, sampling cone voltage 25 V, extraction cone voltage 4.5 V, source temperature 120 °C, desolvation temperature 350 °C, desolvation gas flow 800 L h^{-1}, cone gas flow 60 L h^{-1}). Prior to analyses, the Q-TOF was calibrated by infusing a sodium formate solution. Accurate mass detection was ensured by infusing the internal lockmass reference peptide leucine enkephalin during each run.

Statistical analysis of MS data.

Prior to multivariate analyses, the XCMS R package (v. 3.1.3; Smith *et al.*, 2006) was used to align and integrate raw UPLC-Q-TOF peaks, to correct for total ion current (TIC) and median fold-change. All statistical analyses were performed with median-normalized, cube-root-transformed and Pareto-scaled data, using MetaboAnalyst software (v. 3.0, http://www.metaboanalyst.ca; Xia *et al.*, 2015). Three-dimensional principal component analyses (3D-PCA) were based on the first three principal components (PCs) that explain most variation of the dataset. Supervised partial least square discriminant analyses (PLS-DAs) were conducted to quantify discriminative power between soil types and extraction solutions. PLS-DA models were validated by correlation (R^2) and predictability (Q^2) parameters for both ESI^+ and ESI^- modes ($R^2 > 0.94$ and $Q^2 > 0.59$, respectively). Numbers of total ions were obtained from XCMS output datasets. To quantify metabolic differences between rhizosphere and control soil, volcano plots were constructed at a statistically significant threshold of $P < 0.01$ (Welch's *t*-test) and a fold-difference threshold of 2, using MetaboAnalyst (v. 3.0, http://www.metaboanalyst.ca; Xia *et al.*, 2015). To obtain putative identities of a combined set of ions from all three extraction solutions that are either enriched in Arabidopsis soil, or its corresponding control soil, the top-20 ranking ions from each volcano plot were selected by fold-change (above 2 or below -2) and *P* value, followed by an ANOVA ($P < 0.01$) for statistical differences between all soil/extraction solution combinations, using a Benjamini-Hochberg false discovery rate (FDR) correction for multiple hypothesis testing (Hochberg and Benjamini, 1990). To obtain putative identities from the 50% MeOH extraction solution that are either enriched in the maize rhizosphere, or corresponding control soil, the top-50 ranking ions from each volcano plot (ESI^+ and ESI^-) were selected. For both cultivation systems, ions were corrected for adducts

and/or isotopes, using MarVis (v. 1.0; http://marvis.gobics.de; tolerance: *m*/z = 0.1 Da, RT = 10 s; Kaever *et al.*, 2012). Putative metabolites were identified by referencing the final set of detected accurate *m/z* values against publicly available chemical databases using METLIN, PubChem, MassBank, Lipid Bank, ChemSpider, Kegg, AraCyc and MetaCyc database, as documented in several studies (Kaever *et al.*, 2009; Kaever *et al.*, 2012; Gamir *et al.*, 2014a, 2014b; Pastor *et al.*, 2014; Pétriacq *et al.*, 2016a, 2016b). METLIN (https://metlin.scripps.edu) was used to determine accuracy and chemical formulae for the putative compounds. PubChem (https://pubchem.ncbi.nlm.nih.gov/) was used to check the predicted pathway classification. In cases where multiple ions could be annotated to the same putative metabolite (due to different adducts and ionization modes; Tables S1 and S2), they were counted additively to the metabolite class presented in the pie-charts of Figures 6 and 7.

Experimental system for profiling distant rhizosphere fractions.

To investigate whether the chemical influence of the rhizosphere extends beyond soil that is closely associated with roots, maize plants were grown in mesh bags, which allowed for physical separation of root systems from the distal soil in the periphery of the growth tube. Bags were constructed from a nylon mesh (35 μm diameter holes), folded over and heat sealed to produce bags (6 cm x 11 cm, approximate diameter when filled = 3.5 cm). These bags were filled with 85 cm^3 of a mixture of 75:25 (v/v) agricultural soil:perlite, as used previously for maize experiments. Each mesh bag was placed into the centre of the 150-mL plastic tube (11 cm high and 5 cm diameter; Starlab) with a miracloth sheet covering the bottom hole of the tube. Seventy cm^3 of the same soil substrate was used to fill the

peripheral space between the mesh bag and tube wall. A total of 24 pots were set up in this manner. Pre-germinated maize seeds (W22) were planted into the bags of 12 tubes. The other 12 tubes were left unplanted to serve as plant-free controls. All tubes were wrapped in foil and covered with black plastic beads to prevent algal growth. Sixty mL of distilled water was added to each tube to saturate the soil with water before seeds were planted, after which all pots were transferred to a growth cabinet with the following conditions: 16/8 h light/dark with an average light intensity of 140 µmol m^{-2} s^{-1} at the top of the collection tubes, a relative humidity of 60%, and a constant temperature of 20 °C. Soil metabolites were extracted from the different soil fractions after 24 days of growth. To collect metabolites from the distal soil fractions, black beads were removed, and mesh bags were carefully removed from half of the pots (6 tubes with maize and 6 without). The remaining distal soil in the tube (*i.e.* the soil that had been outside the bag) was tapped to the bottom of the 150-mL tubes and extracted by applying 25 mL of acidified 50% (v/v) MeOH to the top of the soil. The solution was flushed through the tube by applying pressure for 1 min through a modified 150-mL tube lid containing a 50-mL syringe, until ~10 mL of solution was collected from the base of the tube into new 50-mL tubes. To collect metabolites from the whole soil fractions, plastic beads were removed from the remaining 12 pots and maize shoots were cut from the 6 that contained plants. Subsequently, 50 mL of acidified 50% MeOH (v/v) was applied to the top of the tube, keeping the mesh bags in place. The solution was flushed through by applying pressure for 1 min using the modified lid, as previously described, resulting in a least 10 mL of collection volume at the base of the tube. All extracts were centrifuged to pellet soil residues (5 min, 3,500 *g*), after which 8 mL of supernatant were transferred into a new 15-mL centrifuge tube and flash-frozen in liquid nitrogen. All

samples were freeze-dried for two days, after which dried material was re-suspended in 500 µL of methanol: water: formic acid (50: 49.9: 0.1, v/v/v), sonicated at 4 °C for 20 min, vortexed, transferred into 2-mL microtubes and centrifuged (15 min, 14,000 *g*, 4 °C). Final supernatants (180 µL) were transferred into glass vials containing a glass insert before injection through the UPLC system. UPLC-Q-TOF analysis was conducted in ESI^- as described above. For DIMBOA targeted quantitation, a purified and NMR-verified standard (Ahmad *et al.*, 2011) was run alongside the samples. Metabolomics data were normalised for soil amount (n = 6), and subsequent analysis performed with MetaboAnalyst (v. 3.0), as described above (*i.e.* median normalisation, cube-root transformation, Pareto scaling).

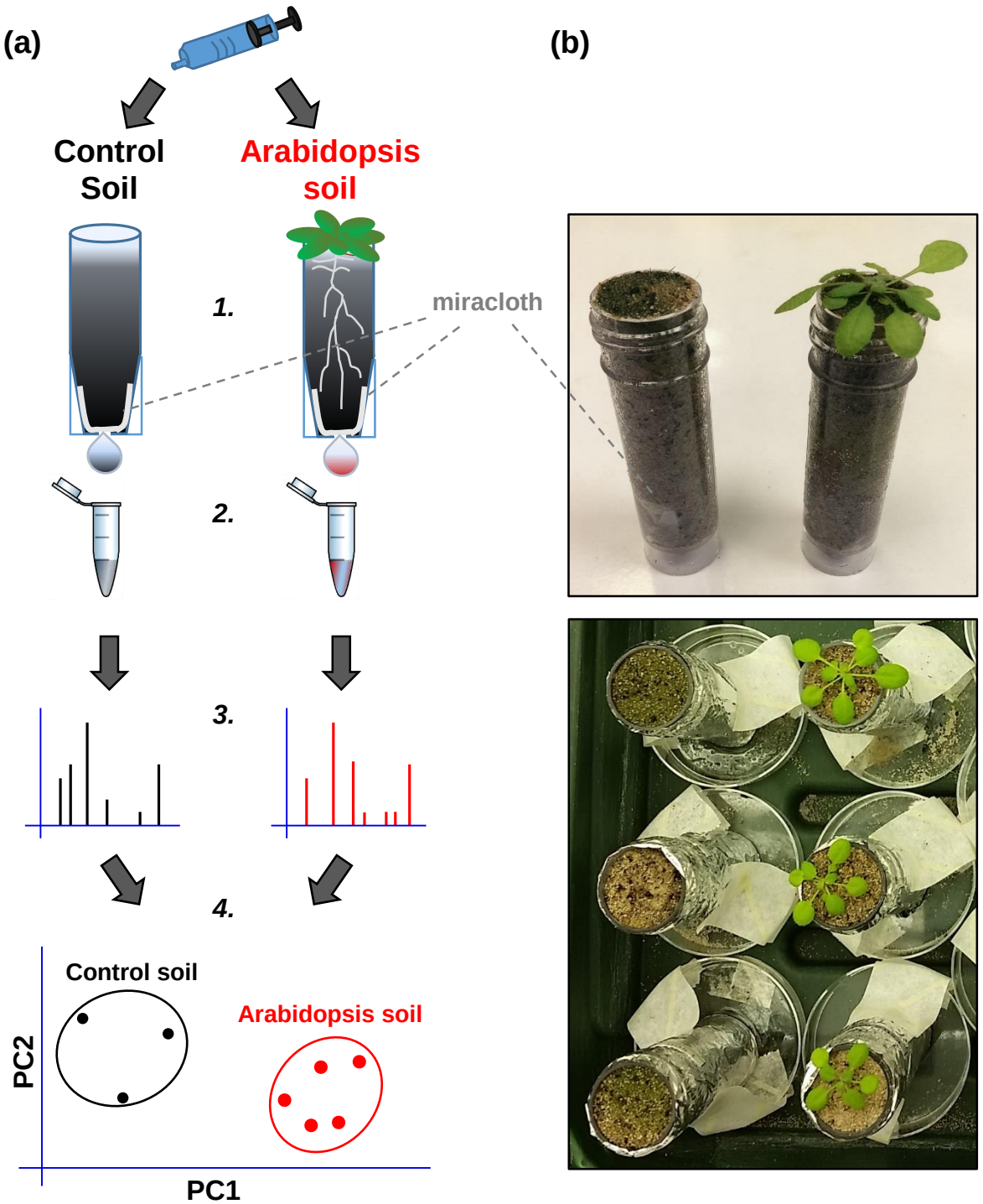


Figure 1. Experimental growth system and analytical approach for comprehensive chemical profiling of non-sterile rhizosphere soil.

(**a**) *1.* Collection tubes (30 mL) with bottom holes (7 mm) covered by miracloth were filled with a sand:compost mixture 9:1 (v/v) and wrapped in aluminium foil to prevent excess algal growth. Individual Arabidopsis plants (Col-0) were grown for 5 weeks in tubes. Additional tubes containing control soil without plants were maintained under similar conditions. *2.* After application of 5 mL of extraction solution, metabolite samples were collected for 1 min, centrifuged and freeze-dried. *3.* Concentrated samples were analysed by ultra-high-pressure liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (UPLC-Q-TOF). *4.* Multi- and univariate statistical methods were used to determine qualitative and quantitative differences between extracts from control soil and Arabidopsis soil. Selection of ions by statistical difference and fold-change between soil types enabled putative identification of metabolites that were enriched in non-sterile rhizosphere soil.
(**b**) Photographs of the experimental system. Top: tubes after 4.5 weeks of growth. Bottom: tubes after 3 weeks of growth taped onto petri-dishes to prevent cross contamination of metabolites and microbes.

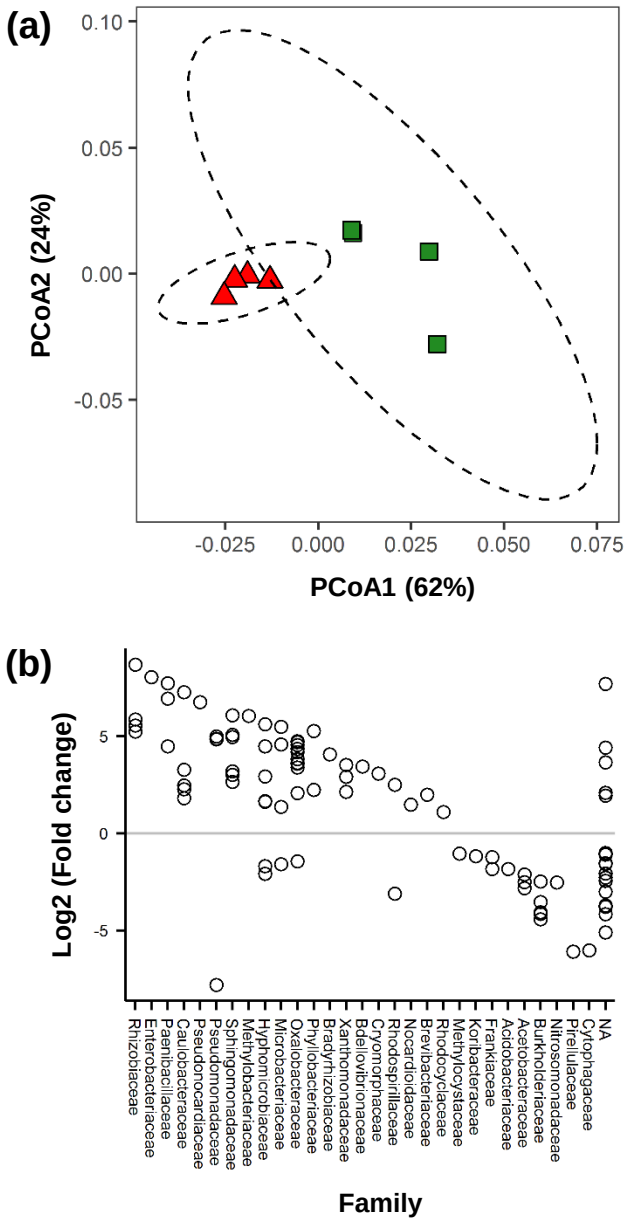


Figure 2. Rhizosphere effect by Arabidopsis in the cultivation system based on 16S rRNA gene sequencing.

Shown are comparisons of bacterial communities between samples from control soil (without roots) and root samples plus adhering rhizosphere soil.
(**a**) Principal coordinate analysis of OTUs in root + rhizosphere samples (red) and control soil samples (green). Ordinations were performed using weighted Unifrac distances. PERMANOVA analysis showed that the root and control soil samples differed significantly (P = 0.023).
(**b**) OTUs that differ in relative abundance between root + rhizosphere samples and control soil samples. OTUs with positive fold changes are more abundant in the root plus rhizosphere samples than control samples. Results are plotted by family for OTUs that showed a significant difference in abundance as calculated using DESeq2, corrected for false discovery. Only OTUs which have a mean count ≥ 20 are shown for clarity. NA, taxonomy not available.

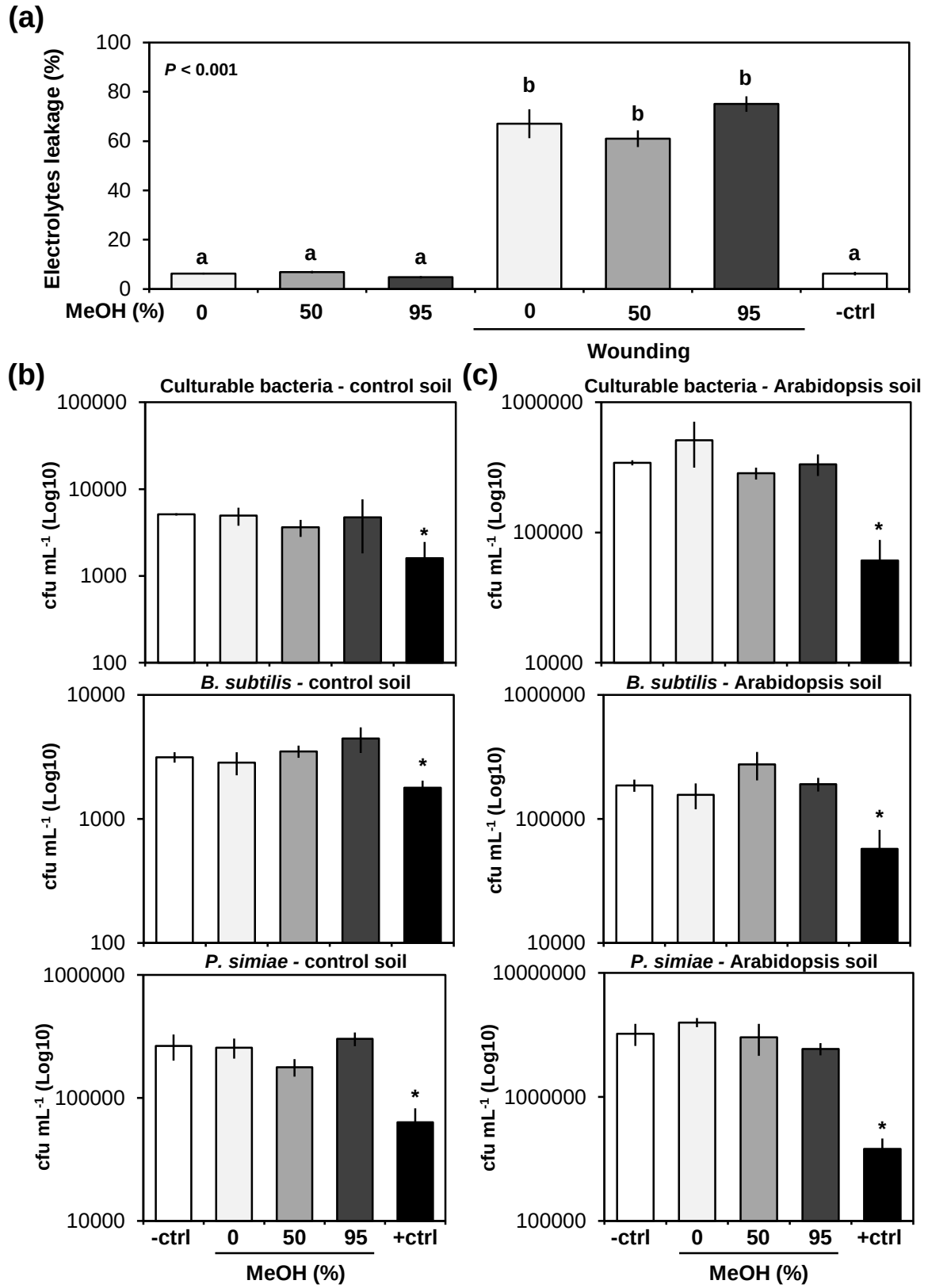


Figure 3. Effects of methanol (MeOH)-containing extraction solutions on electrolytes leakage from Arabidopsis roots (**a**) and viability of soil microbes (**b**, **c**).

