environmental microbiology

Environmental Microbiology (2021) **23**(8), 4706–4725 doi:10.1111/1462-2920.15673

Meta-omic evaluation of bacterial microbial community structure and activity for the environmental assessment of soils: overcoming protein extraction pitfalls

Ana M. Herruzo-Ruiz,[1] Carlos A. Fuentes-Almagro,[2] José M. Jiménez-Pastor,[1] Víctor M. Pérez-Rosa,[1] Julián Blasco,[3] Carmen Michán[1] and José Alhama[1*]

[1]*Departamento de Bioquímica y Biología Molecular, Campus de Excelencia Internacional Agroalimentario CeiA3, Universidad de Córdoba, Campus de Rabanales, Edificio Severo Ochoa, Córdoba, E-14071, Spain.*

[2]*Proteomics Unit, SCAI, University of Córdoba, Córdoba, Spain.*

[3]*Department of Ecology and Coastal Management, ICMAN-CSIC, Campus Rio San Pedro, Puerto Real, E-11510, Spain.*

Summary

Microorganisms play unique, essential and integral roles in the biosphere. This work aims to assess the utility of soil's metaomics for environmental diagnosis. Doñana National Park (DNP) was selected as a natural lab since it contains a strictly protected core that is surrounded by numerous threats of pollution. Culture-independent high-throughput molecular tools were used to evaluate the alterations of the global structure and metabolic activities of the microbiome. 16S rRNA sequencing shows lower bacterial abundance and diversity in areas historically exposed to contamination that surround DNP. For metaproteomics, an innovative post-alkaline protein extraction protocol was developed. After NaOH treatment, successive washing with Tris–HCl buffer supplemented with glycerol was essential to eliminate interferences. Starting from soils with different physicochemical characteristics, the method renders proteins with a remarkable resolution on SDS-PAGE gels. The proteins extracted were analysed by using an in-house database constructed from the rRNA data. LC–MS/MS analysis identified 2182 non-redundant proteins with 135 showing significant differences in relative abundance in the soils around DNP. Relevant global biological processes were altered in response to the environmental changes, such as protective and antioxidant mechanisms, translation, folding and homeostasis of proteins, membrane transport and aerobic respiratory metabolism.

Received 19 March, 2021; revised 8 July, 2021; accepted 11 July, 2021. *For correspondence. E-mail bb2alcaj@uco.es; Tel. +34 957218082.

Introduction

Microorganisms are present in almost every environment in the biosphere where they play unique, essential and integral roles. In soils, microbes preserve homeostasis by mainly contributing to the biogeochemical cycles of the major elements (e.g. carbon, nitrogen, phosphorous and sulfur), nutrients, water and organic waste cycling, matter decomposition and soil fertility (Dominati *et al*., 2010; Starke *et al*., 2019; Margerison *et al*., 2020; Pylro *et al*., 2020). Microbial soil communities are also responsive to anthropogenic pressures (e.g. contamination, deforestation and agricultural management) and climate change-related factors (Bastida and Jehmlich, 2016; Starke *et al*., 2019). However, despite its importance, the soil environment is one of the least studied habitats. Since biological and functional diversity is crucial to maintaining ecosystems and thus for the sustainability of the planet, the study of microbial communities and their relationship with environmental changes is essential to understand soil dynamics (Bastida *et al*., 2014; Zhou *et al*., 2015; Bouchez *et al*., 2016; Wang *et al*., 2016; Starke *et al*., 2019; Margerison *et al*., 2020; Pylro *et al*., 2020). Recent advances in DNA sequencing technologies are increasingly showing that microbial soil communities are enormously diverse ranking among the most abundant, complex and diverse groups of organisms on Earth (Torsvik and Ovreas, 2002; Myrold *et al*., 2013; Delgado-Baquerizo *et al*., 2018; Harrison and Cameron, 2020). It is estimated that up to 10 billion microorganisms belonging to thousands of different species may well be in 1 g of soil (Torsvik and Ovreas, 2002). Currently, the increasing pressure from anthropogenic activities results in a loss of biodiversity which

can have major consequences on ecosystem functioning (Chapin III *et al*., 2000).

Natural environments are receiving an increasing number of contaminants largely as a result of human activities. Thus, Doñana National Park (DNP) is a relevant wildlife reserve and one of the most important natural protected areas in Europe. However, despite its great ecological value, DNP is extremely sensitive. Doñana is surrounded by areas exposed to numerous threats of pollution due to agricultural, mining and industrial undertakings (Abril *et al*., 2014; Fernández-Cisnal *et al*., 2014; García-Sevillano *et al*., 2014; Abril *et al*., 2015; Ruiz-Laguna *et al*., 2016; Michán *et al*., 2019). There is an urgent need to assess and identify early responses to pollution to prevent excessive or even irreversible damage to ecosystems.

Microbial communities have been proposed as sensitive indicator species or biosensors of environmental pollution disturbances. The ubiquitous microbiome is a frontline responder to environmental stresses with changes in its structure, diversity and functional activity (Siggins *et al*., 2012; Zarraonaindia *et al*., 2013; Franzosa *et al*., 2015; Bouchez *et al*., 2016; Czaplicki and Gunsch, 2016; Techtmann and Hazen, 2016). In systems biology, the integration of culture-independent interconnected 'meta-omic' (i.e. metagenomic, metatranscri ptomic, metaproteomic, metabolomic) approaches has evolved to provide detailed information on how microbial communities assemble and interact, their predominant metabolic activities and their responses to environmental perturbations (i.e. contamination) (Zarraonaindia *et al*., 2013; Franzosa *et al*., 2015; Aguiar-Pulido *et al*., 2016; Bouchez *et al*., 2016). Analysis of nucleic acids directly extracted from environmental samples allows us to study microbial communities without the need to cultivate them, a tremendous advantage is given that as many as 99% of the microbes present in many natural environments have neither been successfully cultivated by standard techniques nor characterized (Torsvik and Ovreas, 2002; Riesenfeld *et al*., 2004; Rajendhran and Gunasekaran, 2008; Singh *et al*., 2009; Lloyd *et al*., 2018). Therefore, soil ecosystems are to a large extent uncharted (Torsvik and Ovreas, 2002). Studies based on the direct analysis of 16S rRNA gene sequences provide information on the composition/structure/diversity of a bacterial community (Lane *et al*., 1985; Riesenfeld *et al*., 2004; Rajendhran and Gunasekaran, 2008; Zhou *et al*., 2015; Delgado-Baquerizo *et al*., 2018). However, other approaches are needed to assess and strengthen the connections with actual microbial metabolic activity. Since microbial functions are mediated by proteins and they play key roles in many soil processes, metaproteomics should be the most direct 'omics' estimator of the real microbial activity (Hettich *et al*., 2012; Myrold *et al*., 2013; Zarraonaindia *et al*., 2013; Bastida and Jehmlich, 2016; Wang *et al*., 2016; Greenfield *et al*., 2018; Harrison and Cameron, 2020). Thus, the development of a robust method for protein extraction from soil is of urgent and increasing interest (Greenfield *et al*., 2018). However, unlike the isolation of environmental DNA from soil and sediments for which different efficient and successful methods have been described and easy to use commercial kits are available (Rajendhran and Gunaseka ran, 2008), microbial protein extraction and the efficient isolation and separation of undegraded proteins are still cumbersome processes. In fact, the lack of a standard and reliable protocol for extracting proteins from complex environmental samples significantly limits the routine application of metaproteomics for these samples (Rajend hran and Gunasekaran, 2008; Chen *et al*., 2009; Siggins *et al*., 2012; Myrold *et al*., 2013; Zarraonaindia *et al*., 2013; Bastida *et al*., 2014; Bouchez *et al*., 2016; Wang *et al*., 2016; Greenfield *et al*., 2018; Starke *et al*., 2019). The need for more curated and annotated reference databases and more powerful bioinformatics software for big data analysis adds to the present difficulties in technical extraction due to the presence of inhibitory or interfering substances (e.g. humic acids, clays) in these complex matrices (Siggins *et al*., 2012; Bastida *et al*., 2014; Zhou *et al*., 2015; Bouchez *et al*., 2016; Wang *et al*., 2016).

This study aims to assess the use of bacterial microbial soil communities as biological indicators for environmental diagnosis. DNP was selected as a model natural laboratory since it contains a strictly protected (and presumably pristine) core that is surrounded by areas subjected to different threats of pollution. Culture-independent high-throughput molecular tools were used to evaluate alterations of the global structure and metabolic activities of the microbiomes in response to environmental threats. Since protein extraction from complex samples was a critical technical bottleneck in metaproteomic studies, a highly improved and efficient post-alkaline protein extraction protocol from soils was developed and successfully applied to environmental samples from the area of influence of DNP.

Results

Physicochemical properties of the soils

Soils from three different sites in DNP and its surroundings were used in this study (Fig. 1). The characteristics of the different soils are shown in Supplementary Table 2. The oxidizable organic matter and organic nitrogen were higher in the Lucio del Palacio (LDP) soil

followed by Ajolí (AJO) and finally Matochal (MAT). While the pH value is around 8 in all soils, the electrical conductivity was lower in AJO than at the other sites. The texture varied among the soils: MAT and LDP had a silty clay and silty clay loam texture respectively, while AJO was loamy sand.

Taxonomic analysis of the soils' bacterial community

A different number and diverse assortment of 16S rRNA bacterial sequences (reads) were identified in the different soils (Fig. 2A). The number of identifications was 2.4- and 5.4-fold higher in the soils from LDP (2964 reads) than in those from AJO (1236 reads) and MAT (561 reads) respectively. LDP also shows a large diversity with three main phyla (Proteobacteria, Actinobacteria and Chlorofexi) accounting for 91.7% of the reads while the remaining 8.7% belong to five minority phyla (Firmicutes, Bacteroidetes, Cyanobacteria, Gemmatimonadetes and Verrucomicrobia). A lower assortment was found in the other two soils; thus, apart from the most represented Proteobacteria and Actinobacteria phyla, only Firmicutes and Bacteroides were identified at AJO and MAT respectively. Proteobacteria was found to be the most abundant phylum in all the soils (44.2%–88.95%) closely followed by Actinobacteria in LDP (30.2%) and AJO (44.5%) but not in MAT where this last phylum is highly diminished (8.6%).