(**a**) Quantification of electrolytes leakage from Arabidopsis roots after incubation for 1 min in acidified extraction solutions containing 0%, 50% or 95% MeOH (v/v) and 0.05% formic acid (v/v). The negative control treatment (-ctrl) refers to intact roots that had not been exposed to any extraction solution. As a positive control treatment for cell damage, wounding was inflicted prior to incubation by cutting roots with a razor blade. Shown are average levels of conductivity (n = 4, ± SEM), relative to the maximum level of conductivity after tissue lysis (set at 100%). Statistically significant differences between treatments were determined by a Welch's F test for ranked data (P values indicated in the upper left corner), followed by Games-Howell post-hoc tests (P < 0.05; different letters indicate statistically significant differences).
(**b-c**) Effects of MeOH-containing extraction solutions on viability of soil (**b**) and rhizosphere (**c**) microbes. Shown are average values of colony forming units (CFU) per g of soil for culturable soil bacteria, *Bacillus subtilis* 168 and *Pseudomonas simiae* WCS417r from extraction solution-treated soils (n = 3, ± SEM). Asterisks indicate statistically significant differences between negative control (water-flushed soil) and the corresponding treatment (P < 0.05, Student's t-test); in all cases, only positive controls (*i.e.* incubation in 95% MeOH for 45 min) showed statistically significant differences.

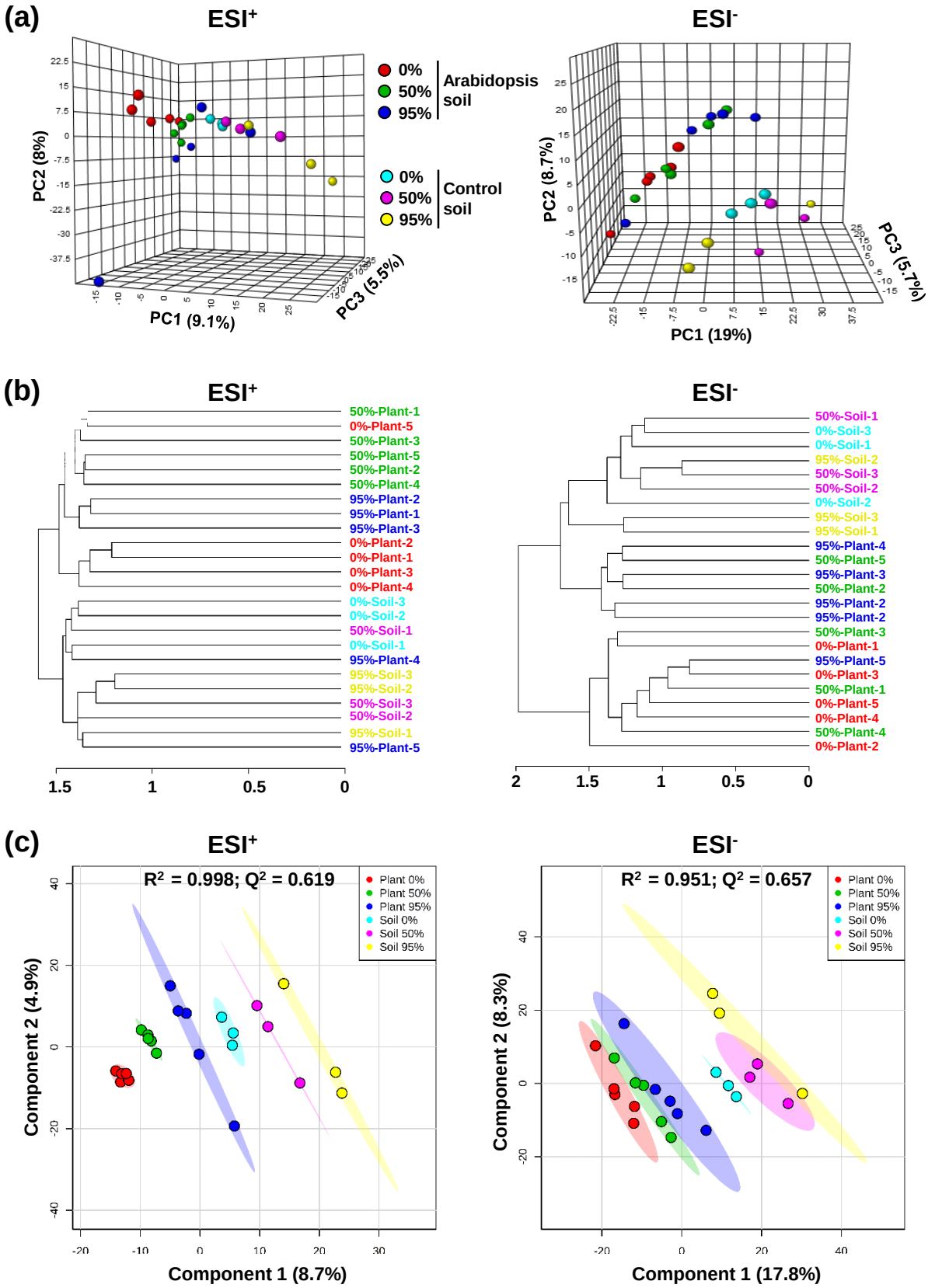


Figure 4. Global differences in metabolite profiles between extracts from control soil ('soil') and Arabidopsis soil ('plant').

Shown are multivariate and hierarchical cluster analyses of mass spectrometry data from extracts with different extraction solutions (indicated by % MeOH). Ions (*m/z* values) were obtained by UPLC-Q-TOF analysis in both positive (ESI^+) and negative (ESI^-) ionization mode. Prior to analysis, data were median-normalized, cube-root-transformed and Pareto-scaled.
(**a**) Unsupervised three-dimensional principal component analysis (3D-PCA). Shown in parentheses are the percentages of variation explained by each principal component (PC).
(**b**) Cluster analysis (Pearson's correlation).
(**c**) Supervised partial least square discriminant analysis (PLS-DA). R^2 and Q^2 values indicate correlation and predictability values of PLS-DA models, respectively.

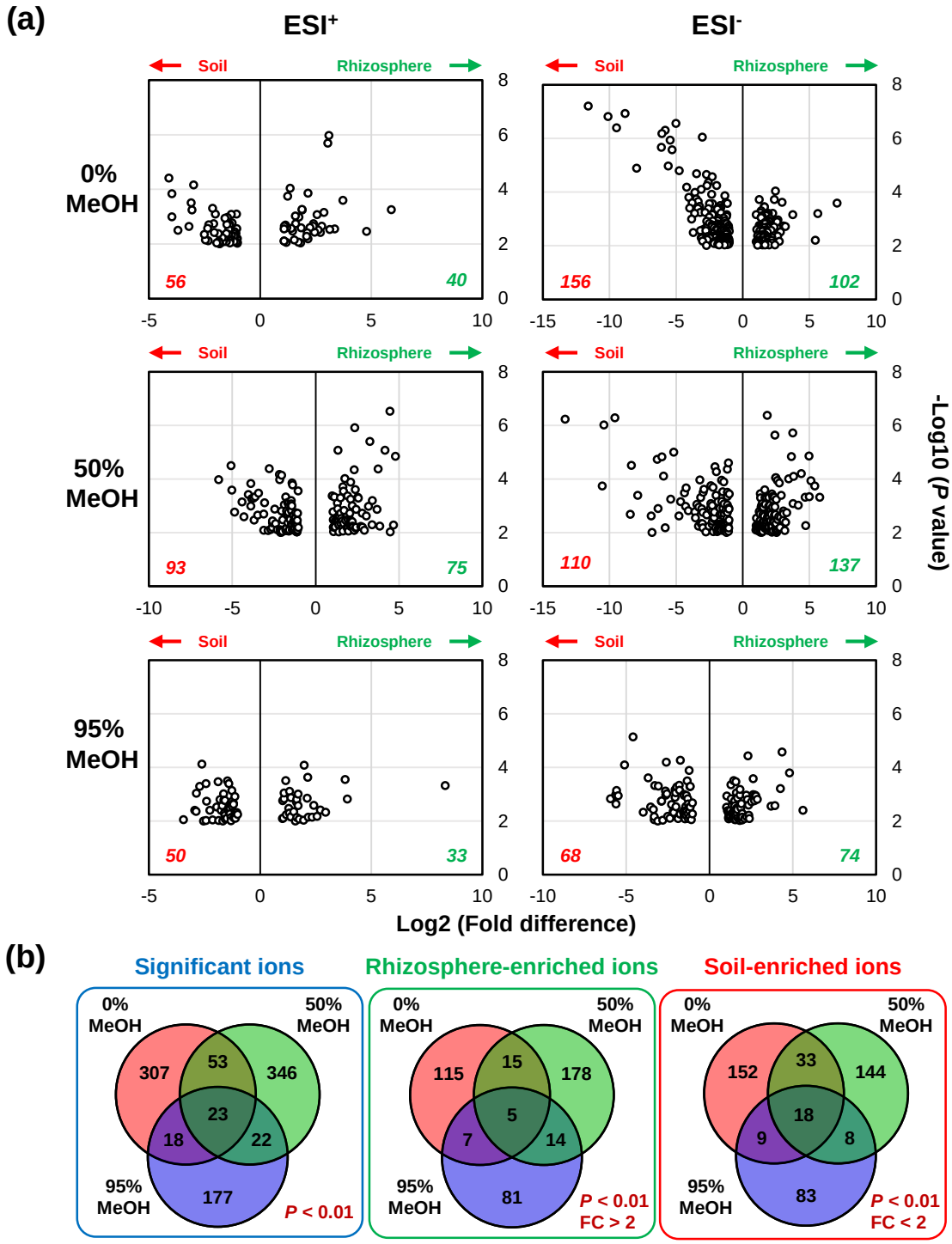


Figure 5. Quantitative differences in metabolite abundance between extracts from control soil and Arabidopsis soil.

(**a**) Volcano plots expressing statistical enrichment of ions (Welch's *t*-test) as a function of fold-difference in control soil (red; 'soil') and Arabidopsis soil (green; 'rhizosphere'). Data shown represent positive (ESI^+) and negative (ESI^-) ions from extractions with different solutions (indicated by % MeOH). Cut-off values were set at $P < 0.01$ (-Log10 = 2) and fold-change > 2 (Log2 = 1).
(**b**) Venn diagrams showing overlap in ions (cations and anions combined) that are significantly different between control and Arabidopsis soil samples (left panel; $P < 0.01$, Welch's *t*-test; without fold-change threshold), enriched in extracts from Arabidopsis soil (middle panel; > 2-fold enrichment to soil at $P < 0.01$, Welch's *t*-test), and enriched in control soil (right panel; < 2-fold enrichment to rhizosphere at $P < 0.01$, Welch's *t*-test).

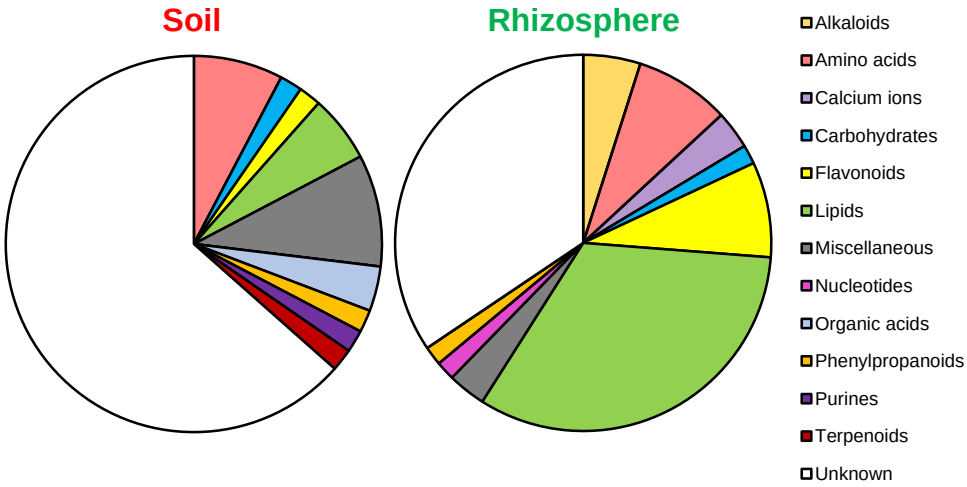


Figure 6. Composition of putative metabolite classes enriched in control soil (left) or Arabidopsis soil (right).

Differentially abundant ions were selected from the top 20-ranking ions of each volcano plot (Figure 5a) and filtered for statistical significance between all soil/extraction solution combinations (ANOVA with Benjamini-Hochberg FDR; $P < 0.01$). The resulting 76 rhizosphere-enriched ions and 75 control soil-enriched ions were corrected for adducts and/or C isotopes (tolerance: m/z = 0.1 Da and RT = 10 s), and cross-referenced against publicly available databases for putative identification. A comprehensive table of all rhizosphere- and soil-enriched markers is presented in Supplemental Table S1. Multiple ions putatively annotating to the same metabolite were counted additively towards the metabolite classes in the pie-charts. Putative metabolites that unlikely accumulate as natural products in (rhizosphere) soil (*e.g.* synthetic drugs, mammalian hormones) were not included in the final selection presented. Miscellaneous: putative metabolites that do not belong to any of the other metabolite classes listed. Unknown: ion markers that could not be assigned to any known compound.

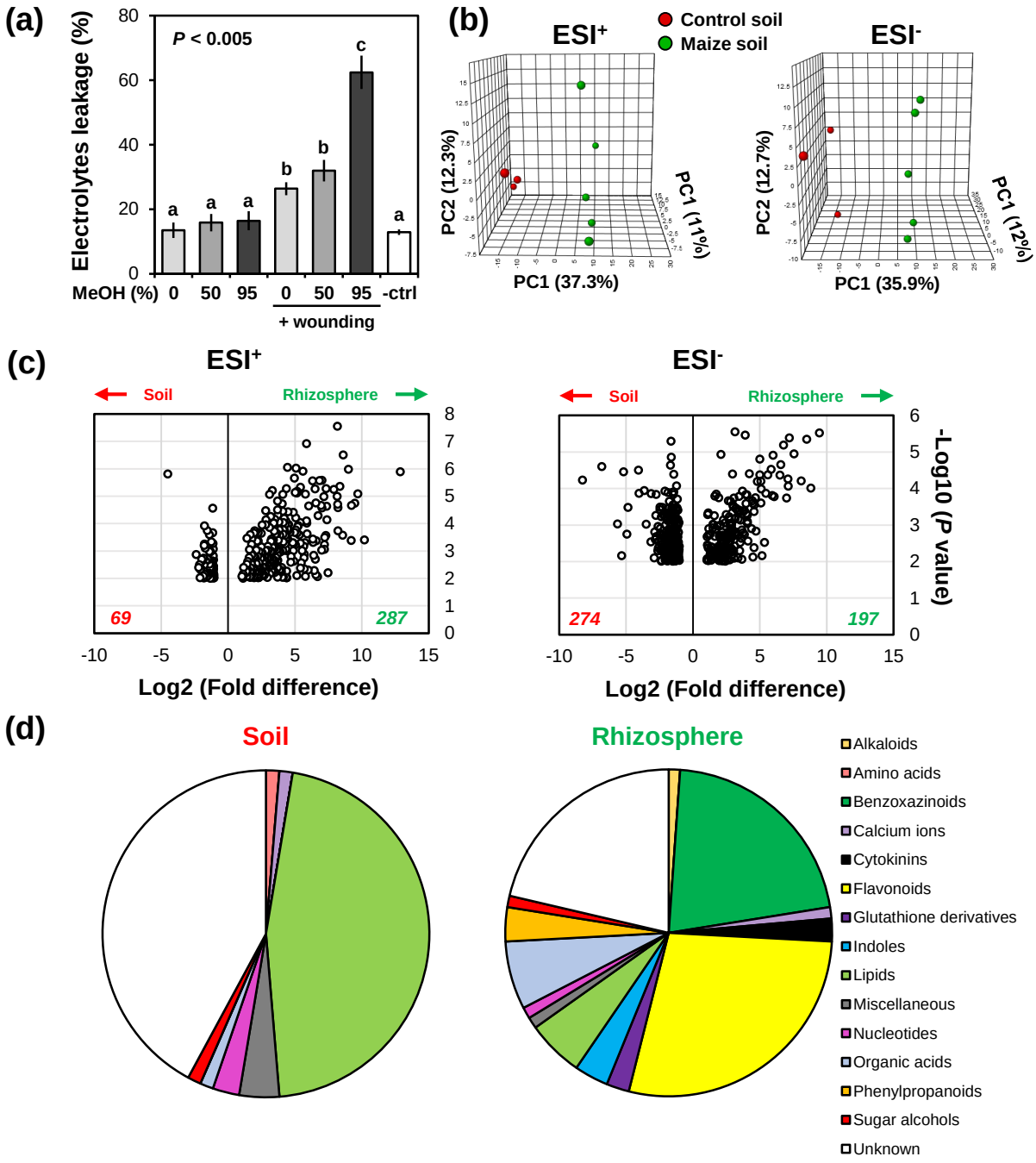
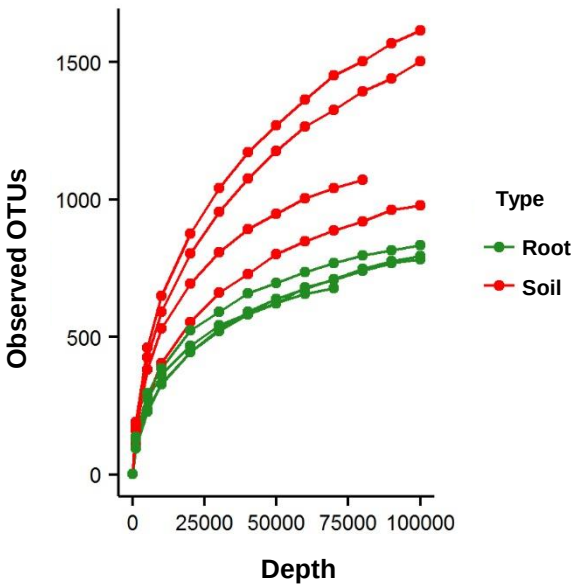


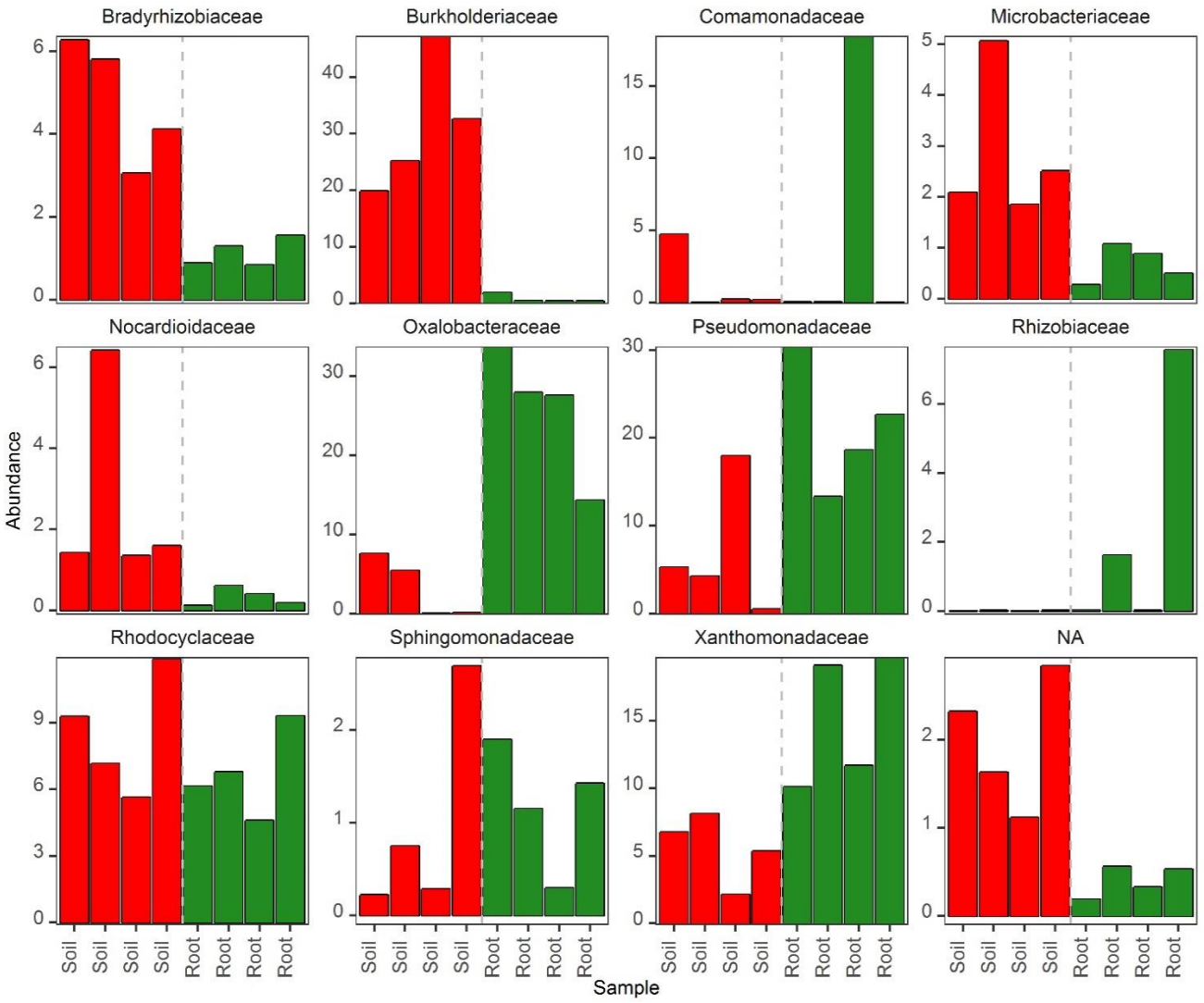
Figure 7. Applicability of the profiling method for maize in agricultural soil.

The experimental system for extracting soil chemistry was based on 50-mL collection tubes filled with a mixture of agricultural soil from arable farmland and perlite (75:25, v/v). Samples were extracted with the 50% MeOH (v/v) solution 17 days after planting.
(**a**) Quantification of maize root damage after direct exposure to the extraction solutions. Five day-old maize roots were incubated for 1 min in acidified extraction solutions containing 0%, 50% or 95% MeOH (v/v) and tested for electrolytes leakage by conductivity. For details, see legend to Figure 3a. Shown are average levels of conductivity (n = 4, ± SEM), relative to the maximum level of conductivity after tissue lysis (set at 100%). Statistically significant differences between treatments were determined by a Welch's F test for ranked data (P values indicated in the upper left corner of each panel), followed by Games-Howell post-hoc tests (P < 0.05; different letters indicate statistically significant differences).
(**b**) Unsupervised 3D-PCA, showing global differences in metabolic profiles between control soil (red) and maize soil (green). Shown are data from extracts with the 50% MeOH (v/v) extraction solution. For further details, see legend to Figure 4.
(**c**) Volcano plots expressing statistical enrichment of ions (Welch's t-test) as a function of fold-difference in control soil (red; 'soil') and maize soil (green; 'rhizosphere'). Cut-off values were set at P < 0.01 (-Log10 = 2) and fold-change > 2 (Log2 = 1).
(**d**) Relative composition of putative metabolite classes enriched in enriched in control soil (left) or maize soil (right). Differentially abundant metabolites were selected from the top 50-ranking ions of each volcano plot (ESI^+ and ESI^-; **c**), corrected for adducts and/or C isotopes (tolerance: m/z = 0.1 Da and RT = 10 s), and cross-referenced against publicly available databases for putative identification. A comprehensive table of all rhizosphere- and soil-enriched markers is presented in Supplemental Table S2. Putative metabolites that unlikely accumulate as natural products in (rhizosphere) soil (*e.g.* synthetic drugs, mammalian hormones) were not included in the final selection presented. For further details, see legend to Figure 6.



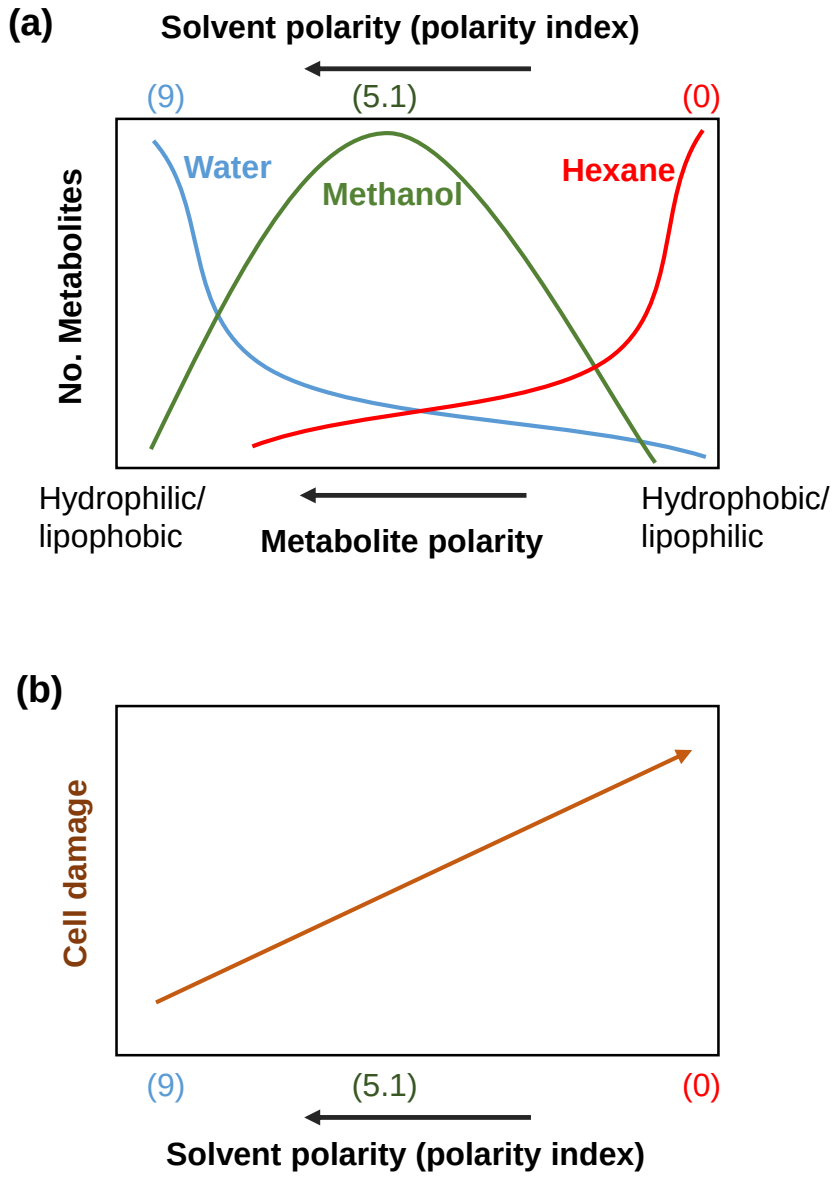
Supplemental Figure S1. Rarefaction curves of detected OTUs.

Shown are curves after removal of singletons for replicate root + rhizosphere samples (green) and control soil samples (red).



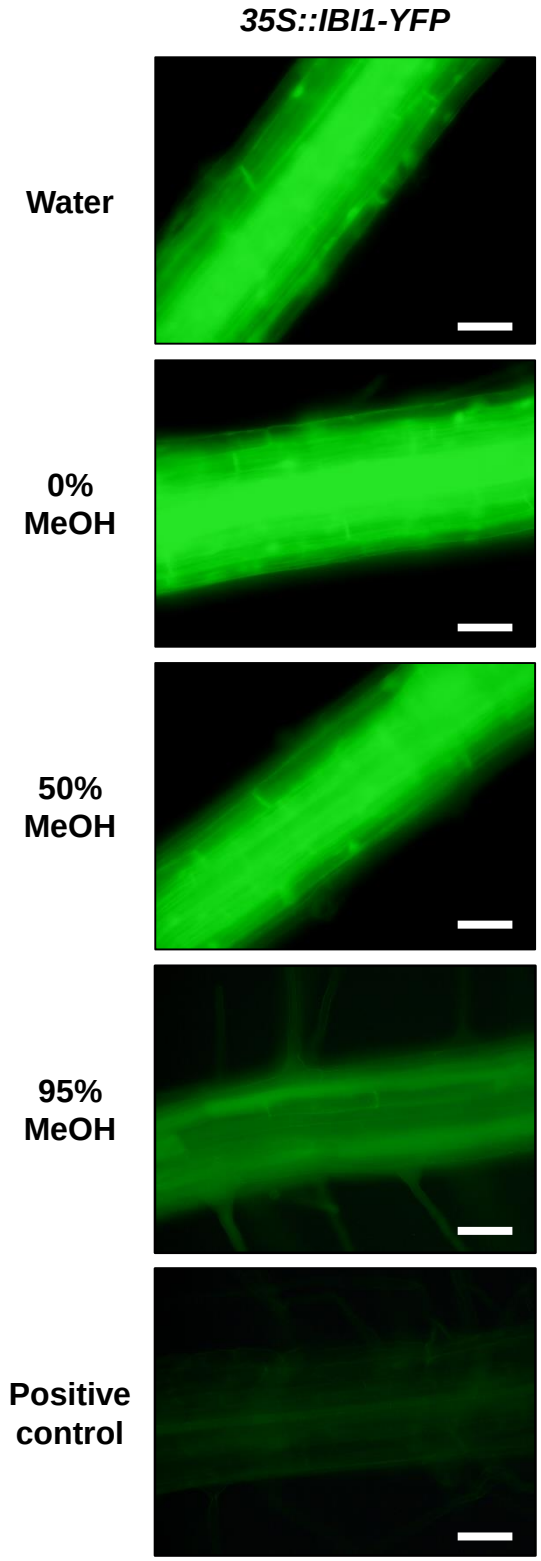
Supplemental Figure S2. Relative abundance (%) of selected families in control soil samples ('Soil'; red) and root + rhizosphere samples ('Root'; green) from the Arabidopsis growth system.

Shown are families containing OTUs with relative abundances > 2% in one or more samples. Each bar represents an individual biological replicate. NA, taxonomy not available.



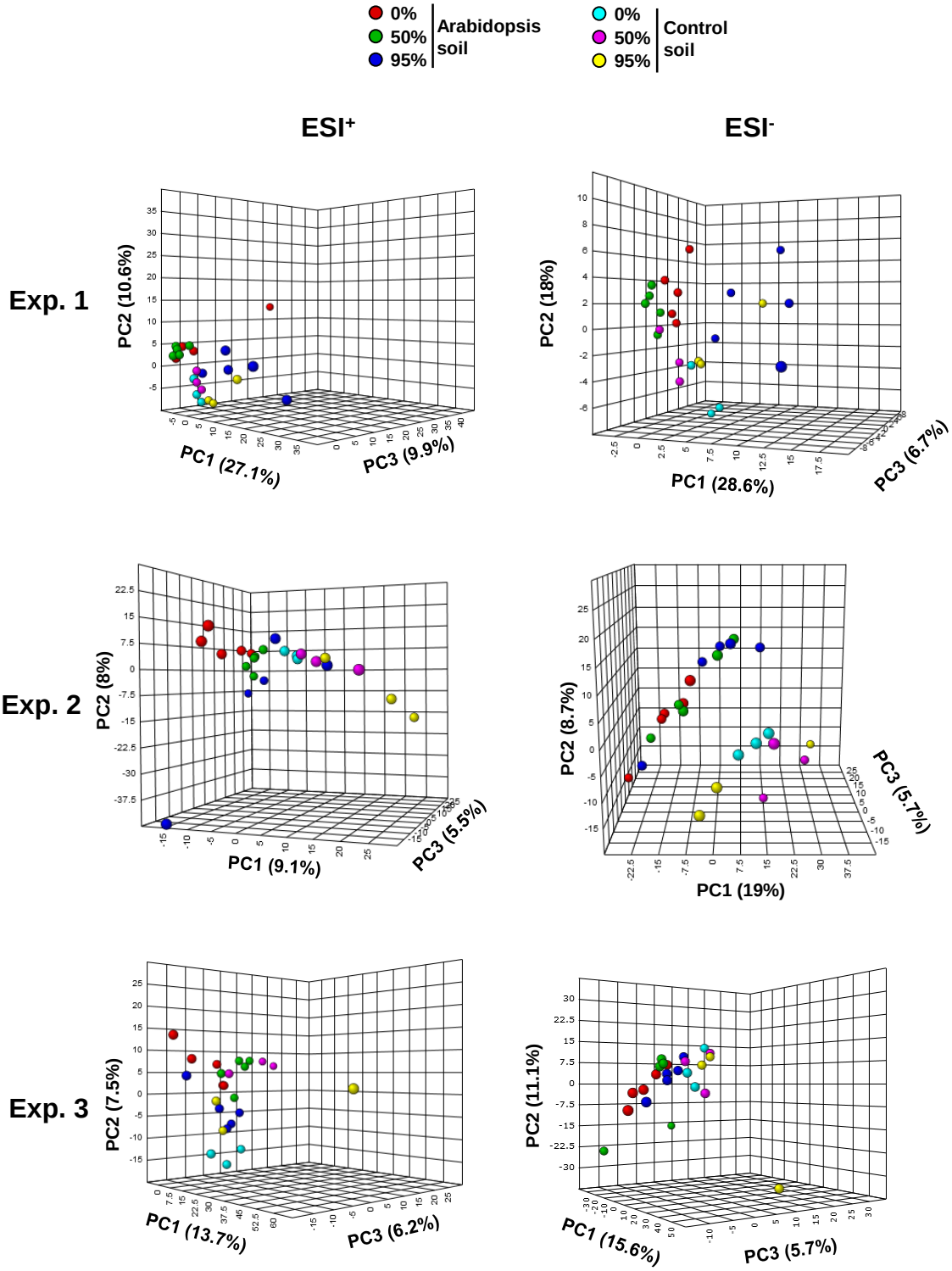
Supplemental Figure S3. Model of expected impacts of solvent polarity on the extraction of soil metabolites.