Site (code)	UTM coordinates (m)	
Lucio del Palacio (LDP)	*X* = 727663.167	*Y* = 4096928.843
Ajolí (AJO)	*X* = 725239.201	*Y* = 4112839.255
Matochal (MAT)	*X* = 744940.117	*Y* = 4109526.314

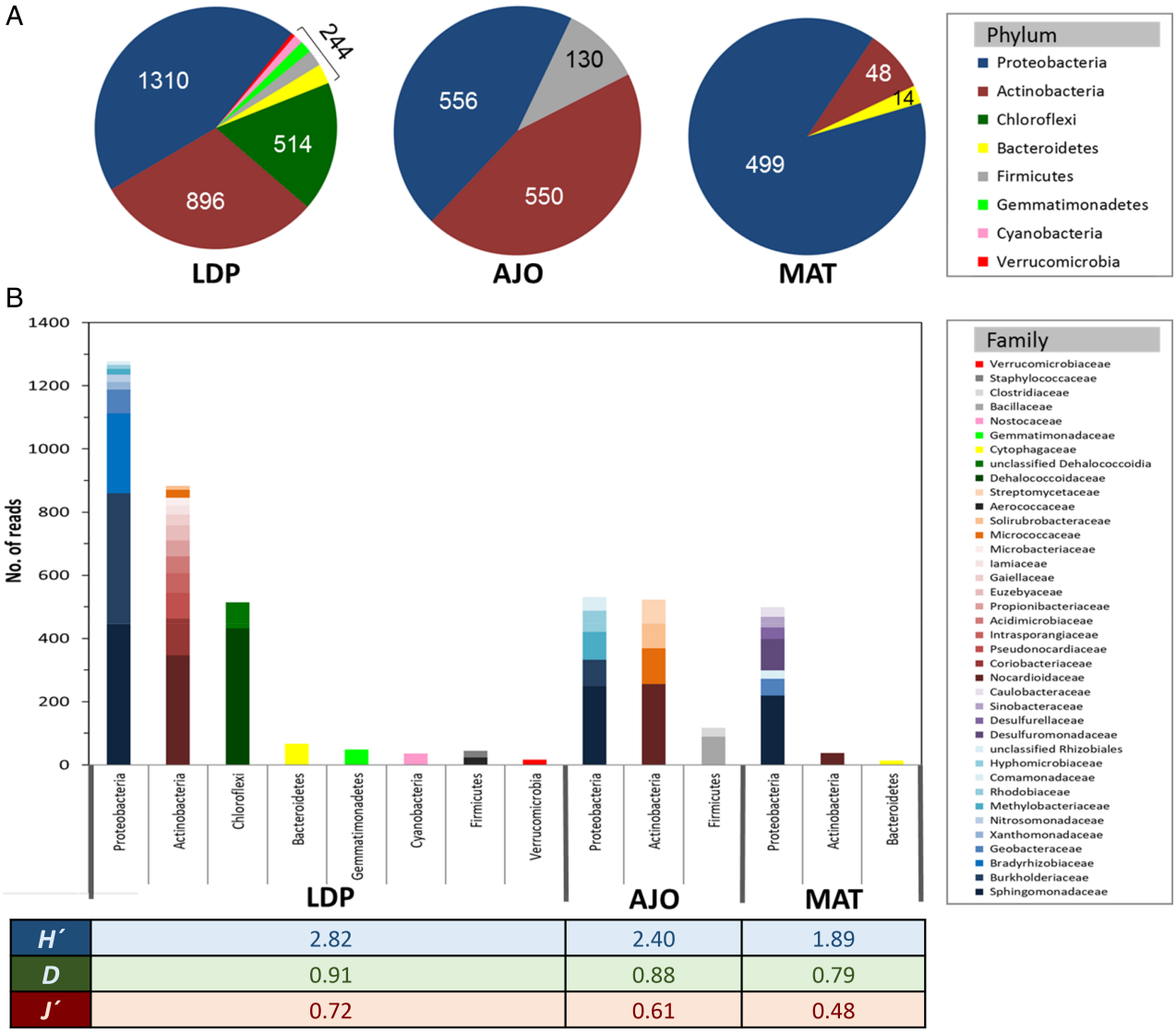
Fig. 1. Map of Doñana National Park (DNP) and its surrounding areas (SW Iberian Peninsula) showing the sites where the soils studied were collected. UTM coordinates of the sites are indicated. [Color figure can be viewed at wileyonlinelibrary.com]

Fifty-one families were identified in these soils of which those identified from more than 15 reads in any of the sampling areas are represented (Fig. 2B). The lower diversity in AJO and MAT microbiota compared to LDP was also observed at the family level with 17, 10 and 36 families respectively. The most abundant family at LDP was Sphingomonadaceae (Proteobacteria) with 444 reads closely followed by Dehalococcoidaceae (Chloroflexi) with 432 and Burkholderiaceae (Proteobacteria) with 415, which all together accounted for 43.6% of the identifications in this soil. At AJO, the Norcardioidaceae family (Actinobacteria) presented the highest numbers with 256 reads but was closely followed by Sphingomonadaceae with 249 and both represented 40.9% of the microorganisms identified for this soil. Sphingomonadaceae was also the predominant family at MAT with 219 reads and just this one family represented 39.0% of the reads.

Although the 16S analysis described above clearly shows a higher diversity of the microorganisms from LPD than from AJO and MAT soils, quantitative characterization of the bacterial communities was done using alpha diversity indices (Fig. 2B, below). The number of microbial species (richness) ranged from 2.82 to 1.89 and 0.91 to 0.79 using the Shannon–Wiener's (H') and Gini–Simpson's indexes (D') respectively, following the expected diversity gradient: LPD > AJO > MAT. The same gradient was found when evenness was estimated by Pielou's (J') method, values ranging from 0.72 to 0.48. Maximum biodiversity/evenness expected values H'_{max}, D'_{max} and J'_{max} are 3.93, 0.98 and 1.0 respectively.

Optimization of protein extraction from soils

Of the soils studied that of AJO has intermediate oxidizable organic matter and organic nitrogen (Supplementary Table 2). Thus, this soil was used for the optimization of a post-alkaline soil protein extraction. All the steps were carried out at room temperature since incubation on ice was also tested and produced a large decrease in protein extraction efficiency (results not shown). After NaOH 0.5 M treatment, centrifugation allowed us to isolate the microbes selectively and separate them from the bulk of the soil's particulate matter as we could verify by visualization under the microscope (results not shown). Washing the isolated microbes with Tris–HCl buffer was an essential step for the lysis of microorganisms to release their proteins. Surprisingly, the presence of 10% glycerol in the washing buffer proved to be highly beneficial for this procedure (Fig. 3). Glycerol supplementation resulted in a 2.72-fold increase in protein extraction efficiency as determined by SDS-PAGE protein gels analysis and quantification (Fig. 3B). To optimize the washing stage required, the effect of the number of successive washing



	LDP	AJO	MAT
H´	2.82	2.40	1.89
D	0.91	0.88	0.79
J´	0.72	0.61	0.48

Fig. 2. Metagenomic analysis showing the abundance and composition of the dominant bacterial phylotypes found in the soil samples.
A. Relative abundance of all phyla found in the soil samples. The number of reads for each phylum is indicated.
B. Taxonomic composition found in the soil samples. The abundance of the phyla/families found is expressed as the number of reads of the corresponding 16S rRNA genes. Only families identified from more than 15 reads are represented. Below, alpha diversity values are shown in terms of richness using the Shannon–Wiener's (*H′*) and Gini–Simpson's (*D*) indexes and evenness estimated by Pielou's (*J′*) method. The soil locations studied are indicated in Fig. 1. [Color figure can be viewed at wileyonlinelibrary.com]

steps on the soil protein extraction efficiency was studied (Fig. 4). A linear increase in protein extraction efficiency ($R = 0.992$) was observed when microbes extracted with NaOH were sequentially washed (0–4 times) with 10% glycerol supplemented buffer and it was maintained after five washes (Fig. 4B). The optimized method was further evaluated starting from different initial amounts of soil (Supplementary Fig. 1). Protein extraction increased linearly in the range of 0.125–1 g of starting soil ($R = 0.9989$).

The efficacy of the microbial protein extraction method developed was further evaluated using soils with different physicochemical characteristics (see Supplementary Table 2). The reliability of the protein extraction method was demonstrated by SDS-PAGE electrophoretic separation and detection of the extracted proteins (Fig. 5). Fluorescent staining with SYPRO Ruby (Fig. 5A) and visible with Coomassie Blue (Fig. 5B) revealed a large number of very intense protein bands throughout the complete range of molecular weight. After image analysis, a similar total protein intensity was obtained in all the soils studied, both when the gels were stained with SYPRO Ruby (Fig. 5C) and with the less sensitive Coomassie Blue staining method (Fig. 5D). A schematic representation of

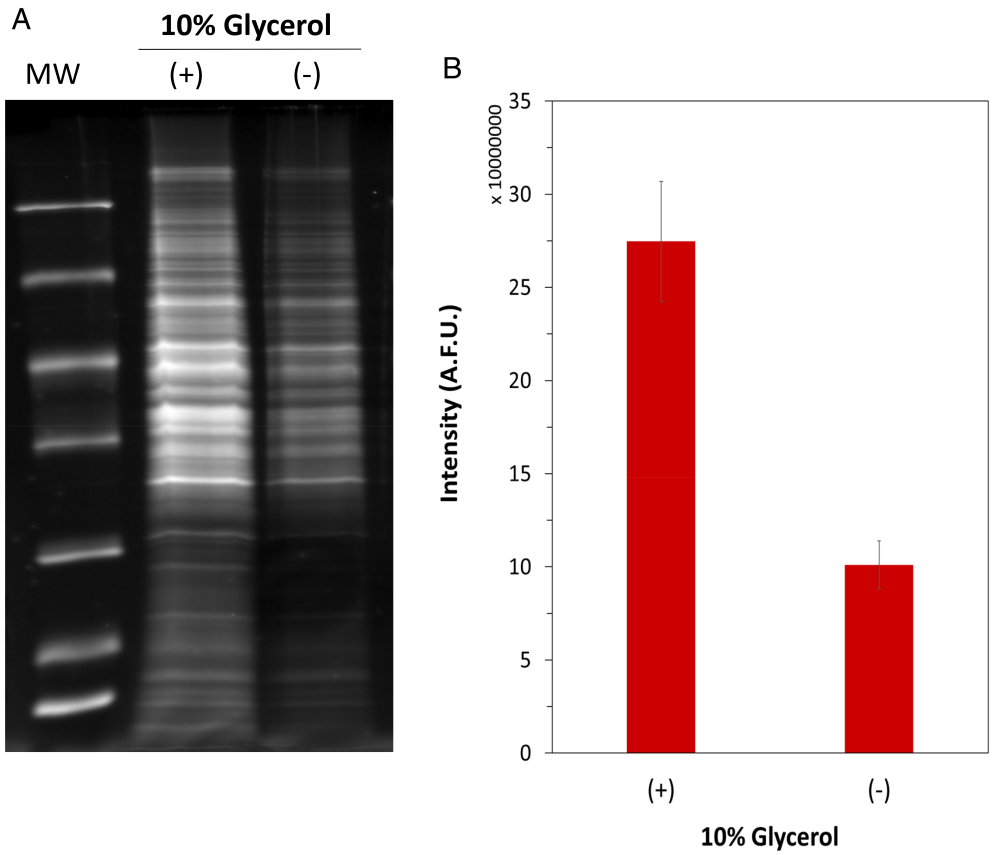


Fig. 3. Effect of the presence of 10% glycerol in the washing buffer on the efficiency of the soil protein extraction. Microbes extracted with 0.5 M NaOH from AJO soil were washed five times with 50 mM Tris–HCl pH 7.5, supplemented or not with 10% glycerol, as described in Experimental procedures.
A. Representative electrophoretic separation of the SYPRO Ruby-stained proteins extracted from 0.5 g of soil using washing buffer with (+) or without (−) 10% glycerol.
B. The efficiency of the protein extraction was evaluated by determining the fluorescence intensity along the gel lanes, using the Image Lab software (version 4.1, Bio-Rad). Bars represent the mean ± SD of eight independent extractions. Molecular weight (MW) markers are specified in Fig. 5. [Color figure can be viewed at wileyonlinelibrary.com]

the global workflow followed for the protein extraction from the soils is shown in Fig. 5E.