(**a**) Examples of solvent polarities and their impact on the type of metabolites extracted. Polarity index of water, methanol (MeOH) and hexane are shown within parentheses.
(**b**) Hypothesized impact of solvent polarity on cell damage of plant roots and soil microbes.



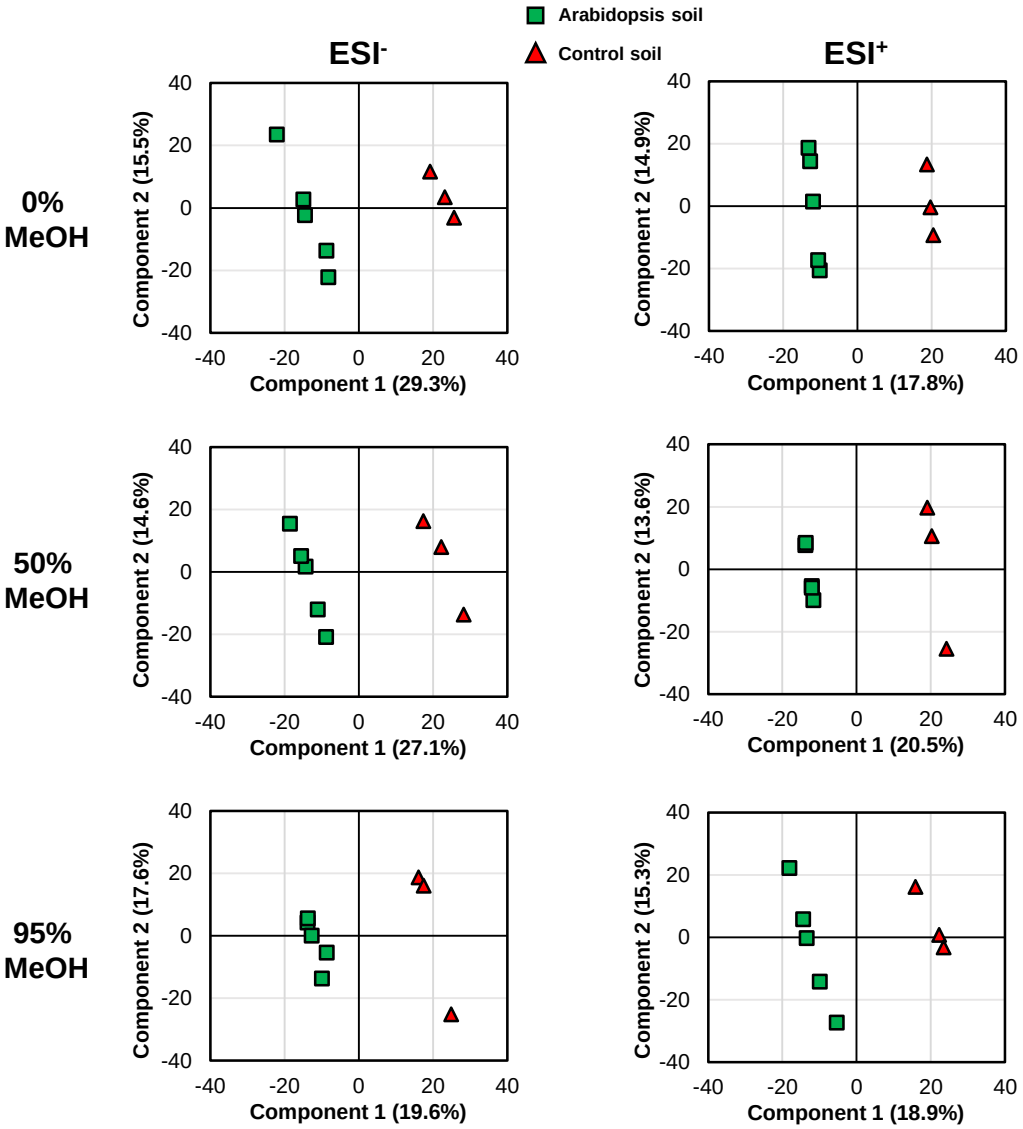
Supplemental Figure S4. Epi-fluorescence microscopy analysis of cell damage in Arabidopsis roots after exposure to MeOH-containing extraction solutions.

Transgenic roots producing the cytoplasmic aspartyl-tRNA synthase IBI1 fused to YFP (*35S::IBI1:YFP*; Luna *et al.*, 2014) were incubated for 1 min in water or acidified extraction solutions with increasing MeOH concentration (0, 50 or 95% MeOH, v/v + 0.05% formic acid, v/v). After incubation, roots were then rinsed in sterile water, and analysed for YFP fluorescence. Photographs show representative examples from observations of at least 12 roots for each treatment. As a positive control for cell damage, roots were incubated in 100% MeOH for 15 min. The experiment was performed four time with similar results. Scale bars: 50 µm.



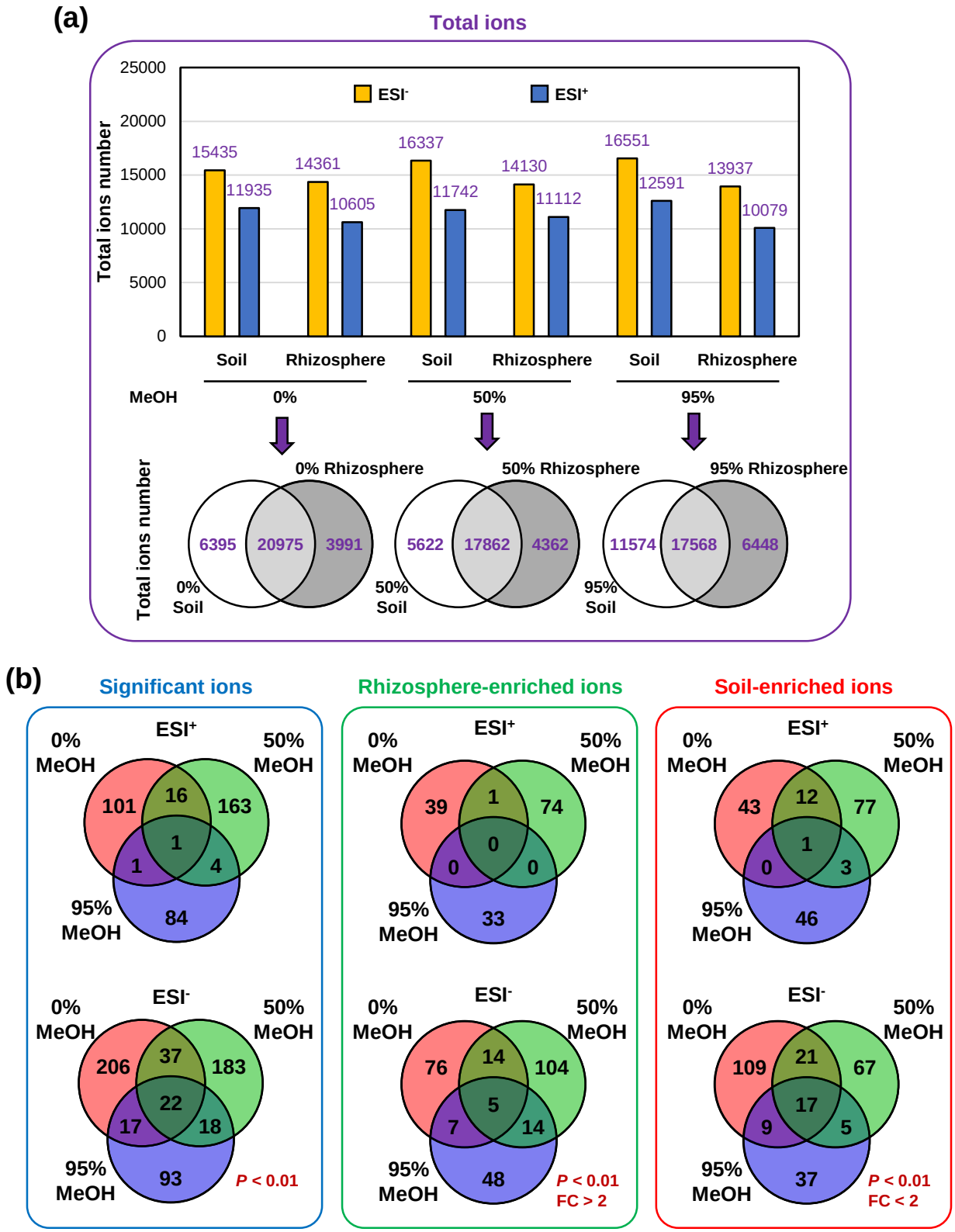
Supplemental Figure S5. Reproducibility of differences in metabolite profiles between control and Arabidopsis soil over three independent experiments.

Shown are unsupervised three-dimensional principal component analyses (3D-PCA) from extracts by the different solutions (indicated by % MeOH). Ions (*m/z* values) were obtained by UPLC-Q-TOF in positive (ESI^+, left panels) and negative (ESI^-, right panel) ionization modes. Analysis was carried out with MetaboAnalyst (v. 3.0), after median normalization, cube-root transformation and Pareto scaling of data. In parentheses are shown the percentages of variation explained by each principal component.



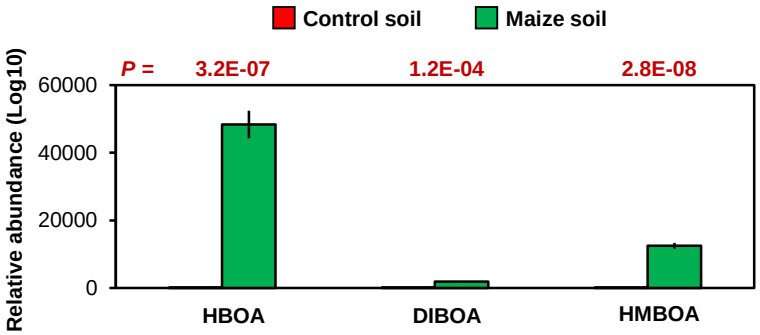
Supplemental Figure S6. Binary PLS-DA analysis of metabolite profiles from control soil and Arabidopsis soil for different extraction solutions (indicated by % MeOH).

Ions (*m/z* values) were obtained by UPLC-Q-TOF analysis in both positive (ESI^+, left panels) and negative (ESI^-, right panel) ionization mode. Prior to analysis, data were median-normalized, cube-root-transformed and Pareto-scaled. All R^2 (correlation) and Q^2 (predictability) values of PLS-DA models were above 0.94 and 0.59, respectively.



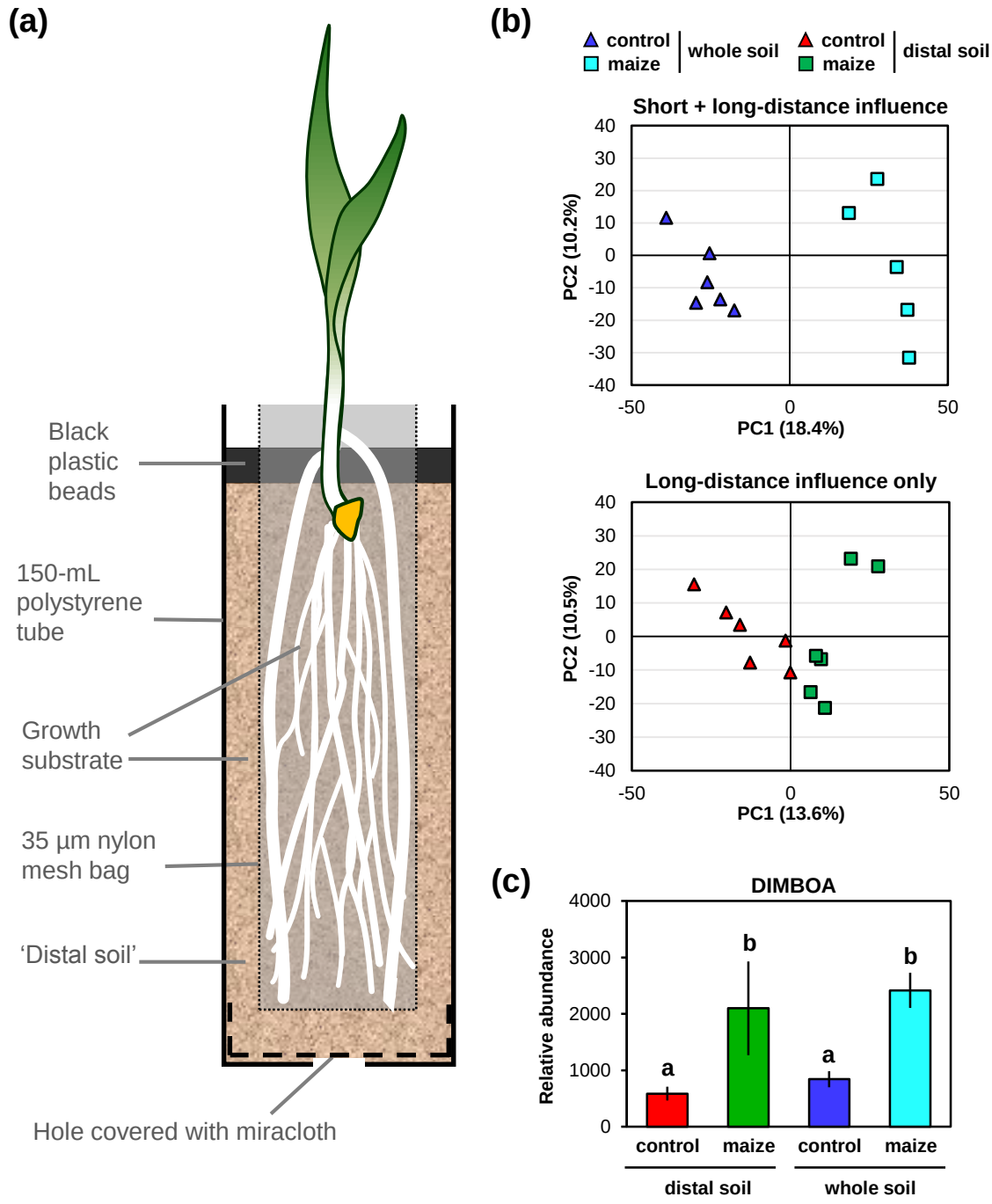
Supplemental Figure S7. Quantitative differences in detected ions (UPLC-Q-TOF) between extracts from control and Arabidopsis soil.

(**a**) Total numbers of ions (top) detected in Arabidopsis soil and control soil after extraction with the different solutions (indicated by % MeOH). Venn diagrams (bottom) show overlap in total ion numbers between extracts for each extraction solution.
(**b**) Venn diagrams showing overlap in cations (ESI^+) and anions (ESI^-) that are statistically different between control and Arabidopsis soil (left panel; $P < 0.01$, Welch's *t*-test), that are enriched in extracts from Arabidopsis soil (middle panel; > 2-fold enrichment to soil at $P < 0.01$, Welch's *t*-test), and that enriched are in extracts from control soil (right panel; < 2-fold enrichment to soil at $P < 0.01$, Welch's *t*-test).



Supplemental Figure S8. Relative quantities of selected benzoxazinoid ions in extracts from maize soil and corresponding control soil.

Selective ions (*m/z*) of HBOA (2-hydroxy-4H-1,4-benzoxazin-3-one), DIBOA (2,4-dihydroxy-1,4-benzoxazin-3-one) and 2-hydroxy-7-methoxy-2H-1,4-benzoxazin-3(4H)-one were detected on the basis of retention time and *m/z* value, using UPLC-Q-TOF (ESI^+, Δppm = 0). Charts indicate means of relative abundances (n = 5, ± SEM). Levels of statistical significance are indicated in red above the corresponding bars (Student's *t*-test).



Supplemental Figure S9. Profiling distal rhizosphere chemistry.

(**a**) Experimental growth system to profile chemistry of distal rhizosphere fractions. Maize was grown within nylon mesh bags inside 150-mL tubes, containing agricultural soil from arable farmland and perlite (75:25, v/v). Similar plant-free tubes were constructed as controls. After 24 days of growth, chemicals were extracted with the acidified 50% MeOH solution from either the entire pot (whole soil), or the soil surrounding the root containing mesh bag after its careful removal (distal soil).
(**b**) Binary PCAs showing chemical rhizosphere effects in whole soil fractions (upper panel; short + long distance influence) and distal soil fractions (lower panel; long distance influence), illustrating that the rhizosphere extends beyond soil that is closely associated with roots.
(**c**) Targeted quantification of DIMBOA by UPLC-Q-TOF. Shown are average ion intensities (± SEM; $n = 6$), normalised by soil weight. Letters indicate statistically significant differences between soil types (Student's t-test, $P < 0.05$).

Supplemental Table S1. Putative identification of Arabidopsis rhizosphere- and control soil-enriched metabolic markers

[1] Percentages indicate relative MeOH contents of the extraction solutions.

[2] *P* values are derived from ANOVA followed by false discovery rate correction (Benjamini-Hochberg).

[3] Retention times (RT) and accurate *m/z* values, detected by UPLC-Q-TOF in negative (-) or positive (+) ion mode.

[4] Predicted parameters from the METLIN database using the detected accurate *m/z*.

[5] Putative metabolites and their corresponding pathways were validated by information from the PubMed chemical database.

[6] Putative metabolites that unlikely accumulate in (rhizosphere) soil.

RHIZOSPHERE -ENRICHED

Sample	Extraction solution[1]	*P* value[2]	RT (min)[3]	Detected *m/z*[3]	Ion mode[3]	Predicted mass[4]	*Adduct*[4]	Δppm[4]	Putative Compound[4]	Predicted Formula[4]	Putative Pathway[5]	soil- or plant-derived[6]
	0%	5.9E-05	7.4	336.117	-	337.131	*[M-H]-*	21	(S)-cis-N-methylstylopine	C20H19NO4	Alkaloids	
	95%	8.2E-03	1.1	612.166	-	567.189	*[M+FA-H]-*	35	Neoacrimarine H	C33H29NO8	Alkaloids	
	0%	3.1E-03	5.2	657.315	+	656.310	*[M+H]+*	3	Disinomenine	C38H44N2O8	Alkaloids	
	0%	1.8E-05	5.9	116.052	+	151.063	*[M+H-2H2O]+*	12	2-Phenylglycine	C8H9NO2	Amino acids	
	95%	1.2E-03	5.0	265.106	+	264.104	*[M+H]+*	21	Thiamine	C12H16N4OS	Amino acids	
	0%	7.5E-03	6.9	109.039	-	110.048	*[M-H]-*	15	Imidazole-4-acetaldehyde	C5H6N2O	Amino acids	
	50%	9.5E-04	7.2	414.322	+	413.314	*[M+H]+*	1	N-stearoyl glutamic acid	C23H43NO5	Amino acids	
	95%	2.4E-03	1.3	668.121	-	669.144	*[M-H]-*	23	Enterochelin	C30H27N3O15	Amino acids	
	50%	3.1E-04	1.4	580.656	-	599.686	*[M-H2O-H]-*	19	Calcium trimetaphosphate	Ca3H6O18P6	Calcium source from Plants	
	50%	6.5E-05	1.2	620.662	-	599.686	*[M+Na-2H]-*	2	Calcium trimetaphosphate	Ca3H6O18P6	Calcium source from Plants	
	0%	2.6E-03	9.3	776.177	+	775.195	*[M+H]+*	32	Reduced coenzyme F420	C29H38N5O18P	Carbohydrates	
	50, 95%	6.0E-04	7.3	383.156	-	384.163	*[M-H]-*	0	2,3-Butanediol apiosylglucoside	C15H28O11	Flavonoids	
	50%	5.2E-05	1.3	699.141	-	654.143	*[M+FA-H]-*	0	Tamarixetin 5-glucoside-7-glucuronide	C28H30O18	Flavonoids	
	50%	8.1E-04	1.2	727.147	-	728.180	*[M-H]-*	35	Primflaside	C31H36O20	Flavonoids	
	0%	9.5E-04	9.0	479.275	+	478.272	*[M+H]+*	8	3-Geranyl-4,2\',4\',6\'-tetrahydroxy-5-prenyldihydrochalcone	C30H38O5	Flavonoids	Unlikely
	0%	7.1E-03	1.2	648.153	-	649.177	*[M-H]-*	25	Pyranodelphinin A	C30H33O16	Flavonoids	
	50%	1.9E-04	1.2	665.136	-	666.143	*[M-H]-*	0	Cyanidin 3-(6\'\'-malonylsambubioside)	C29H30O18	Flavonoids	
	0%	4.3E-03	9.4	452.924	-	453.942	*[M-H]-*	24	4,4'-Diisothiocyano-2,2'-stilbenedisulfonic acid	C16H10N2O6S4	Isothiocyanates	Unlikely
	0%	2.1E-06	8.1	347.327	+	364.334	*[M+H-H2O]+*	12	5,9-tetracosadienoic acid	C24H44O2	Lipids	
	0%	4.8E-03	6.7	594.585	+	593.575	*[M+H]+*	5	Ceramide	C38H75NO3	Lipids	
	0%	4.4E-03	9.1	664.458	+	663.448	*[M+H]+*	4	Phosphoserine	C34H66NO9P	Lipids	
	50%	6.0E-07	1.3	816.548	-	817.562	*[M-H]-*	8	Phosphatidylethanolamine	C47H80NO8P	Lipids	
	50%	3.7E-03	7.3	395.346	+	394.324	*[M+H]+*	38	Dehydroergosterol	C28H42O	Lipids	
	50%	3.1E-03	6.4	420.293	+	437.291	*[M+H-H2O]+*	12	Phosphatidylethanolamine	C21H44NO6P	Lipids	
	95%	2.0E-05	7.7	441.298	+	440.290	*[M+H]+*	1	26,26,26-trifluoro-25-hydroxy-27-norcholecalciferol	C26H39F3O2	Lipids	Unlikely
	50%	6.1E-04	1.6	480.787	-	499.803	*[M-H2O-H]-*	4	Butter acids	C15H8Cl8O2	Lipids	
	50%	1.6E-03	7.2	485.382	+	520.392	*[M+H-2H2O]+*	6	Ginsenoyne A linoleate	C35H52O3	Lipids	
	50%	3.5E-03	9.1	503.461	+	538.460	*[M+H-2H2O]+*	27	Diacylglycerol	C33H62O5	Lipids	
	50%	3.5E-04	9.3	658.604	+	675.593	*[M+H-H2O]+*	21	Diacylglycerol	C43H79D5O5	Lipids	
	95%	3.8E-03	1.1	672.565	-	673.549	*[M-H]-*	34	Glucosylceramide	C38H75NO8	Lipids	
	50%	8.7E-04	9.2	763.519	+	762.520	*[M+H]+*	10	Phosphatidic acid	C44H75O8P	Lipids	
	0%	3.8E-06	9.0	771.658	-	772.658	*[M-H]-*	9	Triacylglycerol	C49H88O6	Lipids	
	95%	1.4E-03	2.8	809.600	-	810.614	*[M-H]-*	8	Phosphatidic acid	C47H87O8P	Lipids	
	95%	6.0E-07	1.3	816.548	-	817.562	*[M-H]-*	8	Phosphatidylethanolamine	C47H80NO8P	Lipids	
	0%	3.1E-03	1.4	834.524	-	835.536	*[M-H]-*	6	Phosphoserine	C46H78NO10P	Lipids	
	50%	8.2E-03	1.2	856.508	-	857.521	*[M-H]-*	6	Phosphoserine	C48H76NO10P	Lipids	
	95%	5.0E-04	6.7	1023.795	+	1022.830	*[M+H]+*	41	Triacylglycerol	C68H110O6	Lipids	
	95%	1.9E-03	1.5	1057.815	-	1012.846	*[M+FA-H]-*	27	Triacylglycerol	C67H112O6	Lipids	
Rhizosphere	95%	< E-09	6.3	265.182	+	264.173	*[M+H]+*	8	12-Oxo-2,3-dinor-10,15-phytodienoic acid	C16H24O3	Lipids/JA metabolism	
	0%	7.8E-04	1.7	92.927	-	111.948	*[M-H2O-H]-*	30	2,2-Dichloroacetaldehyde	C2H2Cl2O	Miscellaneous	Unlikely
	0%	4.8E-03	1.2	306.918	+	305.910	*[M+H]+*	1	Mitobronitol	C6H12Br2O4	Miscellaneous/Alcohol	Unlikely
	0%	2.0E-03	1.1	338.812	-	357.844	*[M-H2O-H]-*	41	2,2\',4,4\',5,5\'-Hexachlorobiphenyl	C12H4Cl6	Miscellaneous/Aromatics	Unlikely
	0, 50, 95%	1.0E-06	1.0	358.836	-	359.837	*[M-H]-*	16	Haloprogin	C9H4Cl3IO	Miscellaneous/Aromatics	
	50%	2.8E-03	1.1	380.818	-	359.837	*[M+Na-2H]-*	15	Haloprogin	C9H4Cl3IO	Miscellaneous/Aromatics	
	50%	5.4E-06	1.5	538.675	-	493.689	*[M+FA-H]-*	21	Decachlorobiphenyl	C12Cl10	Miscellaneous/Aromatics	Unlikely
	50%	< E-09	1.9	203.025	+	202.011	*[M+H]+*	31	4-carboxy-2-hydroxy-cis,cis-muconic acid	C7H6O7	Miscellaneous/Carboxylic acids	Unlikely
	0, 50%	1.9E-03	1.1	216.907	-	235.932	*[M-H2O-H]-*	30	2-Bromomaleylacetate	C6H5BrO5	Miscellaneous/Carboxylic acids	Unlikely
	0%	1.3E-04	1.4	318.838	-	337.863	*[M-H2O-H]-*	19	1,2,3,7,8-Pentachlorodibenzofuran	C12H3Cl5O	Miscellaneous/Dioxins	Unlikely
	95%	7.1E-03	1.2	680.140	-	635.142	*[M+FA-H]-*	0	Aflatoxin B1 exo-8,9-epoxide-GSH	C27H29N3O13S	Miscellaneous/Mycotoxins	Unlikely
	95%	1.1E-05	4.6	1051.437	-	1006.436	*[M+FA-H]-*	2	Oxytocin	C43H66N12O12S2	Miscellaneous/Peptide hormones	Unlikely
	0%	3.2E-03	6.8	1126.544	+	1125.501	*[M+H]+*	31	[Tyr(PO3H2)4]-Angiotensin II	C50H72N13O15P	Miscellaneous/Peptide hormones	Unlikely
	0%	6.9E-05	2.0	181.994	+	217.006	*[M+H-2H2O]+*	3	PROPANIL	C9H9Cl2NO	Miscellaneous/Pesticides	Unlikely
	0%	7.8E-03	4.7	291.058	-	270.089	*[M+Na-2H]-*	20	4-O-Methylpinosylvic acid	C16H14O4	Phenylpropanoids	
	95%	4.3E-03	1.2	646.088	-	665.125	*[M-H2O-H]-*	28	NADH	C21H29N7O14P2	Pyridine nucleotides	
	50%	8.0E-05	5.8	360.978	+	395.986	*[M+H-2H2O]+*	14	Sesquiterpene	C15H23Br2Cl	Terpenoids	Unlikely
	95%	3.7E-03	9.2	74.781	-						Unknown	
	0%	6.7E-05	9.3	87.183	+						Unknown	
	50%	7.8E-03	1.4	218.880	-						Unknown	
	50%	1.5E-05	9.3	225.793	+						Unknown	
	0%	8.6E-04	1.2	262.896	-						Unknown	
	50%	3.3E-03	1.8	279.966	+						Unknown	
	0, 50, 95%	1.8E-09	0.9	396.789	-						Unknown	
	95%	5.2E-03	1.7	400.792	-						Unknown	
	95%	2.6E-05	6.0	430.389	+						Unknown	
	0%	4.1E-05	4.5	444.801	-						Unknown	
	50%	2.9E-05	1.3	470.700	-						Unknown	
	50, 95%	2.2E-04	1.6	518.751	-						Unknown	
	50%	9.0E-03	7.2	540.856	+						Unknown	
	50%	1.6E-03	1.1	558.693	-						Unknown	
	95%	5.8E-03	7.6	598.899	+						Unknown	
	50%	2.7E-06	1.2	618.651	-						Unknown	
	0%	9.5E-03	2.8	620.761	-						Unknown	
	0%	6.6E-03	9.4	766.946	+						Unknown	
	95%	1.7E-04	7.1	862.800	+						Unknown	
	95%	8.4E-07	9.3	1067.185	+						Unknown	
	95%	5.5E-04	9.3	1119.101	+						Unknown	

SOIL-ENRICHED

Sample	Extraction solution[1]	*P* value[2]	RT (min)[3]	Detected *m/z*[3]	Ion mode	Predicted mass[4]	*Adduct*[4]	Δppm[4]	Putative Compound[4]	Predicted Formula[4]	Putative Pathway[5]	soil- or plant-derived[6]
	50%	< E-13	1.2	130.086	-	131.095	*[M-H]-*	10	L-Isoleucine	C6H13NO2	Amino acids	
	50%	1.89E-04	1.0	216.035	+	215.019	*[M+H]+*	38	O-Phospho-4-hydroxy-L-threonine	C4H10NO7P	Amino acids	
	50%	5.77E-05	3.0	245.079	+	280.092	*[M+H-2H2O]+*	0	Methionyl-Methionine	C10H20N2O3S2	Amino acids	
	50%	1.12E-04	1.4	284.061	+	283.046	*[M+H]+*	28	N2-Acetyl-L-aminoadipyl-δ-phosphate	C8H14NO8P	Amino acids	
	50%	1.41E-07	1.3	1103.693	+	215.986	*[M+H]+*	46	S-Methyl-3-phospho-1-thio-D-glycerate	C4H9O6PS	Carbohydrates	
	95%	8.67E-05	6.1	335.033	+	334.032	*[M+H]+*	20	Heptahydroxyflavone	C15H10O9	Flavonoids	
	95%	2.36E-03	7.6	553.287	+	570.289	*[M+H-H2O]+*	1	2-O-(beta-D-galactopyranosyl-(1->6)-beta-D-galactopyranosyl) 2S,3R-dihydroxytridecanoic acid	C25H46O14	Lipids	Unlikely
	95%	6.51E-08	8.0	553.513	+	552.491	*[M+H]+*	27	Linoleyl arachidonate	C38H64O2	Lipids	
	50%	2.05E-05	4.9	620.535	+	619.530	*[M+H]+*	3	Diacylglycerol	C39H71D5O5	Lipids	
	0%	5.33E-03	9.4	701.408	+	700.432	*[M+H]+*	43	Phosphatidylglycerol	C37H65O10P	Lipids	
	95%	5.36E-03	8.1	172.063	-	173.069	*[M-H]-*	8	2,6-Piperidinedicarboxylic acid	C7H11NO4	Miscellaneous	
	0, 50%	6.49E-06	0.8	184.926	-	185.932	*[M-H]-*	8	4-Bromo-3,5-cyclohexadiene-1,2-dione	C6H3BrO2	Miscellaneous	Unlikely
	0%	3.48E-03	7.9	296.181	+	331.194	*[M+H-2H2O]+*	0	Currayanine	C23H25NO	Miscellaneous	Unlikely
	50%	2.64E-03	9.4	495.013	+	494.019	*[M+H]+*	27	Sodium cumeneazo-β-naphthol disulfonate	C19H16N2Na2O7S2	Miscellaneous	Unlikely
	0%	7.92E-07	9.3	163.943	+	198.948	*[M+H-2H2O]+*	47	Bronopol	C3H6BrNO4	Miscellaneous/Antimicrobials	Unlikely
	95%	3.16E-03	9.4	1088.122	+	1123.158	*[M+H-2H2O]+*	21	Mycolic acid	C77H150O3	Miscellaneous/Antimicrobials	Unlikely
	50%	1.01E-05	0.8	110.009	+	145.020	*[M+H-2H2O]+*	18	3,4-Dehydrothiomorpholine-3-carboxylate	C5H7NO2S	Miscellaneous/Carboxylic acids	
	0, 95%	5.65E-13	1.1	148.043	+	147.035	*[M+H]+*	2	Thiomorpholine 3-carboxylate	C5H9NO2S	Miscellaneous/Carboxylic acids	
	0, 50%	7.86E-06	1.1	346.099	+	345.091	*[M+H]+*	0	Clopamide	C14H20ClN3O3S	Miscellaneous/Diuretics	Unlikely
	50%	3.85E-03	1.7	346.096	+	345.091	*[M+H]+*	7	Clopamide	C14H20ClN3O3S	Miscellaneous/Diuretics	Unlikely
	50%	5.41E-03	6.9	382.128	-	383.134	*[M-H]-*	2	Fluazifop butyl	C19H20F3NO4	Miscellaneous/Herbicides	Unlikely
	95%	1.44E-03	2.9	303.049	-	304.058	*[M-H]-*	4	Brompheniramine (monodemethylated)	C15H17BrN2	Miscellaneous/Histamines	Unlikely
	0, 50%	1.44E-04	0.8	183.928	-	184.934	*[M-H]-*	8	Iodoacetamide	ICH2CONH2	Miscellaneous/Lipids	Unlikely
	95%	1.67E-04	8.6	584.329	-	585.345	*[M-H]-*	15	Janthitrem B	C37H47NO5	Miscellaneous/Mycotoxins	Unlikely
	0%	8.26E-03	1.2	234.048	+	269.061	*[M+H-2H2O]+*	0	2-(p-Methoxyphenyl)-3-(m-chlorophenyl)acrylonitrile	C16H12ClNO	Miscellaneous/Nitriles	Unlikely
	50%	1.63E-05	0.8	199.045	+	198.039	*[M+H]+*	5	Nitrofurazone	C6H6N4O4	Miscellaneous/Nitrofurans	Unlikely
	95%	< E-13	1.7	288.046	-	289.054	*[M-H]-*	1	Isocarbophos	C11H16NO4PS	Miscellaneous/Pesticides	Unlikely
	50%	4.74E-04	1.4	303.871	-	258.876	*[M+FA-H]-*	10	Tecnazene	C6HCl4NO2	Miscellaneous/Pesticides	Unlikely
	95%	2.65E-04	7.1	775.262	+	774.247	*[M+H]+*	9	7,8-Dihydromethanopterin	C30H43N6O16P	Miscellaneous/Pteridines	
	0%	5.61E-03	7.7	549.916	+	584.937	*[M+H-2H2O]+*	14	HLo7	C15H17I2N5O4	Miscellaneous/Pyridines	
	95%	1.64E-04	4.2	596.085	-	551.106	*[M+FA-H]-*	32	11-O-Demethylpradimicinone II	C27H21NO12	Miscellaneous/Quinones	Unlikely
	0%	1.88E-03	9.3	405.185	+	404.175	*[M+H]+*	5	Chlormadinone acetate	C23H29ClO4	Miscellaneous/Steroids	Unlikely
	0%	5.20E-03	7.8	457.263	+	456.255	*[M+H]+*	2	Lithocholic acid sulfate	C24H40O6S	Miscellaneous/Steroids	Unlikely
	0%	1.66E-04	1.1	370.066	+	369.059	*[M+H]+*	2	4-Pyridinol, 2-[[[5-(difluoromethoxy)-1H-benzimidazol-2-yl]sulfinyl]methyl]-3-methoxy-	C15H13F2N3O4S	Miscellaneous/Sulfoxides	Unlikely
	0, 50, 95%	2.37E-06	0.9	210.949	-	211.960	*[M-H]-*	18	Thiotropocin	C8H4O3S2	Miscellaneous/Tropolones	Unlikely
	0, 50%	1.00E-04	1.3	216.983	+	215.986	*[M+H]+*	46	S-Methyl-3-phospho-1-thio-D-glycerate	C4H9O6PS	Organic acids	
	50%	2.61E-04	3.3	263.094	+	262.089	*[M+H]+*	8	Thiamine aldehyde	C12H14N4OS	Organic acids	
Control soil	0, 50%	2.26E-07	0.8	182.929	-	183.935	*[M-H]-*	5	Arsonoacetate	C2H5AsO5	Organoarsonic acids	Unlikely
	0%	4.39E-04	6.5	321.094	+	320.090	*[M+H]+*	8	4-Coumaroylshikimate	C16H16O7	Phenylpropanoids	
	0%	6.17E-07	9.4	1049.027	+	1084.067	*[M+H-2H2O]+*	25	Punicalagin	C48H28O30	Phenylpropanoids	Unlikely
	50%	6.61E-03	2.2	1117.153	-	1072.183	*[M+FA-H]-*	25	CoA-glutathione	C31H51N10O22P3S2	Purines	
	50%	1.50E-08	9.3	349.184	+	348.178	*[M+H]+*	4	cis-10-Hydroxylinalyl oxide 7-glucoside	C16H28O8	Terpenoids	
	0, 50%	1.48E-07	0.9	61.987	-						Unknown	
	0, 50%	2.61E-05	1.6	61.988	-						Unknown	
	0, 50%	1.78E-05	1.0	123.941	+						Unknown	
	95%	3.03E-05	1.1	160.841	-						Unknown	
	0%	3.15E-04	1.9	165.936	-						Unknown	
	95%	5.05E-04	2.0	175.968	-						Unknown	
	0, 95%	7.65E-03	0.8	192.957	-						Unknown	
	95%	3.88E-06	4.8	195.925	+						Unknown	
	0, 50%	4.24E-03	1.8	209.948	-						Unknown	
	0, 50%	5.78E-06	0.9	211.946	-						Unknown	
	95%	7.37E-04	9.7	245.894	-						Unknown	
	95%	4.85E-04	1.3	261.870	-						Unknown	
	95%	7.08E-04	1.3	299.842	-						Unknown	
	95%	3.80E-03	5.4	307.150	+						Unknown	
	95%	5.18E-03	4.9	319.154	+						Unknown	
	0%	1.00E-04	1.7	325.894	-						Unknown	
	0%	3.88E-03	1.3	351.873	-						Unknown	
	95%	1.72E-04	1.4	379.788	-						Unknown	
	0%	1.40E-04	1.4	409.834	-						Unknown	
	0, 50%	1.80E-05	1.3	414.120	+						Unknown	
	95%	3.50E-03	1.5	443.807	-						Unknown	
	50%	8.21E-06	1.1	445.806	-						Unknown	
	50%	1.70E-07	1.5	475.778	-						Unknown	
	95%	3.03E-05	9.4	741.928	-						Unknown	
	95%	5.54E-03	8.5	776.674	+						Unknown	
	95%	1.40E-04	6.7	891.384	+						Unknown	
	0%	3.83E-03	2.2	992.858	-						Unknown	
	0%	3.95E-05	0.9	146.965	-						Unknown	
	0, 50%	9.75E-07	1.8	146.965	-						Unknown	
	0, 50%	1.72E-03	0.9	148.968	-						Unknown	
	0, 50%	1.16E-05	0.8	209.948	-						Unknown	
	0, 50%	9.40E-03	1.8	210.949	-						Unknown	
	0, 50%	3.24E-04	1.8	211.946	-						Unknown	