Bacterial database selection and protein identification by LC–MS/MS

In the absence of a specific database, results from the 16S rRNA analysis of soils were used to build an in-house database for the metaproteomic analysis. Table 1 shows the results of the protein search in the databases of the most represented microbial families as obtained from the 16S rRNA analysis. Significantly, there is no correspondence between the number of proteins identified by LC–MS/MS and database search and the number of readings obtained for the different families by 16S rRNA. It should be noted that the protein databases of the different microbial families are very poorly annotated so that much more than 99% corresponds to unreviewed records in most cases. By far, Bacillaceae with a low/intermediate representation in the genomic data was the family database that gave the highest number of protein identifications with 131 417 which represents 95.8% of the total IDs obtained. In contrast, the database of Sphingomonadaceae, the dominant family in genomics, only produced 379 IDs (0.3% of the total). For greater data reliability, the number of proteins identified from ≥2 peptides was calculated and Bacillaceae continued to be the family database that provided the highest number of identifications (32 957 IDs; 92.5% of the total) followed by unclassified Rhizobiales (1361), Bradyrhizobiacea (555), Methylobacteriaceae (183) and Hyphomicrobiaceae (180) (Table 1). Among these reliably identified proteins, many were identified in different species of the same family so the number of non-redundant proteins was calculated (Table 1). Within unique identifications, the number of exclusive IDs for each database was Bacillaceae (6977), unclassified Rhizobiales (224), Bradyrhizobiacea (41), Methylobacteriaceae (117) and Hyphomicrobiaceae

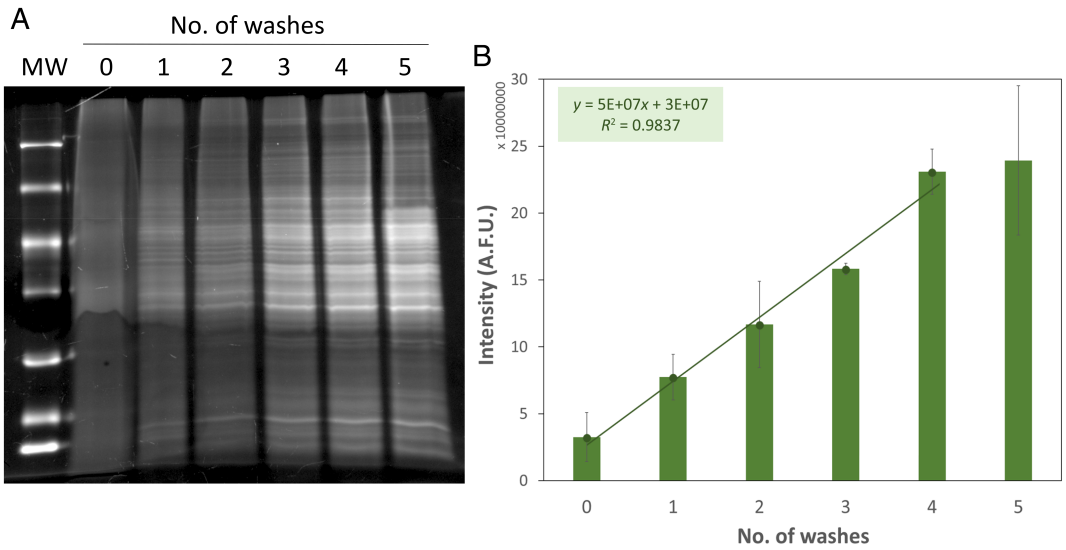


Fig. 4. Effect of the number of successive washing steps on the efficiency of the soil protein extraction. Microbes extracted with 0.5 M NaOH from AJO soil were sequentially washed (1–5 times) in 50 mM Tris–HCl pH 7.5 buffer containing 10% glycerol, as described in Experimental procedures, and compared with non-washed (0) cells.
A. Electrophoretic separation of the SYPRO Ruby-stained proteins extracted from 0.5 g of soil after 0–5 washings steps.
B. The efficiency of the protein extraction was evaluated by determining the fluorescence intensity (A.F.U., arbitrary fluorescence units) along the different lanes in the gel using the Image Lab software (version 4.1, Bio-Rad). Bars represent the mean ± SD of four independent extractions. Molecular weight (MW) markers are specified in Fig. 5. [Color figure can be viewed at wileyonlinelibrary.com]

(8) (Supplementary Fig. 2). Finally, a score was calculated by integrating the number of proteins identified from ≥2 peptides and the average of coverage and identified peptides obtained in each database search (Table 1). The highest scores were obtained after searching against the unclassified Rhizobiales (21.3) and Bradyrhizobiaceae (18.9) databases. To compare, scores obtained after running a human sample and our soil samples against the human database were 35.4 and 6.7 respectively. For the metaproteomic analysis, a combined database was constructed containing the UniProtKB databases of those bacterial families with a score greater than 10 which also coincides with those that gave rise to more than 100 non-redundant identified proteins (Table 1 and Supplementary Fig. 2) that are Bacillaceae, unclassified Rhizobiales, Bradyrhizobiacea, Methylobacteriaceae and Hyphomicrobiaceae.

LC–MS/MS analysis identified 2182 non-redundant proteins. Of these, only those proteins with a covariance of less than 25% in all the soils and showing statistically significant differences ($p < 0.05$) were selected. Supplementary Table 3 lists the 135 unique bacterial proteins showing significant expression differences in the different soils. A hierarchical clustering analysis was carried out to quantify the alterations in the intensity of the proteins in the different soils visually (Fig. 6A). Microbial proteins were grouped into six clusters based on similarities in their expression levels. Clusters B, C and E are composed of proteins that were found only in the soils at AJO (4), LDP (16) and MAT (46 proteins) respectively. Cluster A included 19 proteins with a lower intensity at LDP than all the other sampling sites. Finally, clusters D (24) and F (26) included proteins with a lower intensity in AJO soils following a gradient AJO < MAT<LDP (cluster D) and AJO < LDP < MAT (cluster F). The cluster in which each protein is included is indicated in Supplementary Table 3. When compared to the LDP reference protected area up to 55 proteins changed significantly in the MAT soil with 38 proteins increasing and 17 decreasing in intensity. On the contrary, the soil from AJO showed a greater number of proteins with lower intensity (42) than those with higher intensity (15) for a total of 57 changes compared to LDP. Finally, a very high number of proteins (89) whose intensity increases were obtained at MAT in respect to AJO while only five proteins decreased (Fig. 6A and Supplementary Table 3).

An enrichment analysis was carried out to discover any functional interactions between the differentially expressed proteins. Figure 6B (upper half) shows the protein–protein association network obtained with STRING in the *Rhizobium* sp. LPU83 genome dataset. Individual proteins are identified in the network by its corresponding gene name (for equivalence, see Supplementary Table 3). A 101 node and 284 edge network was obtained with a protein–protein interaction (PPI) enrichment *p*-value <1.0e-16. Hierarchical clustering showed seven (I–VII) tightly connected modules within the network, some of which broadly correspond to functional units as shown in Fig. 6B (lower half). Thus,

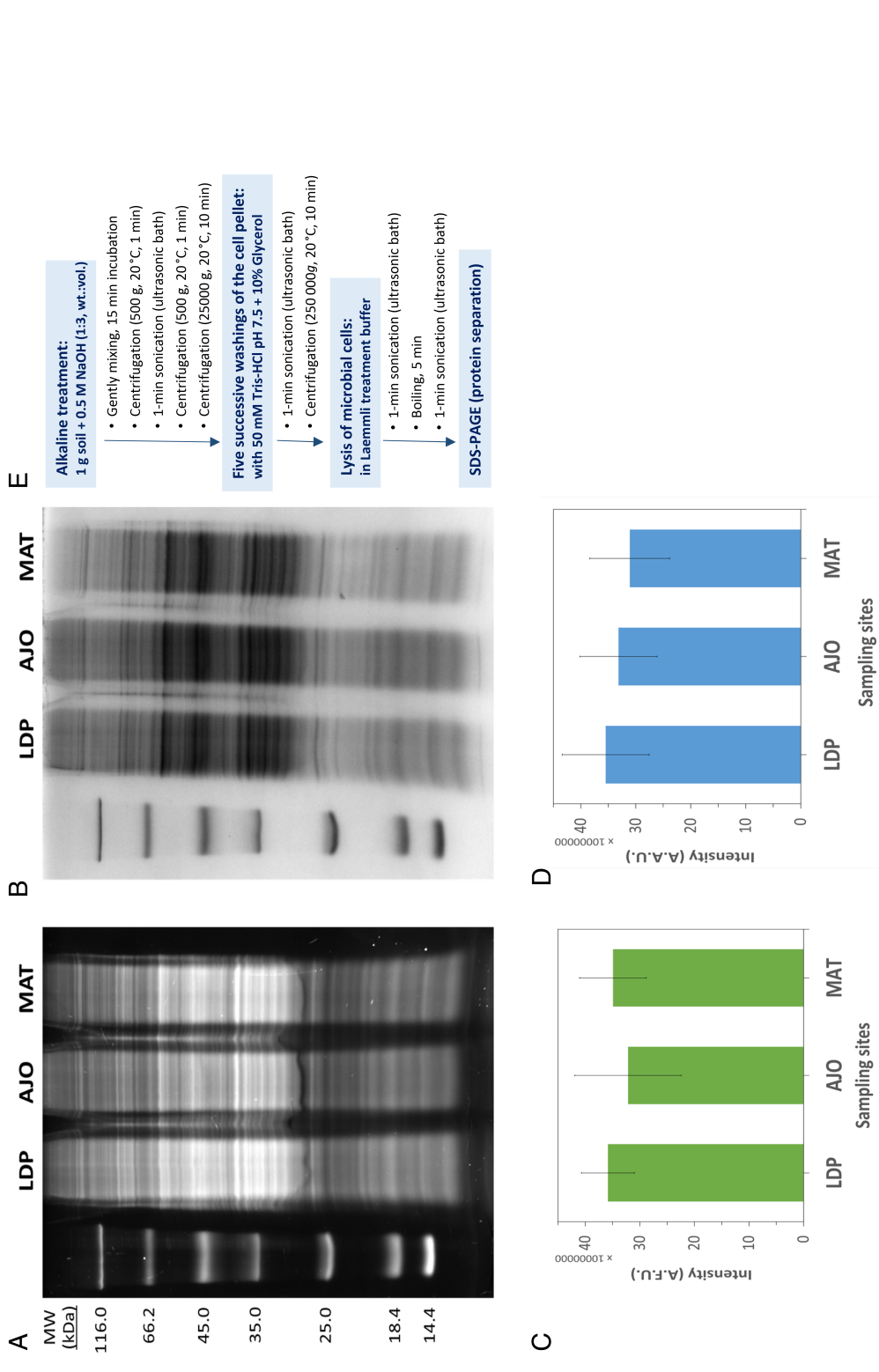


Fig. 5. Electrophoretic separation of proteins extracted from 1 g of the different soils studied. After extraction, the proteins (20 µl) were analysed by 12.5% SDS-PAGE and the fluorescence of SYPRO Ruby (A) and the absorbance of Coomassie Blue (B) stained proteins were assessed and representative gels are shown. Unstained Protein Markers (2 µl; Thermo Scientific, Lithuania, #26610), whose molecular weight (MW) is specified on the left, were loaded on the first lane. The Image Lab software was used to assess the fluorescence intensity and the absorbance intensity, both in arbitrary units, along the different lanes in the gels stained with SYPRO Ruby (C) and Coomassie Blue (D) respectively. The values shown are the mean ± SD of 10 independent extractions. The soil locations studied are indicated in Fig. 1.
E. Schematic workflow of the soil's protein extraction protocol. All the steps were carried out at room temperature. [Color figure can be viewed at wileyonlinelibrary.com]

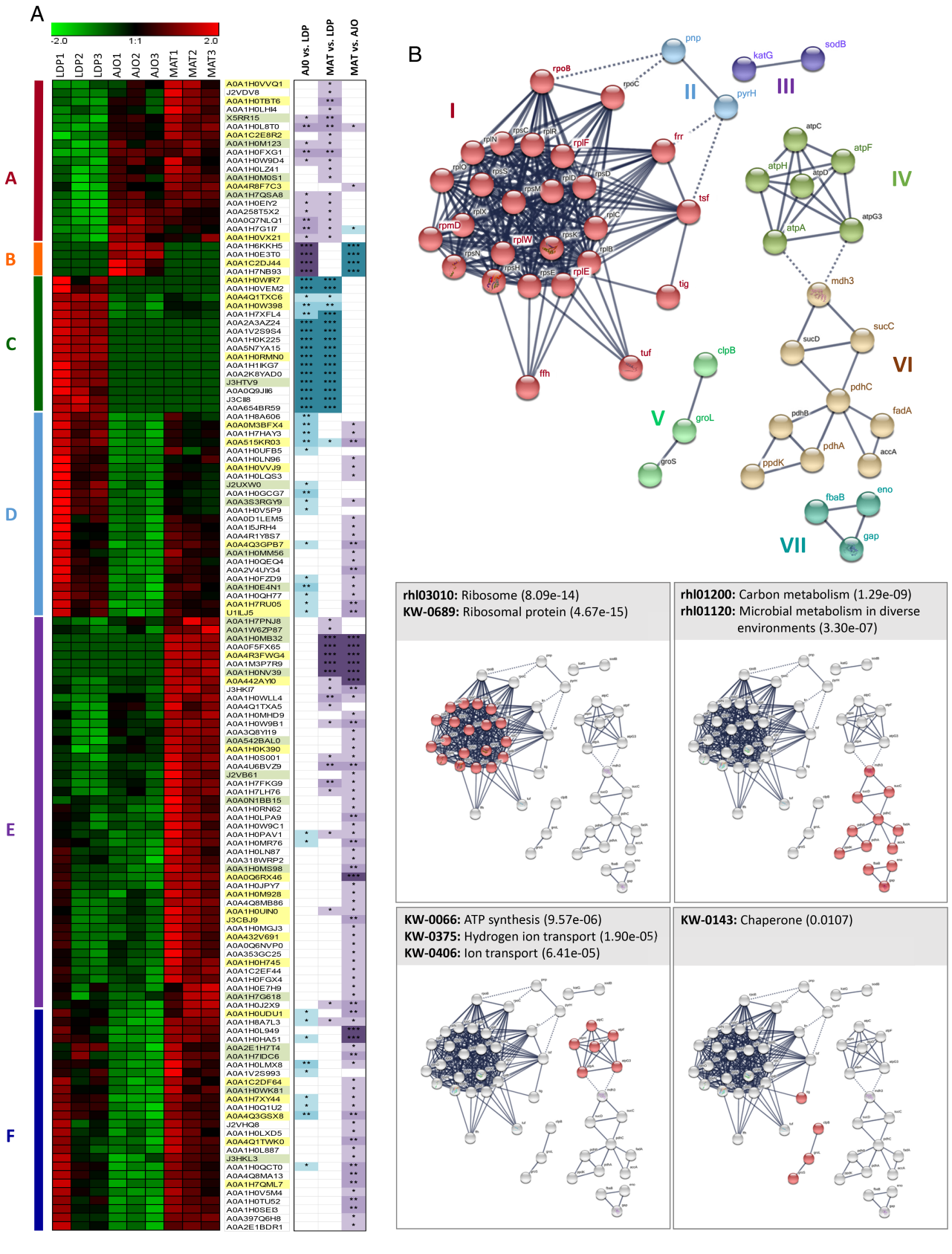
Table 1. Protein identifications obtained from soil samples by a peptide spectra search against the databases of the most represented microbial families identified by metagenomics.

	Metagenomic analysis	Metaproteomic analysis						
Family	No. of reads	UniProtKB database entries	Total protein IDs	Protein IDs (≥2 peptides)	Protein IDs (non-redundant)	$\overline{X}$ coverage	$\overline{X}$ peptides	Score[a]
Sphingomonadaceae	912	1 472 228	379	142	83	8.91	3.07	9.5
Nocardioidaceae	641	647 651	35	2	2	8.39	2.00	2.4
Burkholderiaceae	499	2 770 716	216	57	44	13.49	3.37	9.7
Dehalococcoidaceae	432	14 452	21	2	2	3.30	3.00	2.5
Bradyrhizobiaceae	254	1 274 120	1023	555	327	17.3	4.73	18.9
Micrococcaceae	137	914 208	26	1	1	18.79	4.00	5.1
Geobacteraceae	128	132 155	42	13	7	7.16	2.69	5.3
Coriobacteriaceae	116	158 398	30	3	3	7.05	2.00	2.7
Methylobacteriaceae	106	526 118	480	183	117	11.93	3.73	12.4
Desulfuromonadaceae	99	79 573	50	14	12	7.59	3.00	6.1
Solirubrobacteraceae	91	6590	22	2	2	5.11	2.00	2.0
Bacillaceae	89	3 847 234	131 417	32 957	7386	12.26	2.17	12.2
unclassified Dehalococcoidia	82	156 817	34	2	2	2.46	2.50	1.7
Cytophagaceae	81	449 457	72	16	15	9.07	2.44	5.3
Pseudonocardiaceae	80	1 255 276	44	6	5	6.79	2.17	3.5
Streptomycetaceae	76	5 098 311	29	4	4	6.12	2.00	2.8
Intrasporangiaceae	63	221 682	18	2	2	7.93	2.00	2.4
Acidimicrobiaceae	54	216 887	56	5	5	5.74	2.00	2.9
Propionibacteriaceae	50	216 529	43	7	5	5.94	2.71	4.4
Hyphomicrobiaceae	49	292 693	468	180	102	10.00	3.82	12.4
Gemmatimonadaceae	49	41 699	24	7	5	4.86	2.14	3.3
Euzebyaceae	47	5625	8	0	0	0.00	0.00	0.0
unclassified Rhizobiales	43	4 346 249	2303	1361	701	17.88	4.86	21.3
Desulfurellaceae	37	8275	25	5	5	4.45	2.00	2.7
Nostocaceae	36	289 665	76	19	16	8.41	2.42	5.3
Gaiellaceae	36	915	5	0	0	0.00	0.00	0.0
Sinobacteraceae	33	57 685	65	13	10	7.07	2.54	5.0
Caulobacteraceae	31	467 109	173	59	42	8.10	3.08	8.3
Total	**4351**	**24 968 317**	**137 184**	**35 617**	**8905**			
Human database (human sample)		20 333	3297	2461	2461	21.16	7.51	35.4
Human database (soil samples)		20 333	39	12	12	14.13	3.00	6.7

[a]Score $= \log_{10}$ [Protein IDs (≥2 peptides) $\times$ $\overline{X}$ coverage]$\overline{X}^{\text{peptides}}$.

cluster I, the largest and highest-scored, is composed of a large group of 19 ribosomal proteins. Six proteins involved in (hydrogen) ion transport and ATP synthesis constitute cluster IV, most of which (protein products of *atpA*, *atpC*, *atpD*, *atpH* and *atpG3* genes) are components of the F-type ATP synthase complex CF(1), the catalytic core (UniProt keyword: KW-0139; f.d.r.: 1.90e-05). Clusters VI and VII include proteins that are involved in microbial carbon metabolism in diverse environments including components of the glucolysis/gluconeogenesis (*fbaB*, *eno*, *gap*), of the TCA cycle (*sucC*, *sucD* and *mdh3*) or which connect both routes (*pdhA*, *pdhB*, *pdhC*). All these proteins except *ppdK* product are involved in the biosynthesis of antibiotics (KEGG pathway: rhl01130; f.d.r.: 2.63e-08). Three chaperones form cluster V while clusters II and III are composed of only two proteins. Significantly, two primary antioxidants enzymes, catalase-peroxidase (*katG*) and superoxide dismutase (*sodB*), are part of cluster III. Finally, cluster II which shows partial connections with proteins of cluster I is composed of the protein products of *pnp* involved in mRNA degradation and *pyrH* involved in step 1 of the cytidine 5′-triphosphate (CTP) biosynthesis via the *de novo* pathway.

It is noteworthy that of the 135 bacterial proteins with significant expression changes up to 31 (23%) proteins are membrane transport proteins. Of these, about two-third (21 proteins) are significantly more expressed in the two most contaminated sites MAT and AJO with respect to LDP. Of the transport proteins, 24 proteins (68%) correspond to ATP-binding cassette (ABC) transporters (Fig. 6 and Supplementary Table 3). Additionally, more than 17% of the identified proteins with significant changes among soils are directly related to the defence against



environmental stress (Fig. 6 and Supplementary Table 3). Several of these 21 proteins are involved in the maintenance of the cellular redox state and in defence/repair mechanisms including proteins that participate in: (i) carbohydrate metabolism that also present other relevant non-metabolic functions ('moonlighting proteins'): glyceraldehyde-3-phosphate dehydrogenase, fructose-biphosphate aldolase, malate dehydrogenase, NAD(P)-dependent short-chain alcohol dehydrogenase, isocitrate dehydrogenase, enolase; (ii) antioxidative defence mechanisms: superoxide dismutase, catalase-peroxidase and cysteine synthase; (iii) electron transference: electron transfer flavoprotein, cytochrome *c* and FMN-dependent oxidoreductase; (iv) detoxification: chlorite dismutase and thiosulfate/3-mercaptopyruvate sulfurtransferase; (v) organic carbon compounds degradation pathways: dioxygenase, aryl-alcohol dehydrogenase and aldehyde dehydrogenase; and (vi) protein repair, degradation or misfolding prevention (discussed above): peptide methionine sulfoxide reductase, periplasmic serine endoprotease DegP-like, ATP-dependent zinc metalloprotease FtsH, and 60 kDa and ClpB chaperones. Most of the alteration patterns for these proteins corresponded with increases in MAT and/or AJO soils and only five of them were down-regulated in the polluted spots (Fig. 6 and Supplementary Table 3).

Discussion

By hosting at least one-quarter of all living organisms on the planet, soils are a crucial reservoir of biodiversity (Decaëns *et al*., 2006; Thakur *et al*., 2020). Microbes dominate soil diversity and provide key ecosystem functions and services (Maron *et al*., 2011; Maron *et al*., 2018; Starke *et al*., 2019; Tibbett *et al*., 2020). Bacteria and fungi are generally dominant in soil with 10^2–10^4 times more estimated biomass than protists, archaea and viruses, the other components of the soil microbiome (Fierer, 2017). Furthermore, bacterial communities have been proposed as valuable indicators of soil condition since they may respond in a predictable manner to environmental perturbations and some bacterial individual taxa show key relationships with their environment (Hermans *et al*., 2017).

Recent advances in high-throughput sequencing technologies have enabled investigations of soil microbial composition, structure and biodiversity with greater precision (Maron *et al*., 2011; Zhou *et al*., 2015; Thakur *et al*., 2020). In this study, DNA extraction followed by 16S rRNA sequencing made it possible to analyse the bacterial soil responses in a highly protected natural area, DNP and surroundings. The number of bacterial sequences (reads) identified was higher in soils from the protected area LDP than in those from AJO (2.4-fold) and especially than in those from MAT (5.6-fold). A loss of microbial diversity was found following the same gradient: LDP > AJO > MAT when richness (Shannon–Wiener's and Gini–Simpson's indexes) and evenness (Pielou's method) were estimated (Fig. 2). LDP, in the heart of the National Park, is a low contaminated area (Vioque-Fernández *et al*., 2009; Fernández-Cisnal *et al*., 2014; Gago-Tinoco *et al*., 2014). Surrounding DNP there are several sources of anthropogenic contamination (urban enclaves, and industrial, agricultural and mining activities) that endanger its balance and ecological health (Vioque-Fernández *et al*., 2009; García-Sevillano *et al*., 2014; Abril *et al*., 2015; Fernández-Cisnal *et al*., 2018). AJO is affected by urban wastes and strawberry, citrus fruit and grape fields while MAT, next to the Guadiamar river, is under the influence of paddies (García-Sevillano *et al*., 2014; Abril *et al*., 2015). Previous studies have shown that AJO presents medium levels of contaminants while MAT is a pollution hotspot with high levels of pesticides and metals found in the soil and/or water due to the use of algaecides in the paddies (Supplementary Table 1) (Vioque-Fernández *et al*., 2007; Vioque-Fernández *et al*., 2009; Fernández-Cisnal *et al*., 2014; Gago-Tinoco *et al*., 2014; García-Sevillano *et al*.,