Supplemental Table S2. Putative identification of maize rhizosphere- and control soil-enriched metabolic markers

[1] Fold-change between maize rhizosphere samples and control soil samples.

[2] *P* values are derived from Welch's *t* -test.

[3] Retention times (RT) and accurate *m/z* values, detected by UPLC-Q-TOF in negative (-) or positive (+) ion mode.

[4] Predicted parameters from the METLIN database using the detected accurate *m/z*.

[5] Putative metabolites and their corresponding pathways were validated by information from the PubMed chemical database.

[6] Putative metabolites that unlikely accumulate in (rhizosphere) soil.

RHIZOSPHERE -ENRICHED

Sample	FC	*P* value[2]	RT (min)[3]	Detected *m/z*[3]	Ion mode[3]	Predicted mass[4]	*Adduct*[4]	Δppm[4]	Putative Compound[4]	Predicted Formula[4]	Putative Pathway[5]	Soil- or plant-derived[6]
	96	2.2E-05	2.8	656.303	-	675.325	*[M+H2O-H]-*	6	Jesaconitine	C35H49NO12	Alkaloids	Unlikely
	23	7.6E-03	3.8	512.139	-	513.142	*[M-H]-*	8	Dioxinoacrimarine A	C29H23NO8	Alkaloids	Unlikely
	16	2.6E-04	2.0	434.119	-	413.147	*[M+Na-2H]-*	7	Noscapine	C22H23NO7	Alkaloids	
	833	8.3E-06	2.0	410.106	+	387.117	*[M+Na]+*	0	HDMBOA-Glc	C16H21NO10	Benzoxazinoids	
	392	3.2E-07	2.2	166.050	+	165.043	*[M+H]+*	0	HBOA	C8H7NO3	Benzoxazinoids	
	298	1.8E-04	1.5	426.101	+	403.111	*[M+Na]+*	0	DIM2BOA-Glc	C16H21NO11	Benzoxazinoids	
	292	2.8E-08	2.1	196.061	+	195.053	*[M+H]+*	2	2-Hydroxy-7-methoxy-2H-1,4-benzoxazin-3(4H)-one	C9H9NO4	Benzoxazinoids	
	131	2.7E-06	1.5	410.106	+	387.117	*[M+Na]+*	0	HDMBOA-Glc	C16H21NO10	Benzoxazinoids	
	125	2.2E-05	1.5	380.095	+	357.106	*[M+Na]+*	0	HMBOA-Glc	C15H19NO9	Benzoxazinoids	
	92	5.4E-06	2.2	342.078	+	359.085	*[M+H-H2O]+*	13	TRIBOA-glucoside	C14H17NO10	Benzoxazinoids	
	59	9.7E-05	1.5	396.090	+	373.101	*[M+Na]+*	0	DIMBOA-Glc	C15H19NO10	Benzoxazinoids	
	55	8.5E-06	2.2	188.032	+	165.043	*[M+Na]+*	1	HBOA	C8H7NO3	Benzoxazinoids	
	712	3.0E-06	2.2	164.035	-	165.043	*[M-H]-*	1	HBOA	C8H7NO3	Benzoxazinoids	
	140	8.7E-05	2.0	424.085	-	403.111	*[M+Na-2H]-*	2	DIM2BOA-Glc	C16H21NO11	Benzoxazinoids	
	139	5.5E-05	1.5	402.104	-	403.111	*[M-H]-*	1	DIM2BOA-Glc	C16H21NO11	Benzoxazinoids	
	130	1.8E-04	1.5	356.099	-	357.106	*[M-H]-*	1	HMBOA-Glc	C15H19NO9	Benzoxazinoids	
	92	4.3E-05	1.5	386.109	-	387.117	*[M-H]-*	0	HDMBOA-Glc	C16H21NO10	Benzoxazinoids	
	44	1.4E-04	1.5	372.094	-	373.101	*[M-H]-*	1	DIMBOA-Glc	C15H19NO10	Benzoxazinoids	
	41	3.0E-03	1.6	224.054	-	225.064	*[M-H]-*	10	2-Hydroxy-4,7-dimethoxy-2H-1,4-benzoxazin-3(4H)-one	C10H11NO5	Benzoxazinoids	
	34	2.5E-04	2.0	356.099	-	357.106	*[M-H]-*	0	HMBOA-Glc	C15H19NO9	Benzoxazinoids	
	34	6.3E-05	2.0	386.109	-	387.117	*[M-H]-*	0	HDMBOA-Glc	C16H21NO10	Benzoxazinoids	
	13	1.2E-04	1.5	462.060	-	181.038	*[M-H]-*	22	DIBOA	C8H7NO4	Benzoxazinoids	
	104	8.7E-03	2.0	622.675	+	599.686	*[M+Na]+*	0	Calcium trimetaphosphate	Ca3H6O18P6	Calcium source from Plants	
	64	1.2E-05	2.0	452.097	-	431.121	*[M+Na-2H]-*	3	Ribosylzeatin phosphate	C15H22N5O8P	Cytokinins	
	27	1.5E-03	2.2	398.110	-	377.134	*[M+Na-2H]-*	5	Kinetin-7-N-glucoside	C16H19N5O6	Cytokinins	
	607	2.2E-05	2.0	857.239	+	874.238	*[M+H-H2O]+*	4	Quercetin 3-xylosyl-(1->3)-rhanosyl-(1->6)-[apiosyl-(1->2)-galactoside]	C37H46O24	Flavonoids	
	298	2.7E-05	2.0	828.233	+	863.225	*[M+H-2H2O]+*	25	Cyanidin 3-O-[b-D-Xylopyranosyl-(1->2)-[(4-hydroxybenzoyl)-(->6)-b-D-glucopyranosyl-(1->6)]-b-D-galactopyranoside]	C39H43O22	Flavonoids	
	288	1.0E-05	2.0	379.089	+	356.090	*[M+Na]+*	26	6,7,3\'-Trimethoxy-4\',5\'-methylenedioxyisoflavone	C19H16O7	Flavonoids	
	179	6.2E-03	1.9	824.197	+	801.209	*[M+Na]+*	1	Delphinidin 3-lathyroside 5-(6-acetylglucoside)	C34H41O22	Flavonoids	
	151	3.2E-04	1.9	615.661	+	1185.330	*[M+2Na]2+*	11	Cyanidin 3-(disinapoylsophoroside) 5-glucoside	C55H61O29	Flavonoids	
	129	9.4E-04	1.5	799.186	+	798.164	*[M+H]+*	17	Apigenin 4\'-(2\'\'-feruloylglucuronosyl)-(1->2)-glucuronide	C37H34O20	Flavonoids	
	97	4.3E-04	3.8	565.040	+	542.037	*[M+Na]+*	25	8-Hydroxyapigenin 8-(2\'\'-sulfatoglucuronide)	C21H18O15S	Flavonoids	
	95	7.4E-03	2.0	858.241	+	816.211	*[M+ACN+H]+*	4	Kaempferol 3-(6\'\'-sinapylglucosyl)-(1->2)-galactoside	C38H40O20	Flavonoids	
	84	1.5E-03	2.0	859.168	+	858.170	*[M+H]+*	11	Delphinidin 3-(6\'\'-O-4-malyl-glucoside)-5-(6\'\'\'-O-1-malyl-glucoside)	C35H38O25	Flavonoids	
	81	2.8E-06	2.0	441.120	+	476.132	*[M+H-2H2O]+*	2	Hesperetin 7-O-glucuronide	C23H24O11	Flavonoids	
	60	2.4E-04	2.2	357.584	+	713.157	*[M+2H]2+*	4	Delphinidin 3-(6"-malonyl glucoside) 5-glycoside	C30H33O20	Flavonoids	
	167	2.4E-05	2.2	276.551	+	507.114	*[M+2Na]2+*	17	Delphinidin 3-(acetylglucoside)	C23H23O13	Flavonoids	
	456	9.7E-05	2.0	432.115	-	433.113	*[M-H]-*	20	Pelargonidin 3-galactoside	C21H21O10	Flavonoids	
	110	6.5E-06	2.0	462.126	-	463.124	*[M-H]-*	20	Malvidin-3-O-xyloside	C22H23O11	Flavonoids	
	68	1.7E-04	2.0	433.118	-	434.121	*[M-H]-*	9	Naringenin-7-O-Glucoside	C21H22O10	Flavonoids	
	49	4.3E-05	1.5	440.079	-	419.098	*[M+Na-2H]-*	15	Cyanidin 3-arabinoside	C20H19O10	Flavonoids	
	36	6.9E-03	1.6	594.168	-	595.166	*[M-H]-*	15	Cyanidin 3-rhamnoside 5-glucoside	C27H31O15	Flavonoids	
	32	4.2E-05	1.5	470.091	-	449.108	*[M+Na-2H]-*	17	Cyanidin 3-galactoside	C21H21O11	Flavonoids	
	28	6.4E-05	2.0	463.126	-	464.132	*[M-H]-*	2	Hesperetin 7-O-glucoside	C22H24O11	Flavonoids	
	22	2.2E-04	3.0	176.907	-	433.113	*[M+Na-2H]-*	11	Petunidin 3-arabinoside	C21H21O10	Flavonoids	
	19	3.9E-05	2.0	454.093	-	433.113	*[M+Na-2H]-*	11	Petunidin 3-arabinoside	C21H21O10	Flavonoids	
	17	2.8E-03	1.5	403.109	-	404.111	*[M-H]-*	14	5,3\'-Dihydroxy-3,6,7,4\',5\'-pentamethoxyflavone	C20H20O9	Flavonoids	
	17	5.4E-04	1.5	357.101	-	358.105	*[M-H]-*	8	7-Hydroxy-6,2\',4\',5\'-tetramethoxyisoflavone	C19H18O7	Flavonoids	
	15	1.8E-03	1.7	595.167	-	596.174	*[M-H]-*	0	Naringenin 5,7-di-O-glucoside	C27H32O15	Flavonoids	
Rhizosphere	14	4.3E-04	3.4	979.290	-	958.295	*[M+Na-2H]-*	20	Acacetin 7-rhamnosyl-(1->4\'\')[glucosyl-(1->6\'\')(6\'\'\'-acetylsophoroside)]	C42H54O25	Flavonoids	
	277	6.3E-05	2.0	422.087	-	423.095	*[M-H]-*	1	S-(1,2-dicarboxyethyl)glutathione	C14H21N3O10S	Glutathione derivatives	
	23	2.5E-04	2.0	423.091	-	442.116	*[M-H2O-H]-*	16	S-(4-Nitrobenzyl)glutathione	C17H22N4O8S	Glutathione derivatives	
	297	2.0E-05	2.0	194.045	+	193.038	*[M+H]+*	1	5,6-Dihydroxyindole-2-carboxylic acid	C9H7NO4	Indoles	
	115	2.7E-06	2.1	260.056	+	237.064	*[M+Na]+*	11	Methyl 2,3-dihydro-3,5-dihydroxy-2-oxo-3-indoleacetic acid	C11H11NO5	Indoles	
	82	3.7E-04	1.6	150.056	+	149.048	*[M+H]+*	6	3-Hydroxyindolin-2-one	C8H7NO2	Indoles	
	56	8.6E-03	4.6	982.243	+	959.230	*[M+Na]+*	23	(+)-7-Isojasmonic acid CoA	C33H52N7O18P3S	Jasmonates	
	1178	4.0E-04	1.7	493.281	+	470.301	*[M+Na]+*	18	(25R)-26,26,26-trifluoro-1'alpha;,25-dihydroxycholecalciferol	C27H41F3O3	Lipids	Unlikely
	596	4.2E-04	1.7	494.286	+	511.291	*[M+H-H2O]+*	4	Phosphoserine	C23H46NO9P	Lipids	
	162	5.3E-06	2.0	231.048	+	208.059	*[M+Na]+*	1	Lipoic acid	C8H16O2S2	Lipids	
	83	4.8E-04	2.0	824.725	+	859.739	*[M+H-2H2O]+*	1	Phosphatidylcholine	C50H102NO7P	Lipids	
	16	1.5E-04	2.2	165.038	-	166.048	*[M-H]-*	14	Arabinonic acid	C5H10O6	Lipids	
	59	1.2E-07	2.2	165.042	+	142.053	*[M+Na]+*	1	4-Cyanoindole	C9H6N2	Miscellaneous/Drugs	Unlikely
	62	1.6E-04	1.5	438.083	-	439.090	*[M-H]-*	0	Hydroxytinidazole glucuronide	C14H21N3O11S	Miscellaneous/Drugs	Unlikely
	24	2.8E-04	1.5	460.065	-	415.066	*[M+FA-H]-*	0	Cephaloridine	C19H17N3O4S2	Miscellaneous/Drugs	Unlikely
	131	2.7E-05	2.2	275.533	+	527.071	*[M+H+Na]2+*	2	Indoxacarb	C22H17ClF3N3O7	Miscellaneous/Oxazines	Unlikely
	17	8.0E-04	1.5	424.086	-	425.086	*[M-H]-*	16	5,12-Dihydroxanthommatin	C20H15N3O8	Miscellaneous/Oxazines	Unlikely
	15	4.0E-04	3.8	572.058	-	527.071	*[M+FA-H]-*	18	Indoxacarb	C22H17ClF3N3O7	Miscellaneous/Oxazines	Unlikely
	64	5.0E-03	1.6	572.158	+	571.160	*[M+H]+*	16	Cyclochlorotine	C24H31Cl2N5O7	Miscellaneous/Peptides	
	149	5.9E-04	3.8	535.028	+	534.029	*[M+H]+*	15	UDP-L-Ara4O	C14H20N2O16P2	Nucleotides	
	720	1.8E-05	2.0	827.233	+	804.258	*[M+Na]+*	17	5-Formyl-5,6,7,8-tetrahydromethanopterin	C31H45N6O17P	Organic acid	
	381	2.7E-04	2.0	797.222	+	774.247	*[M+Na]+*	18	7,8-Dihydromethanopterin	C30H43N6O16P	Organic acid	
	159	2.8E-04	4.7	538.029	+	515.045	*[M+Na]+*	10	3-(ADP)-glycerate	C13H19N5O13P2	Organic acid derivatives	
	150	4.1E-06	2.2	149.011	-	150.016	*[M-H]-*	15	Tartaric acid	C4H6O6	Organic acids	
	33	1.6E-05	2.1	179.022	-	134.022	*[M+FA-H]-*	14	Malic acid	C4H6O5	Organic acids	
	26	1.8E-04	2.2	150.015	-	169.038	*[M-H2O-H]-*	28	Dihydrodipicolinic acid	C7H7NO4	Organic acids	
	24	4.3E-04	1.5	408.072	-	387.099	*[M+Na-2H]-*	3	Pyraclostrobin	C19H18ClN3O4	Organic acids/Carbamates	Unlikely
	368	4.6E-06	2.1	194.046	-	195.053	*[M-H]-*	1	N-acetyl-4-aminosalicylic acid	C9H9NO4	Phenylpropanoids/SA derivatives	
	189	1.1E-05	1.5	194.045	-	195.053	*[M-H]-*	2	N-acetyl-4-aminosalicylic acid	C9H9NO4	Phenylpropanoids/SA derivatives	
	16	1.5E-04	2.1	195.049	-	214.063	*[M-H2O-H]-*	20	Phenyl salicylate	C13H10O3	Phenylpropanoids/SA derivatives	
	16	4.9E-04	3.8	542.061	-	521.081	*[M+Na-2H]-*	9	4-(Cytidine 5\'-diphospho)-2-C-methyl-D-erythritol	C14H25N3O14P2	Sugar alcohols	
	24	4.2E-03	3.1	671.291	-	650.330	*[M+Na-2H]-*	20	Thapsigargin	C34H50O12	Terpenoids	Unlikely
	7523	1.3E-06	2.2	267.546	+						Unknown	
	518	1.1E-06	2.2	259.557	+						Unknown	
	329	4.5E-06	2.2	268.558	+						Unknown	
	241	5.8E-06	2.2	274.563	+						Unknown	
	168	7.3E-05	2.0	391.588	+						Unknown	
	148	8.8E-04	2.0	607.672	+						Unknown	
	145	1.1E-05	2.2	260.556	+						Unknown	
	106	5.2E-05	2.0	630.666	+						Unknown	
	96	1.8E-05	2.2	282.550	+						Unknown	
	93	2.0E-03	2.0	592.672	+						Unknown	
	87	5.2E-03	2.2	342.578	+						Unknown	
	66	2.3E-05	2.2	350.566	+						Unknown	
	59	8.1E-05	2.0	376.582	+						Unknown	
	58	3.0E-05	2.7	669.273	-						Unknown	
	35	1.0E-04	2.7	670.279	-						Unknown	
	21	1.3E-03	1.5	798.190	-						Unknown	
	17	9.0E-04	2.9	1067.243	-						Unknown	
	15	3.5E-06	2.2	1179.206	-						Unknown	
	14	1.6E-03	1.5	1111.044	-						Unknown	