Fig. 6. Clustering and functional analyses of the differentially expressed soils' bacterial proteins identified by LC–MS/MS.
A. K-means clustering as obtained by the Genesis analysis (left). Proteins are grouped into six clusters (A–F). Each row in the heatmap represents one differentially expressed protein. Green rectangles indicate samples with a lower intensity relative to other soils while red rectangles represent higher levels. The colour intensity is proportional to the fold-change as represented by the scale. Proteins involved in the transport of a wide variety of substrates are highlighted in yellow and those proteins that have a relevant protective role against environmental stress are in light green (see Supplementary Table 3). Statistically significant differences between the different soils are shown (right). Protein intensities increasing (purple) or decreasing (blue) in soils from AJO (vs. LDP), MAT (vs. LDP) and MAT (vs. AJO) are highlighted. Statistical significances are as follows: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.
B. Functional analysis as obtained by STRING (version 11.0). The upper screenshot shows the protein–protein association network obtained in STRING (version 11.0) using *Rhizobium* sp. LPU83 as the model soil microorganism. The gene names of the proteins identified are coloured in the figure and their correspondence is indicated in Supplementary Table 3. For simplification, the disconnected nodes in the network are not shown. The confidence cutoff has been set to 'highest' (0.900) and the maximum number of interactions has been set at no more than 10 in the first and second shell. The network obtained was clustered to an MCL inflation parameter of 2 and seven (I–VII) clusters were obtained. The insets at the bottom show the functional enrichments in the network, indicating: the KEGG Pathways or UniProt Keywords, their codes (in bold), a description of the enriched network and its False Discovery Rate (FDR-corrected *p*-value; in parenthesis) which shows the expected proportion of discoveries (rejected null hypotheses) that are false (incorrect rejections of the null). Screenshots show those genes that contribute to the enrichment of each pathway or functional subsystem. [Color figure can be viewed at wileyonlinelibrary.com]

2014; Fernández-Cisnal *et al*., 2017; Fernández-Cisnal *et al*., 2018). It is worthy of note that the MAT area suffered an accidental input of high levels of toxic metals (e.g. Fe, Hg, Pb, Cd, Ag, As, Zn, Cu) transported by the Guadiamar river as a result of the rupture of the Aznalcollar mine tailing pond in 1998 (Grimalt *et al*., 1999; Bonilla-Valverde *et al*., 2004). Several threats to the soil, such as pollution, intensive human exploitation, land-use change, soil erosion or climate change have negative effects on soil diversity (Jiao *et al*., 2019; Tibbett *et al*., 2020). Thus, indigenous microbial communities of agricultural soils were shown to be strongly affected by organic and inorganic chemical contamination (Ventorino *et al*., 2018; Jiao *et al*., 2019; Beaumelle *et al*., 2020). The impact of different pesticides has been assessed by evaluating changes in gene expression and biochemical enzymatic activities and in the diversity and composition of soil bacterial communities (Feld *et al*., 2015; Bacmaga *et al*., 2018; Storck *et al*., 2018; Bacmaga *et al*., 2019). A recent study has shown that pesticides decrease the bacterial diversity and abundance in irrigated paddies (Onwona-Kwakye *et al*., 2020). Bioactivity, richness and microbial diversity also decrease when the concentration of heavy metals increase (Yin *et al*., 2015; Xie *et al*., 2016; Feng *et al*., 2018; Liu *et al*., 2018; Luo *et al*., 2018; Fatimawali *et al*., 2020; Thomas IV *et al*., 2020; Zhao *et al*., 2020). Significantly, it has been shown that heavy metals contribute to the dissemination of antimicrobial resistance in contaminated soils due to the abundance and co-occurrence of antibiotic and metal resistant genes (Chen *et al*., 2019; Thomas *et al*., 2020). The decline in the microbial diversity of the soil can significantly affect the soil's ability to function normally, to respond to perturbations and to recover (Loreau *et al*., 2001; Maron *et al*., 2011; Cardinale *et al*., 2012; Maron *et al*., 2018; Beaumelle *et al*., 2020; Tibbett *et al*., 2020).

The bacterial composition of the microbiomes associated with the three soils analysed in this work shows that Proteobacteria and Actinobacteria, which are also the two bacterial phylotypes predominant across the globe (Delgado-Baquerizo *et al*., 2018), dominated the communities (Fig. 2). In the pristine LDP area, we must add the presence of other globally ubiquitous although less abundant phyla (i.e. Chloroflexi, Bacteroidetes, Gemmatimonadetes, Cyanobacteria, Firmicutes, Verrucomicrobia) to the nine and 12 families identified of the Proteobacteria and Actinobacteria phyla respectively, and that contribute to increasing the microbial diversity in this area. Proteobacteria was the most abundant phylum given its relevant role in soils which is to provide basic functions in the biogeochemical cycles (Delgado-Baquerizo *et al*., 2018; Feng *et al*., 2018). In general, Proteobacteria have shown a significant tolerance to heavy metals (Zhao *et al*., 2019; Zhao *et al*., 2020). The disappearance of many phyla and families with respect to LDP explains the great loss of diversity in AJO and MAT, two areas that have been historically exposed to contamination and that surround DNP as discussed above. As shown here, a strong decrease in the Actinobacteria, Chloroflexi, Gemmatimonadetes and Firmicutes phyla in contaminated soils has been previously reported (Ventorino *et al*., 2018). The greatest reduction in abundance observed for Actinobacteria in the MAT soil confirms the presence of pollution in this area, given its known global susceptibility to environmental stress (Yin *et al*., 2015; Ventorino *et al*., 2018).

Metagenomic analysis would inform of the metabolic potential (He *et al*., 2017), however, metaproteomics is a more appropriate approach to unveil the active functional role of microbes in different ecosystems and to reveal the metabolic adaptations in response to environmental stress. However, the extraction of soil proteins in enough quantities and of the adequate purity has long been a challenging task in great need of and with much room for improvement (Starke *et al*., 2019; Chiapello *et al*., 2020). To the low amounts of protein present in the soil, it is worth highlighting the significant amounts of interfering substances (Chen *et al*., 2009; Keiblinger *et al*., 2012; Bastida *et al*., 2014; Qian and Hettich, 2017; Greenfield *et al*., 2018; Mandalakis *et al*., 2018). Nevertheless, high-quality sample preparation is crucial to obtain a good resolution in proteomic analysis (Keiblinger *et al*., 2012; Bastida *et al*., 2014; Qian and Hettich, 2017). Until now, the recovery of proteins from a matrix as complex as soil has been shown to be highly dependent on the extractant solvent used and the type of soil (Bastida *et al*., 2014; Greenfield *et al*., 2018; Mandalakis *et al*., 2018). Here, we describe the development and optimization of an alkaline-based protocol for the extraction of proteins from soil. It has been previously described that an initial NaOH treatment facilitates the extraction of humic compounds from the soil (Benndorf *et al*., 2007; Greenfield *et al*., 2018). Alkaline pretreatment with 0.1 M NaOH has been used for the efficient extraction of proteins of two different yeast strains, *Saccharomyces cerevisiae* and *Hansenula polymorpha*. After an alkali treatment, yeast cells were pelleted and directly boiled in standard electrophoresis loading buffer for virtually complete protein extraction as shown by SDS-PAGE (Kushnirov, 2000). The alkaline method was later adapted by increasing the NaOH concentration to 0.3 M for protein extraction from fission yeast cells (Matsuo *et al*., 2006). An alkaline pretreatment of the soil with 0.5 M NaOH followed by sequential centrifugation allowed the separation of the microbes from the bulk of the soil's particulate matter including interfering substances. Unlike what occurs with cultured yeasts, direct protein extraction from the pelleted microorganisms by boiling in an electrophoresis-loading

buffer was completely inefficient. However, successive washing of the cells with Tris–HCl buffer increased the protein extraction greatly (Fig. 4). Furthermore, the efficiency of extraction was significantly improved when the washing buffer was supplemented with 10% glycerol (Fig. 3). Unlike as has been previously described (Benndorf *et al*., 2007), NaOH-alkaline treatment did not disrupt the microorganisms as their isolation by selective centrifugation and visualization under the microscope indicated (results not shown). The sequential washing of the microbial pellet favoured subsequent cell breakage and the efficient extraction of proteins by boiling in SDS-PAGE loading buffer and, more importantly, made it possible to avoid the need for a subsequent phenol extraction and washing with organic solvents (Benndorf *et al*., 2007) which makes the protocol easier, faster, non-toxic and more environmentally friendly. The toxic and hazardous waste-generating phenol treatment has so far been the most widely used method to remove interfering substances and to extract soil proteins (Benndorf *et al*., 2007; Chen *et al*., 2009; Keiblinger *et al*., 2012; Bastida *et al*., 2014; Gunnigle *et al*., 2014; Mandalakis *et al*., 2018; Thorn *et al*., 2018). Recently, other methods promising comparable average yields in terms of proteins extracted have been proposed. Thus, coagulation with Al^{3+} was less efficient than the phenol treatment to remove humic acids while the recently marketed NoviPure Soil Protein Extraction kit is quite expensive (Mandalakis *et al*., 2018; Bona *et al*., 2019). Here, very high resolved gels with similar band profiles, number of protein bands and total band intensities were obtained from the different soils studied after SDS-PAGE separation and staining with SYPRO Ruby or Coomassie Blue (Fig. 5). The texture and organic composition varied among the three soils studied, with that of LDP (silty clay loam) containing the highest amount of oxidizable organic matter and organic nitrogen and that of MAT (silty clay) the lowest, while the soil of AJO (loamy sand) showing intermediate values (Supplementary Table 2). Despite the physicochemical differences between the soils, the amount of protein extracted was similar in all of them as shown after SDS-PAGE (Fig. 5) obtaining a good number and intensity of protein bands with a resolution of proteins ranging in size from the bottom to the top of the gel not previously achieved (Benndorf *et al*., 2007; Chen *et al*., 2009; Taylor and Williams, 2010; Thorn *et al*., 2018; Renu *et al*., 2019) even when using cultured bacteria-amended soil as the starting material (Benndorf *et al*., 2007; Taylor and Williams, 2010; Mandalakis *et al*., 2018). The main reason for the success of the post-alkaline method lies in the separation of the microbial cells and elimination of interfering substances prior to the protein extraction. This was indicated previously by a greater extractability and detection of protein on SDS-PAGE gels when the microorganisms were separated from soils using density gradient centrifugation as a preliminary step to protein extraction. Nevertheless, this method was limited by large differences in the efficiency of the extraction depending on the type of soil and by the inevitable need for soil amendment (Taylor and Williams, 2010). The efficient breakdown of the microbial cells is another critical step in protein extraction. With this intent, the boiling and repeated sonication with the detergent SDS-containing loading buffer allowed effective cellular lysis and maximized the protein extraction (Chen *et al*., 2009; Chourey *et al*., 2010; Keiblinger *et al*., 2012; Bastida *et al*., 2014). Moreover, the complete protein extraction from soil protocol described here (Fig. 5E) lasts less than 2 h, a clear improvement over previous time-consuming processes (2 days) which require overnight precipitation (Chourey *et al*., 2010; Keiblinger *et al*., 2012; Starke *et al*., 2019).