SOIL-ENRICHED

Sample	FC	*P* value[2]	RT (min)[3]	Detected *m/z*[3]	Ion mode[3]	Predicted mass[4]	*Adduct*[4]	Δppm[4]	Putative Compound[4]	Predicted Formula[4]	Putative Pathway[5]	Soil- or plant-derived[4]
	0.28	1.8E-03	8.7	309.175	+	286.189	*[M+Na]+*	11	N-Acetyl-leucyl-leucine	C14H26N2O4	Amino acids	
	0.12	1.4E-04	1.2	644.697	-	599.686	*[M+FA-H]-*	20	Calcium trimetaphosphate	Ca3H6O18P6	Calcium ions	
	0.43	6.3E-03	8.6	171.149	+	170.142	*[M+H]+*	1	3-Acrylamidopropyl trimethylammonium	C9H18N2O	Inorganic compounds	Unlikely
	0.04	1.6E-06	6.6	783.518	+	782.510	*[M+H]+*	1	Phosphatidylglycerol	C43H75O10P	Lipids	
	0.19	1.4E-03	6.4	607.394	+	624.400	*[M+H-H2O]+*	5	Phosphatidylglycerol	C31H61O10P	Lipids	
	0.22	4.1E-03	5.5	521.315	+	498.332	*[M+Na]+*	12	Phosphatidylglycerol	C24H51O8P	Lipids	
	0.23	3.2E-03	6.1	588.422	+	565.447	*[M+Na]+*	24	Phosphatidylcholine	C30H64NO6P	Lipids	
	0.24	4.7E-03	6.5	382.315	+	417.324	*[M+H-2H2O]+*	9	N-arachidonoyl leucine	C26H43NO3	Lipids	
	0.25	4.5E-03	5.2	483.292	+	482.285	*[M+H]+*	1	Vitamin D3 derivative (2716)	C30H42O3S	Lipids	Unlikely
	0.26	2.0E-03	6.8	741.523	+	740.536	*[M+H]+*	26	Phosphatidic acid	C42H77O8P	Lipids	
	0.26	2.3E-03	6.6	649.450	+	666.447	*[M+H-H2O]+*	8	Phosphatidylglycerol	C34H67O10P	Lipids	
	0.28	8.3E-03	6.6	693.405	+	670.406	*[M+Na]+*	14	Phosphatidylglycerol	C32H63O12P	Lipids	
	0.29	9.8E-03	6.7	566.452	+	565.447	*[M+H]+*	4	Phosphatidylcholine	C30H64NO6P	Lipids	
	0.30	1.2E-04	8.4	418.271	+	395.282	*[M+Na]+*	1	N-(3-hydroxyphenyl)-Arachidonoyl amide	C26H37NO2	Lipids	
	0.33	9.1E-03	6.9	652.440	+	669.437	*[M+H-H2O]+*	8	Phosphatidylethanolamine	C36H64NO8P	Lipids	
	0.33	7.5E-03	5.3	737.512	+	736.504	*[M+H]+*	0	Phosphatidic acid	C42H73O8P	Lipids	
	0.34	1.8E-04	4.1	405.199	+	382.212	*[M+Na]+*	5	Phosphatidic acid	C17H35O7P	Lipids	
	0.36	7.9E-03	6.0	660.481	+	659.468	*[M+H]+*	9	Diacylglycerol	C43H63D5O5	Lipids	
	0.36	7.0E-03	6.1	629.418	+	628.410	*[M+H]+*	0	Phosphatidic acid	C34H61O8P	Lipids	
	0.37	2.1E-03	8.5	662.465	+	661.459	*[M+H]+*	1	2-carboxy-2-amino-3-O-(13\'-methyltetradecanoyl)-4-hydroxy-17-methyloctadec-5-ene-1-sulfonic acid	C35H67NO8S	Lipids	Unlikely
	0.37	6.4E-03	8.8	1010.743	+	987.759	*[M+Na]+*	4	Galabiosylceramide	C55H105NO13	Lipids	
	0.38	1.1E-03	6.4	648.561	+	647.561	*[M+H]+*	11	Diacylglycerol	C41H75D5O5	Lipids	
	0.38	2.6E-04	8.8	929.775	+	928.752	*[M+H]+*	16	Triacylglycerol	C61H100O6	Lipids	
	0.38	3.1E-03	8.6	518.351	+	495.369	*[M+Na]+*	13	Phosphatidylcholine	C25H54NO6P	Lipids	
	0.40	9.9E-03	5.7	420.283	+	437.291	*[M+H-H2O]+*	11	Phosphatidylethanolamine	C21H44NO6P	Lipids	
	0.40	1.1E-03	8.7	766.578	+	765.567	*[M+H]+*	4	Phosphatidylcholine	C44H80NO7P	Lipids	
	0.41	4.8E-03	7.7	684.483	+	683.474	*[M+H]+*	2	Phosphatidylglycerol	C34H67O10P	Lipids	
	0.41	8.8E-03	6.4	1032.690	+	1009.707	*[M+Na]+*	5	3-O-acetyl-sphingosine-2,3,4,6-tetra-O-acetyl-GalCeramide	C56H99NO14	Lipids	
	0.42	5.8E-03	8.3	764.596	+	799.609	*[M+H-2H2O]+*	0	Phosphatidylethanolamine	C45H86NO8P	Lipids	
	0.44	4.1E-03	6.8	528.394	+	563.395	*[M+H-2H2O]+*	22	Phosphatidylcholine	C29H58NO7P	Lipids	
	0.44	5.7E-03	6.6	902.749	+	901.750	*[M+H]+*	9	Phosphatidylcholine	C52H104NO8P	Lipids	
	0.06	1.4E-04	1.3	691.689	-	692.705	*[M-H]-*	12	Hydroxyphthioceranic acid (C46)	C46H92O3	Lipids	Unlikely
	0.09	9.6E-04	6.3	1154.726	-	1133.780	*[M+Na-2H]-*	24	Galalpha1-4Galbeta1-4Glcbeta-Ceramide	C60H111NO18	Lipids	
	0.20	6.0E-04	7.5	897.675	-	898.705	*[M-H]-*	25	Triacylglycerol	C59H94O6	Lipids	
	0.20	3.4E-03	2.3	719.725	-	720.736	*[M-H]-*	4	Hydroxyphthioceranic acid (C48)	C48H96O3	Lipids	Unlikely
	0.20	4.7E-03	6.8	1007.732	-	1026.768	*[M-H2O-H]-*	17	Triacylglycerol	C69H102O6	Lipids	
	0.22	2.2E-03	8.9	312.836	-	291.860	*[M+Na-2H]-*	4	Ethyl 2-Bromo-2-iodoacetate	C4H6BrIO2	Lipids	Unlikely
	0.23	6.1E-03	8.7	693.481	-	712.504	*[M-H2O-H]-*	6	Phosphatidic acid	C40H73O8P	Lipids	
	0.24	9.1E-03	8.5	647.673	-	648.678	*[M-H]-*	2	Octacosyl-palmitate	C44H88O2	Lipids	
	0.24	1.0E-03	1.2	937.711	-	938.736	*[M-H]-*	19	Triacylglycerol	C62H98O6	Lipids	
	0.42	3.1E-03	7.6	520.315	+	519.294	*[M+H]+*	25	Vignatic acid B	C27H41N3O7	Miscellaneous	Unlikely
	0.44	6.5E-03	6.9	493.453	+	510.455	*[M+H-H2O]+*	1	Nb-Lignoceroyltryptamine	C34H58N2O	Miscellaneous/Alkylindoles	Unlikely
	0.03	3.5E-05	1.1	208.936	-	209.941	*[M-H]-*	13	2,4,6-Trichloroanisole	C7H5Cl3O	Miscellaneous/Aromatics	Unlikely
	0.15	5.5E-04	1.6	596.816	-	597.811	*[M-H]-*	20	Iopodic acid	C12H13I3N2O2	Miscellaneous/Aromatics	Unlikely
	0.22	5.1E-04	1.6	244.894	-	223.920	*[M+Na-2H]-*	3	2,3,6-Trichlorobenzoic acid	C7H3Cl3O2	Miscellaneous/Aromatics	Unlikely
	0.31	8.1E-04	8.6	476.308	+	475.301	*[M+H]+*	0	Netilmicin	C21H41N5O7	Miscellaneous/Drugs	Unlikely
	0.35	2.0E-03	8.6	319.202	+	336.205	*[M+H-H2O]+*	0	Acebutolol	C18H28N2O4	Miscellaneous/Drugs	Unlikely
	0.44	2.7E-03	4.5	450.171	+	427.180	*[M+Na]+*	5	5-hydroxyfluvastatin	C24H26FNO5	Miscellaneous/Drugs	Unlikely
	0.17	4.2E-03	7.7	493.102	-	448.103	*[M+FA-H]-*	1	N,N\'-(((4-methyl-1,3-phenylene)bis(azanediyl))bis(carbonothioyl))dibenzamide	C23H20N4O2S2	Miscellaneous/Drugs	Unlikely
	0.24	4.7E-03	8.8	1015.228	-	994.249	*[M+Na-2H]-*	4	Calcein AM	C46H46N2O23	Miscellaneous/Fluoresceins	Unlikely
Control soil	0.00	6.0E-05	0.8	209.949	-	188.975	*[M+Na-2H]-*	2	2,6-Dichlorobenzamide	C7H5Cl2NO	Miscellaneous/Insecticides	Unlikely
	0.01	2.5E-05	1.0	225.926	-	244.950	*[M-H2O-H]-*	21	7-N,N-Dimethylamino-1,2,3,4,5-pentathiocyclooctane	C5H11NS5	Miscellaneous/Insecticides	Unlikely
	0.23	9.5E-03	6.5	633.269	+	668.283	*[M+H-2H2O]+*	2	Filicin	C36H44O12	Miscellaneous/Ketones	Unlikely
	0.29	6.3E-04	6.4	939.584	+	974.593	*[M+H-2H2O]+*	4	Megalomicin C2	C49H86N2O17	Miscellaneous/Macrolides	Unlikely
	0.02	7.0E-03	1.1	306.938	-	325.964	*[M-H2O-H]-*	25	Butonate	C8H14Cl3O5P	Miscellaneous/Organophosphonates	Unlikely
	0.03	3.3E-04	0.9	338.895	-	317.928	*[M+Na-2H]-*	24	Tiludronic acid	C7H9ClO6P2S	Miscellaneous/Organophosphonates	Unlikely
	0.43	7.1E-03	6.9	696.468	+	731.483	*[M+H-2H2O]+*	3	Microcolin B	C39H65N5O8	Miscellaneous/Peptides	Unlikely
	0.21	3.4E-03	9.0	887.232	-	888.232	*[M-H]-*	7	Cyanidin 3-[6-(6-p-coumarylglucosyl)-2-xylosylgalactoside]	C41H44O22	Miscellaneous/Pigments	
	0.29	5.7E-04	8.8	1031.546	+	1066.556	*[M+H-2H2O]+*	2	3-O-(Glcb1-2(Xylb1-3)Glcb1-4Galb)-(25R)-5alpha-spirostan-3beta-ol	C51H86O23	Miscellaneous/Polysaccharides	
	0.44	3.6E-03	8.7	361.237	+	360.230	*[M+H]+*	0	11beta-17-Dihydroxy-6alpha-methylpregn-4-ene-3,20-dione	C22H32O4	Miscellaneous/Steroids	Unlikely
	0.14	2.5E-03	8.9	859.065	-	878.098	*[M-H2O-H]-*	16	Mn(III) tetrakis(4-benzoic acid) porphyrin chloride	C48H28ClMnN4O8	Miscellaneous/Tetrapyrroles	
	0.32	2.4E-03	6.6	500.401	+	535.403	*[M+H-2H2O]+*	22	25-dihydroxycholecalciferol	C35H53NO3	Miscellaneous/Vitamins	Unlikely
	0.40	6.5E-03	6.6	541.358	+	576.366	*[M+H-2H2O]+*	8	1-Hydroxyvitamin D3 3-D-glucopyranoside	C33H52O8	Miscellaneous/Vitamins	Unlikely
	0.16	5.1E-03	5.6	723.070	-	724.065	*[M-H]-*	16	Adenylated molybdopterin	C20H26N10O12P2S2	Nucleotides	
	0.23	4.9E-03	6.7	666.049	-	621.061	*[M+FA-H]-*	14	UDP-N-acetyl-D-galactosaminuronic acid	C17H25N3O18P2	Nucleotides	
	0.21	5.3E-03	1.3	432.844	-	433.851	*[M-H]-*	0	3-(3,5-Diiodo-4-hydroxyphenyl)lactate	C9H8I2O4	Organic acids	
	0.40	1.1E-03	6.4	491.189	+	468.200	*[M+Na]+*	0	Paucin	C23H32O10	Terpenoids	Unlikely
	0.15	1.5E-04	3.9	443.132	-	444.141	*[M-H]-*	3	1-O-[2-(L-Cysteinamido)-2-deoxy-alpha-D-glucopyranosyl]-1D-myo-inositol	C15H28N2O11S	Sugar alcohols	
	0.29	3.3E-03	5.4	383.217	+						Unknown	
	0.30	7.2E-03	1.1	217.987	+						Unknown	
	0.39	4.1E-03	6.4	1049.598	+						Unknown	
	0.40	3.3E-03	6.2	664.622	+						Unknown	
	0.42	9.8E-03	6.5	988.586	+						Unknown	
	0.43	4.7E-04	8.8	768.714	+						Unknown	
	0.45	4.8E-03	6.4	636.900	+						Unknown	
	0.02	9.3E-04	1.1	192.959	-						Unknown	
	0.03	1.8E-03	1.1	322.917	-						Unknown	
	0.06	3.2E-05	1.3	747.725	-						Unknown	
	0.08	1.1E-04	1.3	910.685	-						Unknown	
	0.11	2.9E-03	1.9	629.726	-						Unknown	
	0.14	7.9E-03	2.0	434.817	-						Unknown	
	0.14	1.6E-03	1.9	610.816	-						Unknown	
	0.15	2.1E-03	1.5	319.839	-						Unknown	
	0.17	1.1E-03	1.1	311.877	-						Unknown	
	0.17	2.8E-03	9.1	557.078	-						Unknown	
	0.18	3.8E-04	1.3	328.864	-						Unknown	
	0.18	3.3E-04	1.5	255.929	-						Unknown	
	0.19	1.2E-03	2.1	587.806	-						Unknown	
	0.19	1.7E-03	1.9	568.840	-						Unknown	
	0.19	4.8E-03	1.1	1021.633	-						Unknown	
	0.20	9.8E-03	1.3	782.734	-						Unknown	
	0.20	9.4E-03	1.1	295.900	-						Unknown	
	0.21	3.3E-04	8.9	1098.495	-						Unknown	
	0.21	1.9E-03	2.3	320.835	-						Unknown	
	0.23	3.9E-03	1.3	575.804	-						Unknown	
	0.23	6.0E-04	2.7	652.759	-						Unknown	
	0.24	2.0E-03	1.2	574.803	-						Unknown	
	0.24	1.2E-04	8.9	1178.917	-						Unknown	
	0.24	4.7E-03	9.3	566.779	-						Unknown	
	0.24	7.0E-03	3.9	261.886	-						Unknown	

SUPPLEMENTAL METHODS

DNA extraction, 16S rRNA gene sequencing and analysis.