Once an efficient and reliable protein extraction method has been achieved, a proper database designed for protein identification is crucial (Keiblinger *et al*., 2012; Siggins *et al*., 2012). Large databases have a higher risk of giving false-positive matches and make identification extremely time-consuming and hardware-demanding (Wang *et al*., 2016). A promising solution to the absence of complete protein databases is to build in-house databases based on genomic data previously obtained from the same environmental samples (Zampieri *et al*., 2016; Mattarozzi *et al*., 2017; Chiapello *et al*., 2020). A reference database was created for this study by combining the UniProtKB databases of the dominant bacterial families (i.e. Bacillaceae, unclassified Rhizobiales, Bradyrhizobiaceae, Methylobacteriaceae and Hyphomicrobiaceae) in the soils studied as determined from the previous 16S rRNA analysis. It should be noted that Sphingomonadaceae, the dominant family, was not included in the database since it had a very low number of protein identifications and score; on the contrary, Bacillaceae was by far the family with the highest number of identifications (131 417 total protein IDs, 7386 non-redundant protein IDs) and unclassified Rhizobiales had the highest score (21.3) even though both were underrepresented in genomic data (Table 1, Supplementary Fig. 2). Afterwards, LC–MS/MS analysis identified 2182 unique proteins using the reference database constructed which is a high number considering that redundancies were removed and only confident identifications (≥2 peptides) were considered. Functional redundancy is common in complex environmental samples as similar proteins are expressed by a range of microbial species (Siggins *et al*., 2012; Liu *et al*., 2019). In 2010, the identification of 716 redundant and 333 non-redundant proteins, the deepest proteome coverage to that date, was obtained after direct protein extraction from a natural non-spiked soil by the SDS-TCA

method (Chourey *et al*., 2010; Becher *et al*., 2013). Later studies showed that the number of proteins detected varied depending on the soil and the extraction protocol used with 494 non-redundant proteins (Keiblinger *et al*., 2012) and 1048 total proteins (Bastida *et al*., 2014) identified at best. Even recently, only a total of 579 and 696 bacterial proteins had been identified associated with two soils from a vineyard (Bona *et al*., 2019) and with the rhizosphere of maize (Renu *et al*., 2019) respectively.

Of the proteins identified, a total of 135 unique bacterial proteins showed significant expression differences in the soils studied. Although the total number of significant differences is similar, MAT (38) soil had a higher number of upregulated proteins than AJO soil (15) in comparison to LDP soil. Furthermore, comparing both sampling sites surrounding DNP, up to 89 proteins were upregulated at MAT versus the only five proteins with a higher intensity at AJO (Fig. 6A and Supplementary Table 3). As discussed above, AJO has historically presented intermediate levels of contaminants while MAT has been a more polluted hotspot. PPI network analysis revealed the global biological processes, based on molecular interactions and functional associations, which are altered in response to the environmental changes (Fig. 6B). Coincident global responses were obtained in the soil microbiome with those that showed higher organisms (e.g. *Mus spretus* mice) in the same area studied as determined by label-free quantitative proteomics analysis (Michán *et al*., 2019). Matching responses include (i) Increased synthesis of protective antioxidant enzymes to fight against the oxidative stress generated by pollutants (Cluster III). (ii) *De novo* synthesis of proteins, including protective proteins and the recovery of those damaged under oxidative stress conditions, to preserve cellular homeostasis. This requires the activation and assembly of the ribosomal translational machinery (Clusters I and II). (iii) The correct folding of the newly synthesized protein that requires the sequential actions of multiple molecular chaperones (Cluster V). (iv) The increase in aerobic respiratory metabolism to address the huge demand for metabolic energy of all the previous processes, highlighting protein translation. Metabolic changes include components of the glycolysis and the TCA cycle (Clusters VI and VII) and the mitochondrial machinery of ATP synthesis (Cluster IV). A previous transcriptomic study showed that exposure of the actinomycete *Rhodococcuss aetherivorans* I24 to contaminated sediments or PCBs in pure cultures increased the expression of genes encoding several antioxidant enzymes, chaperones and ribosomal proteins, among others (Puglisi *et al*., 2010).

The highest number of proteins that changed in response to pollution were those involved in membrane transport, highlighting the ABC transporters that form one of the largest and possibly oldest protein families (Supplementary Table 3). Bacterial ABC transporters have environmental relevance since they are involved in many important and diverse processes, including multidrug and antibiotics resistance, environmental sensing, growth under stress conditions, osmosensing, protein secretion and nutrient uptake (Hosie *et al*., 2002; Dawson and Locher, 2006; Basavanna *et al*., 2009). Thus, ABC transporters were the most abundant functional group observed in Cd-contaminated soils (Feng *et al*., 2018). They couple ATP hydrolysis not only to the active import of nutrients, biosynthetic precursors, trace metals and vitamins but also to the export of antibiotics, toxins, xenobiotics, drugs, bacteriocins, hydrolytic enzymes, siderophores, and so on. Therefore, it is a rather expensive process that highly contributes to increasing the energy demand of bacterial cells (Davidson and Chen, 2004; Basavanna *et al*., 2009). Anyhow, the different uptake or extrusion mechanisms and the broad and diverse range of substrate exchanged allow bacteria to quickly adapt to and colonize changing environments (Jeckelmann and Erni, 2020) Furthermore, several proteins that showed significant increases in the two polluted sites, AJO and MAT, are directly related to the defence against environmental stress (Fig. 6 and Supplementary Table 3), many of which have previously been shown to respond to environmental pollution, including dioxygenase (Sharma *et al*., 2019), glyceraldehyde-3-phosphate dehydrogenase (Reyes-Hernández *et al*., 2009; Michán *et al*., 2019), fructose-biphosphate aldolase (Fernández-Cisnal *et al*., 2014, 2017; Michán *et al*., 2019), cysteine synthase (Ding *et al*., 2021), isocitrate dehydrogenase (Fernández-Cisnal *et al*., 2017), enolase (Puglisi *et al*., 2010; Michán *et al*., 2019), ATP-dependent zinc metalloprotease FTSH (Tang *et al*., 2016), catalase-peroxidase (Puglisi *et al*., 2010; Michán *et al*., 2019), chlorite dismutase (Hofbauer *et al*., 2014), thiosulfate sulfurtransferase (Michán *et al*., 2019), peptide methionine sulfoxide reductase MsrA (Weissbach *et al*., 2002), superoxide dismutase (Puglisi *et al*., 2010; Fernández-Cisnal *et al*., 2014), FMN-dependent oxidoreductase (Ellis, 2010; Puglisi *et al*., 2010), aldehyde dehydrogenase (Puglisi *et al*., 2010; Abril *et al*., 2015), and so on. All these proteins contribute to detoxification processes, to defence against oxidative stress or repair of cell damage after exposure to xenobiotics.

In summary, knowing how microbial communities are affected by environmental alterations is an expanding field where new culture-independent high-throughput molecular tools have much to contribute. To overcome development limitations, a reliable post-alkaline protein extraction protocol has been developed and evaluated.

The proteins extracted were analysed by using an in-house database based on a previous 16S rRNA analysis which reported on alterations of the composition and on a decrease in diversity of a bacterial microbial community in response to pollution. Proteomic analysis made it possible to discover the global biological processes that are altered in response to the environmental changes: the synthesis of protective antioxidant/detoxification enzymes and molecular chaperones, ribosomal machinery for protein biosynthesis, active transport of a wide variety of substances and aerobic respiratory metabolism (glycolysis, TCA cycle and ATP synthesis). Aerobic respiration allows bacteria to cope with the huge demand for metabolic energy necessary to defend themselves against environmental threats, especially to ensure protein translation, membrane transport and protein homeostasis. The impact of pollutants on the abundance, structure, diversity and metabolic activities of soil bacterial populations could be used as potential indicators of their environmental toxicity and subsequent effects on the ecosystem. However, we are fully aware that there is still much room for improvement, with an urgent need for more curated and consistently annotated databases as well as more powerful bioinformatic software for big data analysis.

Experimental procedures

Experimental area and soil characteristics

Soils were sampled in May 2015 at three sites of DNP (SW Iberian Peninsula) or its surroundings (Fig. 1). LDP, located in the heart of Doñana, is a highly protected area that has usually been considered as a pristine ecosystem although several studies have nuanced this assumption (García-Sevillano *et al*., 2012; Michán *et al*., 2019). However, DNP is surrounded by many anthropogenic activities that threaten its ecological balance and environmental health (Abril *et al*., 2011; García-Sevillano *et al*., 2012; Fernández-Cisnal *et al*., 2014; García-Sevillano *et al*., 2014; Abril *et al*., 2015; Fernández-Cisnal *et al*., 2017; Michán *et al*., 2019). We studied two areas that border the National Park: in the northwest, AJO is close to extensive areas covered with strawberry greenhouses and vineyards and is also affected by wastes from several urban enclaves; in the east, MAT is affected by paddies and suffers the input of pesticides and metals used as algaecides during the rice-growing process (García-Sevillano *et al*., 2012; García-Sevillano *et al*., 2014; Abril *et al*., 2015). A summary of the records published of the pollutants determined at the sites studied can be found in Supplementary Table 1. Soils were sampled from the top 20 cm of the soil surface at least in quadruplicate and transported to the laboratory in plastic bags. After removing the plant remains, the soil samples were air-dried, ground, passed through a 2 mm sieve and finally kept at −30°C until protein extraction or analysis.

Physicochemical analysis of soils

The physicochemical characteristics of the soil samples were analysed in the 'Laboratorio Agroalimentario de Córdoba' (Junta de Andalucía). The organic matter was determined by oxidation with potassium dichromate ($K_2Cr_2O_7$) in an acid medium and titration of the excess dichromate with ferrous sulfate ($FeSO_4$) (United States Salinity Laboratory Staff, 1954). The organic nitrogen was determined using Kjeldahl's method as modified by Bremner (1965). The electrical conductivity and pH were measured in a 1/5 and 1/2.5 (wt./vol.) aqueous solution respectively. Soil particle-size analysis was performed by the Bouyoucos method (Day, 1965) and the textural classification was based on the soil texture triangle according to the USDA nomenclature (Soil Survey Division Staff, 1993).

Soil DNA extraction and 16S rRNA analysis

DNA was isolated using a protocol adapted from .Dong *et al*., (2006)). Briefly, 600 mg of soil samples were mixed thoroughly with 300 μl of 0.1 M phosphate buffer pH 6.6 and 200 μl of 0.1 M aluminium sulfate to precipitate humic substances. Then, the soil cells were disrupted adding 150 μl of 1 M NaOH, 350 mg of sterile sand and 300 μl of 0.1 M phosphate buffer pH 8.0, vortexed and further supplemented with 250 μl of SDS lysis buffer (100 mM NaCl, 500 mM Tris, 10% SDS pH 8.0). The samples were vortexed horizontally for 10 min and centrifuged at 10 000*g* for 30 s to remove debris. The DNA containing supernatant was transferred to a clean tube and treated with chloroform/isoamyl alcohol (24:1), precipitated with ammonium acetate + isopropanol and washed three times with 70% ethanol to clean the samples. Finally, the DNA was resuspended in sterile milliQ-water, quantified by spectrophotometry and visualized on a 1% agarose gel to check DNA integrity.

Before genomic analysis, all the samples were PCR amplified using specific bacteria 16S primers to discard the presence of PCR inhibitors in the DNA samples. The 16S rRNA sequencing was performed at the Genomic Unit of the Central Service for Research Support (SCAI) of the University of Córdoba using an Ion Torrent sequencer (PGM) and their specific Ion 16S™ Metagenomics Kit. To identify the microorganisms, the amplification results were analysed with the Ion Reporter™ 5.0 software Ion 16S Metagenomic analyses module. Primers V3 provided the maximal number of

identifications for all the samples and, thus, their amplicons were selected for the taxonomy assignments.

Alpha diversity was evaluated by determining richness and evenness at the family level. The Shannon–Wiener's (H') and Gini–Simpson's (D) indexes were applied to evaluate richness (abundance), as follows: $H'(\text{nats}) = -\sum \frac{n_i}{N} \ln \frac{n_i}{N} = -\sum Pi \ln Pi$, and $D = 1 - \sum Pi^2$, n_i is the number of reads of the ith families, N the total number of reads of all the families in the samples and $Pi = \frac{n_i}{N}$. Evenness, which implies equality in the number of reads of families, was estimated by the Pielou method (J') as: $J' = \frac{H'}{H'\max} = \frac{H'}{\ln S}$, where S is the number of families and $H'\max$ is the maximum expected diversity (Thukral, 2017).