For microbial profiling, eight additional growth tubes were set up, as described in the Experimental Procedures, but were not used for the collection of chemicals. Four of these tubes contained one Arabidopsis plant and four contained only growth substrate. After 5 weeks, plants were sampled by carefully loosening the soil around the edges of the growth tube, pulling up the roots and removing excess soil by shaking. Soil samples were also taken from the tubes without plants, using a sterile spatula and avoiding surface material. DNA was extracted from the resulting samples consisting of either roots covered in their closely adhering soil (root plus rhizosphere samples), or only soil (control soil), using a PowerSoil DNA extraction kit (MoBio Laboratories Inc., Carlsbad, CA, USA) according to the manufacturer's instructions. Partial prokaryotic 16S rRNA genes were amplified from this extract, using primers 799F and 1193R (Chelius and Triplett, 2001; Bodenhausen *et al.*, 2013), which were modified to include the Illumina overhang adapter nucleotide sequences (adapters shown in normal typeface, locus specific primers in bold letter font):

799F:

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**AACMGGATTAGATACCCKG**

1193R:

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**ACGTCATCCCCACCTTCC**.

PCRs were carried out, using 0.4 U of KAPA HiFi HotStart DNA polymerase (Kapa Biosystems Ltd, London, UK) on 2 µL of DNA extract in the presence of 2.5 mM

$MgCl_2$, 1.2 mM deoxynucleoside triphosphates (dNTPs), 0.2 µM of each primer, and the manufacturer's reaction buffer in a total reaction volume of 20 µL (PCR conditions: 95 °C for 3 min; 25 cycles at 95 °C for 30 s, 58 °C for 30 s and 72 °C for 30 s; and 72 °C for 5 min). To reduce PCR bias, the PCR was performed in triplicate and amplicons were pooled. A sequencing library was constructed by cleaning up pooled PCR products, using AMPure XP beads (Beckman Coulter (UK) Ltd, High Wycomb, UK), followed by attachment of dual indices and Illumina sequencing adapters, using the Nextera XT Index Kit (Illumina Inc. Essex UK) and following the manufacturer's instructions. The indexed PCR products were cleaned using AMPure XP beads and sequencing was performed using a paired end 2 x 250 bp cycle kit v2 on a MiSeq machine running v2 chemistry (Illumina Inc, at The Genome Analysis Centre, Norwich, UK). Raw sequencing data were deposited in the European Nucleotide Archive (ENA) under accession number PRJEB17782. Sequences were analysed by USEARCH (Edgar, 2010) and Qiime pipelines (Caporaso *et al.*, 2010a). Sequences were filtered using USEARCH, retaining those with a maxEE value of 1 (equivalent to 1 in 1,000 errors) and 251 bp long. Chimeras were detected using UCHIME (Edgar *et al.*, 2011), using both reference based and *de novo* detection methods. After selection of OTUs by USEARCH (97% similarity), the representative sequences were aligned to the Greengenes 13_8 core reference alignment (DeSantis *et al.* 2006) using PyNAST (Caporaso *et al.*, 2010b). All other steps leading to the generation of OTU abundance tables were performed using Qiime. All statistical analyses of community data were performed using the R programming language (R Development Core Team, 2016; https://www.R-project.org/) and with the packages phyloseq (McMurdie and Holmes, 2013), vegan (https://github.com/vegandevs/vegan) and DESeq2 (Love *et al.*, 2014).

Quantification of plant tissue damage by electrolytes leakage.

Tissue damage by the acidified extraction solutions was quantified by conductivity of cell electrolytes leakage, as described previously (Pétriacq *et al.*, 2016a, 2016b). For Arabidopsis, roots were collected from plants cultivated in half strength Murashige-Skoog, solidified with 0.8% Phytagel (Sigma-Aldrich, UK) and adjusted to pH 5.8. Root replicates consisted of one intact root system from 2-week-old plants, which was removed carefully from the agar medium. For maize, roots were collected from surface sterilised seeds, germinated and grown for five days on wetted filter paper in sealed petri-dishes. Tissues were incubated for 1 min in 10 mL of different acidified extraction solutions, containing 0.05% formic acid (v/v) and 0%, 50% or 95% methanol (v/v). As a negative control, tissues were incubated in double-distilled sterile water. As a positive control for cell damage, tissues were wounded prior to extraction solution incubation by cutting roots into 10 pieces with a razor blade. Directly after incubation, tissues were rinsed in double-distilled sterile water, then transferred into glass bottles containing 5 mL of double-distilled sterile water, and subsequently agitated at room temperature for 2 hours on an orbital shaker (200 rpm). Conductivity was then measured in the balanced solution, using a CMD 500 WPA conductivity meter. Subsequently, all samples were boiled for 30 min and re-measured for conductivity of lysed tissue. Cell damage was expressed as the average level of conductivity, relative to the maximum level of conductivity after tissue lysis (set at 100%). Each treatment was based on 4 replicated samples (n = 4). Data were analysed in IBM SPSS (v. 22), using a Welch's F test for ranked data, followed by Games-Howell tests to assess individual differences ($P < 0.05$). The experiments were repeated three times with similar results.

Analysis of microscopic root cell damage by extraction solutions.

Transgenic Arabidopsis plants (Col-0) expressing the *35S:IBI1:YFP* construct, encoding the cytoplasmic aspartyl-tRNA synthatase IBI1 with a C-terminal fusion to Yellow Fluorescent Protein (Luna *et al.*, 2014; *35S::IBI1:YFP*), were cultivated for two weeks (8.5/15.5 h light/dark at 21/19 °C, 120 µmol m^{-2} s^{-1} photons, 70% relative humidity) on half strength Murashige-Skoog agar plates, solidified with 0.8% Phytagel (Sigma-Aldrich, UK) and adjusted to pH 5.8. Roots were extracted carefully form agar plates, and incubated for 1 min in the acidified MeOH-containing extraction solutions (0, 50, 95% MeOH with 0.05% formic acid, v/v). As negative and positive controls for cell damage, roots were incubated for 1 min in double-distilled sterile water, or for 15 min in 100% MeOH, respectively. After incubation, roots were rinsed in double-distilled sterile water prior to epi-fluorescence microscopy analysis. Fluorescence was observed using an epi-fluorescence microscope (Olympus BX51, excitation filter BP 470/40 nm, barrier filter BP 525/50 nm). For each treatment, root systems form 12 different plants were analysed and photos were taken of representative samples. The experiment was performed four times with similar results.

Analysis of impacts on soil and rhizosphere bacteria by extraction solutions.

Tubes (30 mL; n = 3) containing the sand:compost mixture (9:1 v/v) with or without 5-week-old Arabidopsis were left untreated, or were bacterized by syringe injection with 5 mL of 10 mM $MgSO_4$, containing either YFP-expressing *P. simiae* WCS417r (Zamioudis *et al.*, 2014), or rifampicin-resistant *B. subtilis* 168 (Yi *et al.*, 2016), to a final density of 10^7 colony CFU g^{-1}. After 48 h, tubes were flushed with extraction

solution (as detailed in Experimental procedures). Additional tubes were flushed with double-distilled sterile water (control), or 95% MeOH and left for 45 min (positive control for cell damage). Subsequently, 1 g of either control soil (without roots), or Arabidopsis roots plus adhering rhizosphere soil, was sampled from the tubes, suspended for 5 min into 50 mL of 10 mM $MgSO_4$, and centrifuged (5 min, 3,500 *g*). Pellets were re-suspended in 1 mL of 10 mM $MgSO_4$, and subjected to dilution plating onto Luria Broth (LB) agar medium supplemented with 5 µg mL^{-1} of the anti-fungal cycloheximide. For testing impacts on culturable soil bacteria, LB agar contained no further antibiotics; for testing impacts on *P. simiae* WCS417r and *B. subtilis* 168, plates were supplemented with 5 µg mL^{-1} tetracycline + 25 µg mL^{-1} rifampicin and 50 µg mL^{-1} rifampicin, respectively. Plates were kept for 24 - 48 h at 28 °C. Each biologically replicated sample was plated four times, after which the technical replicates were averaged to minimize confounding effects of heterogeneity in suspended pellets. Experiments were repeated twice with comparable results.

UPLC-Q-TOF mass spectrometry.

Untargeted metabolic profiling by UPLC-Q-TOF mass spectrometry (MS) was performed as described previously (Pétriacq *et al.*, 2016b) using an ACQUITY ultra-high-pressure liquid chromatography (UPLC) system coupled to a SYNAPT G2 Q-TOF mass spectrometer with an electrospray (ESI) ionization source (Waters, UK). The system was controlled by MassLynx v. 4.1 software (Waters). Chromatographic separation of samples was carried out at a flow rate of 0.4 mL min^{-1} using an ACQUITY UPLC BEH C18 column (2.1 × 50 mm, 1.7 µm, Waters) coupled to a C18 VanGuard pre-column (2.1 x 5 mm, 1.7 µm, Waters). The mobile phase consisted of solvent A (0.05 %, formic acid v/v, in water) and solvent B (0.05 % formic acid v/v in

acetonitrile) with the following gradient: 0 – 3 min 5 – 35 % B, 3 – 6 min 35 – 100 % B, holding at 100 % B for 2 min, 8 – 10 min, 100 – 5 % B. The column was maintained at 45 °C and the injection volume was 10 µL. Between each condition, a blank was injected with 50% methanol (v/v) to clean the column. Sample runs in negative and positive ionization mode (ESI^- and ESI^+) were separated by two consecutive injections with 50% methanol (v/v) to allow stabilization of the ionization modes. An ACQUITY PDA detector (Waters) was used to monitor the UV trace (range 205 – 400 nm, sampling rate 40 points s^{-1}, resolution 1.2 nm). MS detection of ions was operated in sensitivity mode by SYNAPT G2 (50 - 1200 Da, scan time = 0.2 s) in both ESI^- and ESI^+, using a full MS scan (*i.e.* no collision energy) and applying the MS^E function with a ramp in the transfer cell in elevated energy mode (5 to 45 eV). The following conditions were applied for ESI^- (capillary voltage - 3 kV, sampling cone voltage - 25 V, extraction cone voltage -4.5 V, source temperature 120 °C, desolvation temperature 350 °C, desolvation gas flow 800 L h^{-1}, cone gas flow 60 L h^{-1}), and for ESI^+ (capillary voltage 3.5 kV, sampling cone voltage 25 V, extraction cone voltage 4.5 V, source temperature 120 °C, desolvation temperature 350 °C, desolvation gas flow 800 L h^{-1}, cone gas flow 60 L h^{-1}). Prior to analyses, the Q-TOF was calibrated by infusing a sodium formate solution. Accurate mass detection was ensured by infusing the internal lockmass reference peptide leucine enkephalin during each run.

Statistical analysis of MS data.

Prior to multivariate analyses, the XCMS R package (v. 3.1.3; Smith *et al.*, 2006) was used to align and integrate raw UPLC-Q-TOF peaks, to correct for total ion current (TIC) and median fold-change. All statistical analyses were performed with

median-normalized, cube-root-transformed and Pareto-scaled data, using MetaboAnalyst software (v. 3.0, http://www.metaboanalyst.ca; Xia *et al.*, 2015). Three-dimensional principal component analyses (3D-PCA) were based on the first three principal components (PCs) that explain most variation of the dataset. Supervised partial least square discriminant analyses (PLS-DAs) were conducted to quantify discriminative power between soil types and extraction solutions. PLS-DA models were validated by correlation (R^2) and predictability (Q^2) parameters for both ESI^+ and ESI^- modes ($R^2 > 0.94$ and $Q^2 > 0.59$, respectively). Numbers of total ions were obtained from XCMS output datasets. To quantify metabolic differences between rhizosphere and control soil, volcano plots were constructed at a statistically significant threshold of $P < 0.01$ (Welch's *t*-test) and a fold-difference threshold of 2, using MetaboAnalyst (v. 3.0, http://www.metaboanalyst.ca; Xia *et al.*, 2015). To obtain putative identities of a combined set of ions from all three extraction solutions that are either enriched in Arabidopsis soil, or its corresponding control soil, the top-20 ranking ions from each volcano plot were selected by fold-change (above 2 or below -2) and *P* value, followed by an ANOVA ($P < 0.01$) for statistical differences between all soil/extraction solution combinations, using a Benjamini-Hochberg false discovery rate (FDR) correction for multiple hypothesis testing (Hochberg and Benjamini, 1990). To obtain putative identities from the 50% MeOH extraction solution that are either enriched in the maize rhizosphere, or corresponding control soil, the top-50 ranking ions from each volcano plot (ESI^+ and ESI^-) were selected. For both cultivation systems, ions were corrected for adducts and/or isotopes, using MarVis (v. 1.0; http://marvis.gobics.de; tolerance: *m*/z = 0.1 Da, RT = 10 s; Kaever *et al.*, 2012). Putative metabolites were identified by referencing the final set of detected accurate *m/z* values against publicly available

chemical databases using METLIN, PubChem, MassBank, Lipid Bank, ChemSpider, Kegg, AraCyc and MetaCyc database, as documented in several studies (Kaever *et al.*, 2009; Kaever *et al.*, 2012; Gamir *et al.*, 2014a, 2014b; Pastor *et al.*, 2014; Pétriacq *et al.*, 2016a, 2016b). METLIN (https://metlin.scripps.edu) was used to determine accuracy and chemical formulae for the putative compounds. PubChem (https://pubchem.ncbi.nlm.nih.gov/) was used to check the predicted pathway classification. In cases where multiple ions could be annotated to the same putative metabolite (due to different adducts and ionization modes; Tables S1 and S2), they were counted additively to the metabolite class presented in the pie-charts of Figures 6 and 7.

Experimental system for profiling distant rhizosphere fractions.

To investigate whether the chemical influence of the rhizosphere extends beyond soil that is closely associated with roots, maize plants were grown in mesh bags, which allowed for physical separation of root systems from the distal soil in the periphery of the growth tube. Bags were constructed from a nylon mesh (35 μm diameter holes), folded over and heat sealed to produce bags (6 cm x 11 cm, approximate diameter when filled = 3.5 cm). These bags were filled with 85 cm^3 of a mixture of 75:25 (v/v) agricultural soil:perlite, as used previously for maize experiments. Each mesh bag was placed into the centre of the 150-mL plastic tube (11 cm high and 5 cm diameter; Starlab) with a miracloth sheet covering the bottom hole of the tube. Seventy cm^3 of the same soil substrate was used to fill the peripheral space between the mesh bag and tube wall. A total of 24 pots were set up in this manner. Pre-germinated maize seeds (W22) were planted into the bags of 12 tubes. The other 12 tubes were left unplanted to serve as plant-free controls. All

tubes were wrapped in foil and covered with black plastic beads to prevent algal growth. Sixty mL of distilled water was added to each tube to saturate the soil with water before seeds were planted, after which all pots were transferred to a growth cabinet with the following conditions: 16/8 h light/dark with an average light intensity of 140 µmol m^{-2} s^{-1} at the top of the collection tubes, a relative humidity of 60%, and a constant temperature of 20 °C. Soil metabolites were extracted from the different soil fractions after 24 days of growth. To collect metabolites from the distal soil fractions, black beads were removed, and mesh bags were carefully removed from half of the pots (6 tubes with maize and 6 without). The remaining distal soil in the tube (*i.e.* the soil that had been outside the bag) was tapped to the bottom of the 150-mL tubes and extracted by applying 25 mL of acidified 50% (v/v) MeOH to the top of the soil. The solution was flushed through the tube by applying pressure for 1 min through a modified 150-mL tube lid containing a 50-mL syringe, until ~10 mL of solution was collected from the base of the tube into new 50-mL tubes. To collect metabolites from the whole soil fractions, plastic beads were removed from the remaining 12 pots and maize shoots were cut from the 6 that contained plants. Subsequently, 50 mL of acidified 50% MeOH (v/v) was applied to the top of the tube, keeping the mesh bags in place. The solution was flushed through by applying pressure for 1 min using the modified lid, as previously described, resulting in a least 10 mL of collection volume at the base of the tube. All extracts were centrifuged to pellet soil residues (5 min, 3,500 *g*), after which 8 mL of supernatant were transferred into a new 15-mL centrifuge tube and flash-frozen in liquid nitrogen. All samples were freeze-dried for two days, after which dried material was re-suspended in 500 µL of methanol: water: formic acid (50: 49.9: 0.1, v/v/v), sonicated at 4 °C for 20 min, vortexed, transferred into 2-mL microtubes and centrifuged (15

min, 14,000 *g*, 4 °C). Final supernatants (180 µL) were transferred into glass vials containing a glass insert before injection through the UPLC system. UPLC-Q-TOF analysis was conducted in ESI^- as described above. For DIMBOA targeted quantitation, a purified and NMR-verified standard (Ahmad *et al.*, 2011) was run alongside the samples. Metabolomics data were normalised for soil amount (n = 6), and subsequent analysis performed with MetaboAnalyst (v. 3.0), as described above (*i.e.* median normalisation, cube-root transformation, Pareto scaling).