Soil protein extraction

A novel reproducible alkaline-based protocol was developed and optimized here for the highly efficient extraction of proteins from soils. All the steps were carried out at room temperature. The soil samples in the extraction solution (0.5 M NaOH; 1:3, wt.:vol.) were incubated for 15 min, gently mixing by inversion each 5 min and centrifuged at 500*g* (1 min, 20°C). After 1-min sonication of the supernatant in an ultrasonic bath (Ultrasons 6, J.P. Selecta), it was centrifuged again as before. This step separates the bulk of the particulate matter in the pellet from the microbes that remain in suspension. The last supernatant was centrifuged at 25 000*g* (10 min, 20°C) to pellet the microorganisms. The cells were sequentially washed by resuspending the pellets in 1.5 ml of 50 mM Tris–HCl pH 7.5 buffer containing 10% glycerol followed by 1-min sonication and centrifugation as before. The washing conditions were optimized by evaluating the effect of the number of successive washings and the presence of 10% glycerol in the washing buffer on the efficiency of the soil protein extraction. For SDS-PAGE, Laemmli treatment buffer (20 μl) was added to lyse the final pellet of microbial cells and protein was extracted by a 5-min boiling step that was preceded and followed by a 1-min sonication step.

Electrophoretic separation of proteins

Microbial proteins extracted using Laemmli treatment buffer (20 μl) were loaded on 12.5% SDS-PAGE gels and separated in a Mini-PROTEAN® 3 Cell (Bio-Rad, Hercules, CA, USA). After electrophoresis, the gels were fluorescent stained with SYPRO Ruby® (Bio-Rad) and scanned using a ChemiDoc™ MP Imaging System (Bio-Rad) at 532/555 nm excitation/emission detection wavelengths. The gels were additionally stained using the least sensitive Coomassie Blue R-250 to compare the different soils. Image Lab software (version 4.1, Bio-Rad) was used for acquisition of the gel images and all subsequent image analyses.

Sample preparation, LC–MS/MS analyses and protein identification

All the analyses were performed at the Proteomic Unit at the Research Support Central Facility (SCAI) at the University of Cordoba. Microbial proteins extracted from 1 g of soil in Laemmli treatment buffer (30 μl) were loaded on 1 mm thick 8% SDS-PAGE gels and run until the front reached the start of the resolving gel. After Coomassie Blue staining, the protein bands were cut and in-gel tryptic digestion was performed as described previously (Fajardo *et al*., 2019).

The analysis of the peptides obtained from these digestions was carried out as described in a previous work (Alhama *et al*., 2018). LC–MS/MS analyses were performed in an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific) equipped with a nanoelectrospray source operating in positive mode. A Dionex Ultimate 3000 nano HPLC (Thermo Fisher Scientific) was used for peptide separation.

The raw data files were analysed with the Proteome Discoverer software (version 2.1.0.81, Thermo Fisher Scientific) including the SEQUEST HT algorithm. For the identification of the peptides, bacterial family databases were obtained from the UniProtKB repository updated on June 12th, 2020, selecting only those families with more than 30 reads in the previous genomic analysis. Search engine parameters were set as follows: up to one missed cleavage and cysteine carbamidomethylation and methionine oxidation were set as fixed and variable modifications respectively. Percolator was applied as the false discovery rate (FDR) validator. Peptide identifications were grouped into the same protein ID according to the law of parsimony and filtered to 1% FDR. A confidence score was calculated to select the best family databases by considering the number of proteins identified from ≥2 peptides and the average of coverage and of the number of peptides obtained from all the proteins identified, as follows: $\log_{10}[\text{Protein IDs } (\geq 2 \text{ peptides}) \times \overline{X}\ \text{coverage}]^{\overline{X}\ \text{peptides}}$. A database integrating all those from the bacterial families with a score higher than 10 was used for the metaproteomic analysis.

Statistics

The proteins identified were filtered not taking into consideration those that showed a covariance value above 25% in any of the soils studied (Microsoft Excel).

Then, protein quantification data were transformed to a logarithmic scale with base 2 for normalization. Statistical significance was evaluated using two-way ANOVA followed by *post hoc* multiple comparison according to Tukey, using Graphpad InStat software. Statistically significant differences are expressed as $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

Cluster and functional analysis

The Genesis package (Sturn *et al*., 2002) was used for the cluster analysis of the differentially expressed proteins. Intensity data were normalized and the distance measure employed was Pearson's correlation. Complete linkage hierarchical clustering and k-means clustering analyses were performed.

The online STRING v11 resource (https://string-db.org/) was used to find out the protein–protein structural and functional association networks (Szklarczyk *et al*., 2019). A functional enrichment analysis was carried out by downloading the list of differentially expressed proteins by names in the multiple proteins option and then searching using *Rhizobium* sp. LPU83 as the soil model organism selected. The network obtained was clustered to a Markov Cluster inflation parameter of 2.

Acknowledgements

This study was supported by the Spanish Ministry of Economy and Competitiveness (MINECO, CTM2016-75908-R), by the Spanish Ministry of Science and Innovation (MICINN, PID2019-110049RB-I00), by the Andalusian Plan of Research, Development, and Innovation (PAIDI) and by the University of Cordoba 'Plan Propio' program for group BIO187. A. M. Herruzo-Ruiz received a predoctoral contract of the University of Cordoba ('Plan Propio'). Funding for open access charge: Universidad de Córdoba / CBUA.

Abbreviations

AJO	Ajolí
DNP	Doñana National Park
FDR-adjusted *p*	false discovery rate
LDP	Lucio del Palacio
MAT	Matochal
PPI	protein–protein interaction

References

Abril, N., Chicano-Gálvez, E., Michán, C., Pueyo, C., and López-Barea, J. (2015) iTRAQ analysis of hepatic proteins in free-living *Mus spretus* mice to assess the contamination status of areas surrounding Doñana National Park (SW Spain). *Sci Total Environ* **523**: 16–27.

Abril, N., Ruiz-Laguna, J., Garcia-Sevillano, M.A., Mata, A. M., Gomez-Ariza, J.L., and Pueyo, C. (2014) Heterologous microarray analysis of transcriptome alterations in *Mus spretus* mice living in an industrial settlement. *Environ Sci Technol* **48**: 2183–2192.

Abril, N., Ruiz-Laguna, J., Osuna-Jiménez, I., Vioque-Fernández, A., Fernández-Cisnal, R., Chicano-Gálvez, E., *et al*. (2011) Omic approaches in environmental issues. *J Toxicol Environ Health A* **74**: 1001–1019.

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., and Narasimhan, G. (2016) Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. *Evol Bioinform Online* **12**: 5–16.

Alhama, J., Fuentes-Almagro, C.A., Abril, N., and Michán, C. (2018) Alterations in oxidative responses and post-translational modification caused by *p,p′*-DDE in *Mus spretus* testes reveal Cys oxidation status in proteins related to cell-redox homeostasis and male fertility. *Sci Total Environ* **636**: 656–669.

Bacmaga, M., Kucharski, J., and Wyszkowska, J. (2019) Microbiological and biochemical properties of soil polluted with a mixture of spiroxamine, tebuconazole, and triadimenol under the cultivation of *Triticum aestivum* L. *Environ Monit Assess* **191**: 416.

Bacmaga, M., Wyszkowska, J., and Kucharski, J. (2018) The influence of chlorothalonil on the activity of soil microorganisms and enzymes. *Ecotoxicology* **27**: 1188–1202.

Basavanna, S., Khandavilli, S., Yuste, J., Cohen, J.M., Hosie, A.H.F., Webb, A.J., *et al*. (2009) Screening of *Streptococcus pneumoniae* ABC transporter mutants demonstrates that LivJHMGF, a branched-chain amino acid ABC transporter, is necessary for disease pathogenesis. *Infect Immun* **77**: 3412–3423.

Bastida, F., Hernández, T., and García, C. (2014) Metaproteomics of soils from semiarid environment: functional and phylogenetic information obtained with different protein extraction methods. *J Proteomics* **101**: 31–42.

Bastida, F., and Jehmlich, N. (2016) It's all about functionality: how can metaproteomics help us to discuss the attributes of ecological relevance in soil? *J Proteomics* **144**: 159–161.

Beaumelle, L., De Laender, F., and Eisenhauer, N. (2020) Biodiversity mediates the effects of stressors but not nutrients on litter decomposition. *Elife* **9**: e55659.

Becher, D., Bernhardt, J., Fuchs, S., and Riedel, K. (2013) Metaproteomics to unravel major microbial players in leaf litter and soil environments: challenges and perspectives. *Proteomics* **13**: 2895–2909.

Benndorf, D., Balcke, G.U., Harms, H., and von Bergen, M. (2007) Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *ISME J* **1**: 224–234.

Bona, E., Massa, N., Novello, G., Boatti, L., Cesaro, P., Todeschini, V., *et al*. (2019) Metaproteomic characterization of the *Vitis vinifera* rhizosphere. *FEMS Microbiol Ecol* **95**: fiy204.

Bonilla-Valverde, D., Ruiz-Laguna, J., Muñoz, A., Ballesteros, J., Lorenzo, F., Gómez-Ariza, J.L., and López-Barea, J. (2004) Evolution of biological effects of Aznalcollar mining spill in the Algerian mouse (*Mus spretus*) using biochemical biomarkers. *Toxicology* **197**: 123–138.

Bouchez, T., Blieux, A.L., Dequiedt, S., Domaizon, I., Dufresne, A., Ferreira, S., *et al*. (2016) Molecular microbiology methods for environmental diagnosis. *Environ Chem Lett* **14**: 423–441.

Bremner, J.M. (1965) Total nitrogen. In *Methods of Soil Analysis*, Black, C.A. (ed). Madison, Wisconsin: American Society of Agronomy, pp. 1149–1178.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., *et al*. (2012) Biodiversity loss and its impact on humanity. *Nature* **486**: 59–67.

Chapin, F.S., III, Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., *et al*. (2000) Consequences of changing biodiversity. *Nature* **405**: 234–242.

Chen, J., Li, J., Zhang, H., Shi, W., and Liu, Y. (2019) Bacterial heavy-metal and antibiotic resistance genes in a copper tailing dam area in northern China. *Front Microbiol* **10**: 1916.

Chen, S., Rillig, M.C., and Wang, W. (2009) Improving soil protein extraction for metaproteome analysis and glomalin-related soil protein detection. *Proteomics* **9**: 4970–4973.

Chiapello, M., Zampieri, E., and Mello, A. (2020) A small effort for researchers, a big gain for soil metaproteomics. *Front Microbiol* **11**: 88.

Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K.L., Tom, L.M., *et al*. (2010) Direct cellular lysis/protein extraction protocol for soil metaproteomics. *J Proteome Res* **9**: 6615–6622.

Czaplicki, L.M., and Gunsch, C.K. (2016) Reflection on molecular approaches influencing state-of-the-art bioremediation design: culturing to microbial community fingerprinting to omics. *J Environ Eng (New York)* **142**: 10.1061. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001141.

Davidson, A.L., and Chen, J. (2004) ATP-binding cassette transporters in bacteria. *Annu Rev Biochem* **73**: 241–268.

Dawson, R.J.P., and Locher, K.P. (2006) Structure of a bacterial multidrug ABC transporter. *Nature* **443**: 180–185.

Day, P.R. (1965) Particle fractionation and particle-size analysis. In *Methods of Soil Analysis*, Black, C.A. (ed). Madison, Wisconsin: American Society of Agronomy, pp. 545–546.

Decaëns, T., Jiménez, J.J., Gioia, C., Measey, G.J., and Lavelle, P. (2006) The values of soil animals for conservation biology. *Eur J Soil Biol* **42**: S23–S38.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-Gonzalez, A., Eldridge, D.J., Bardgett, R.D., *et al*. (2018) A global atlas of the dominant bacteria found in soil. *Science* **359**: 320–325.

Ding, Z., Fang, Q., Daraz, U., and Sun, Q. (2021) Physiological responses and metal distributions of different organs of *Phragmites australis* shoots under acid mine drainage stress. *Environ Sci Pollut Res Int* **28**: 3375–3385.

Dominati, E., Patterson, M., and Mackay, A. (2010) A framework for classifying and quantifying the natural capital and ecosystem services of soils. *Ecol Econ* **69**: 1858–1868.

Dong, D., Yan, A., Liu, H., Zhang, X., and Xu, Y. (2006) Removal of humic substances from soil DNA using aluminium sulfate. *J Microbiol Methods* **66**: 217–222.

Ellis, H.R. (2010) The FMN-dependent two-component monooxygenase systems. *Arch Biochem Biophys* **497**: 1–12.

Fajardo, C., Amil-Ruiz, F., Fuentes-Almagro, C., De Donato, M., Martinez-Rodriguez, G., Escobar-Nino, A., *et al*. (2019) An "omic" approach to *Pyrocystis lunula*: new insights related with this bioluminescent dinoflagellate. *J Proteomics* **209**: 103502.

Fatimawali., Kepel, B.J., Gani, M.A., and Tallei, T.E. (2020). Comparison of bacterial community structure and diversity in traditional gold mining waste disposal site and Rice field by using a Metabarcoding approach. *Int J Microbiol* 1858732.

Feld, L., Hjelmso, M.H., Nielsen, M.S., Jacobsen, A.D., Ronn, R., Ekelund, F., *et al*. (2015) Pesticide side effects in an agricultural soil ecosystem as measured by amoA expression quantification and bacterial diversity changes. *PLoS One* **10**: e0126080.

Feng, G., Xie, T., Wang, X., Bai, J., Tang, L., Zhao, H., *et al*. (2018) Metagenomic analysis of microbial community and function involved in cd-contaminated soil. *BMC Microbiol* **18**: 11.

Fernández-Cisnal, R., Alhama, J., Abril, N., Pueyo, C., and López-Barea, J. (2014) Redox proteomics as biomarker for assessing the biological effects of contaminants in crayfish from Doñana National Park. *Sci Total Environ* **490**: 121–133.

Fernández-Cisnal, R., García-Sevillano, M.A., García-Barrera, T., Gómez-Ariza, J.L., and Abril, N. (2018) Metabolomic alterations and oxidative stress are associated with environmental pollution in *Procambarus clarkii*. *Aquat Toxicol* **205**: 76–88.

Fernández-Cisnal, R., García-Sevillano, M.A., Gómez-Ariza, J.L., Pueyo, C., López-Barea, J., and Abril, N. (2017) 2D-DIGE as a proteomic biomarker discovery tool in environmental studies with *Procambarus clarkii*. *Sci Total Environ* **584-585**: 813–827.

Fierer, N. (2017) Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol* **15**: 579–590.

Franzosa, E.A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X.C., and Huttenhower, C. (2015) Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat Rev Microbiol* **13**: 360–372.

Gago-Tinoco, A., González-Domínguez, R., García-Barrera, T., Blasco-Moreno, J., Bebianno, M.J., and Gómez-Ariza, J.L. (2014) Metabolic signatures associated with environmental pollution by metals in Doñana National Park using *P. clarkii* as bioindicator. *Environ Sci Pollut Res Int* **21**: 13315–13323.

García-Sevillano, M.A., García-Barrera, T., Abril, N., Pueyo, C., López-Barea, J., and Gómez-Ariza, J.L. (2014) Omics technologies and their applications to evaluate metal toxicity in mice *M. spretus* as a bioindicator. *J Proteomics* **104**: 4–23.

García-Sevillano, M.A., González-Fernández, M., Jara-Biedma, R., García-Barrera, T., López-Barea, J., Pueyo, C., and Gómez-Ariza, J.L. (2012) Biological response of free-living mouse *Mus spretus* from Doñana National Park under environmental stress based on assessment of metal-binding biomolecules by SEC-ICP-MS. *Anal Bioanal Chem* **404**: 1967–1981.

Greenfield, L.M., Hill, P.W., Paterson, E., Baggs, E.M., and Jones, D.L. (2018) Methodological bias associated with soluble protein recovery from soil. *Sci Rep* **8**: 11186.

Grimalt, J.O., Ferrer, M., and Macpherson, E. (1999) The mine tailing accident in Aznalcollar. *Sci Total Environ* **242**: 3–11.

Gunnigle, E., Ramond, J.B., Frossard, A., Seeley, M., and Cowan, D. (2014) A sequential co-extraction method for DNA, RNA and protein recovery from soil for future system-based approaches. *J Microbiol Methods* **103**: 118–123.

Harrison, X.A., and Cameron, S.J.S. (2020) Analytical approaches for microbiome research. In *Microbiomes of Soils, Plants and Animals: An Integrated Approach*, Antwis, R.E., Harrison, X.A., and Cox, M.J. (eds). Cambridge, UK: Cambridge University Press, pp. 8–28.

He, T., Li, H., and Zhang, X. (2017) Deep-sea hydrothermal vent viruses compensate for microbial metabolism in virus-host interactions. *mBio* **8**: e00893-17.

Hermans, S.M., Buckley, H.L., Case, B.S., Curran-Cournane, F., Taylor, M., and Lear, G. (2017) Bacteria as emerging indicators of soil condition. *Appl Environ Microbiol* **83**: e02826-16.

Hettich, R.L., Sharma, R., Chourey, K., and Giannone, R.J. (2012) Microbial metaproteomics: identifying the repertoire of proteins that microorganisms use to compete and cooperate in complex environmental communities. *Curr Opin Microbiol* **15**: 373–380.

Hofbauer, S., Schaffner, I., Furtmüller, P.G., and Obinger, C. (2014) Chlorite dismutases - a heme enzyme family for use in bioremediation and generation of molecular oxygen. *Biotechnol J* **9**: 461–473.

Hosie, A.H., Allaway, D., Galloway, C.S., Dunsby, H.A., and Poole, P.S. (2002) *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (bra/LIV) of the ABC family. *J Bacteriol* **184**: 4071–4080.

Jeckelmann, J.M., and Erni, B. (2020) Transporters of glucose and other carbohydrates in bacteria. *Pflugers Arch - Eur J Physiol* **472**: 1129–1153.

Jiao, S., Chen, W., and Wei, G. (2019) Resilience and assemblage of soil microbiome in response to chemical contamination combined with plant growth. *Appl Environ Microbiol* **85**: e02523-02518.

Keiblinger, K.M., Wilhartitz, I.C., Schneider, T., Roschitzki, B., Schmid, E., Eberl, L., *et al*. (2012) Soil metaproteomics - comparative evaluation of protein extraction protocols. *Soil Biol Biochem* **54**: 14–24.

Kushnirov, V.V. (2000) Rapid and reliable protein extraction from yeast. *Yeast* **16**: 857–860.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R. (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* **82**: 6955–6959.

Liu, J., Meng, Z., Liu, X., and Zhang, X.-H. (2019) Microbial assembly, interaction, functioning, activity and diversification: a review derived from community compositional data. *Mar Life Sci Technol* **1**: 112–128.

Liu, Y.R., Delgado-Baquerizo, M., Bi, L., Zhu, J., and He, J. Z. (2018) Consistent responses of soil microbial taxonomic and functional attributes to mercury pollution across China. *Microbiome* **6**: 183.

Lloyd, K.G., Steen, A.D., Ladau, J., Yin, J., and Crosby, L. (2018) Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *mSystems* **3**: e00055-00018.

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J.P., Hector, A., *et al*. (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**: 804–808.

Luo, Z., Ma, J., Chen, F., Li, X., and Zhang, S. (2018) Effects of Pb smelting on the soil bacterial community near a secondary Lead plant. *Int J Environ Res Public Health* **15**: 1030.

Mandalakis, M., Panikov, N.S., Polymenakou, P.N., Sizova, M.V., and Stamatakis, A. (2018) A simple cleanup method for the removal of humic substances from soil protein extracts using aluminum coagulation. *Environ Sci Pollut Res Int* **25**: 23845–23856.

Margerison, R., Nicolitch, O., and Zhang, Y. (2020) Microbiomes of soils. In *Microbiomes of Soils, Plants and Animals*, Antwis, R.E., Harrison, X.A., and Cox, M.J. (eds). Cambridge, UK: Cambridge University Press, pp. 29–54.

Maron, P.A., Mougel, C., and Ranjard, L. (2011) Soil microbial diversity: methodological strategy, spatial overview and functional interest. *C R Biol* **334**: 403–411.

Maron, P.A., Sarr, A., Kaisermann, A., Leveque, J., Mathieu, O., Guigue, J., *et al*. (2018) High microbial diversity promotes soil ecosystem functioning. *Appl Environ Microbiol* **84**: e02738-17.

Matsuo, Y., Asakawa, K., Toda, T., and Katayama, S. (2006) A rapid method for protein extraction from fission yeast. *Biosci Biotechnol Biochem* **70**: 1992–1994.

Mattarozzi, M., Manfredi, M., Montanini, B., Gosetti, F., Sanangelantoni, A.M., Marengo, E., *et al*. (2017) A metaproteomic approach dissecting major bacterial functions in the rhizosphere of plants living in serpentine soil. *Anal Bioanal Chem* **409**: 2327–2339.

Michán, C., Chicano-Gálvez, E., Fuentes-Almagro, C.A., and Alhama, J. (2019) Redox and global interconnected proteome changes in mice exposed to complex environmental hazards surrounding Donana National Park. *Environ Pollut* **252**: 427–439.

Myrold, D.D., Zeglin, L.H., and Jansson, J.K. (2013) The potential of metagenomic approaches for understanding soil microbial processes. *Soil Sci Soc Am* **78**: 3–10.

Onwona-Kwakye, M., Plants-Paris, K., Keita, K., Lee, J., Brink, P., Hogarh, J.N., and Darkoh, C. (2020) Pesticides decrease bacterial diversity and abundance of irrigated Rice fields. *Microorganisms* **8**: 318.

Puglisi, E., Cahill, M.J., Lessard, P.A., Capri, E., Sinskey, A. J., Archer, J.A., and Boccazzi, P. (2010) Transcriptional response of *Rhodococcus aetherivorans* I24 to polychlorinated biphenyl-contaminated sediments. *Microb Ecol* **60**: 505–515.

Pylro, V.S., Fulthorpe, R., and Roesch, L.F.W. (2020) Editorial: microbe-mediated processes in soils. *Front Environ Sci* **8**: 124.

Qian, C., and Hettich, R.L. (2017) Optimized extraction method to remove humic acid interferences from soil

samples prior to microbial proteome measurements. *J Proteome Res* **16**: 2537–2546.

Rajendhran, J., and Gunasekaran, P. (2008) Strategies for accessing soil metagenome for desired applications. *Biotechnol Adv* **26**: 576–590.

Renu, Gupta, S.K., Rai, A.K., Sarim, K.M., Sharma, A., Budhlakoti, N., *et al*. (2019) Metaproteomic data of maize rhizosphere for deciphering functional diversity. *Data Brief* **27**: 104574.

Reyes-Hernández, O.D., Mejía-García, A., Sánchez-Ocampo, E.M., Castro-Muñozledo, F., Hernández-Muñoz, R., and Elizondo, G. (2009) Aromatic hydrocarbons upregulate glyceraldehyde-3-phosphate dehydrogenase and induce changes in Actin cytoskeleton. Role of the aryl hydrocarbon receptor (AhR). *Toxicology* **266**: 30–37.

Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* **38**: 525–552.

Ruiz-Laguna, J., Vélez, J.M., Pueyo, C., and Abril, N. (2016) Global gene expression profiling using heterologous DNA microarrays to analyze alterations in the transcriptome of *Mus spretus* mice living in a heavily polluted environment. *Environ Sci Pollut Res Int* **23**: 5853–5867.

Sharma, P.K., Sharma, V., Sharma, S., Bathia, G., Singh, C., and Sharma, R. (2019) Comparative metatranscriptome analysis revealed broad response of microbial communities in two soil types, agriculture versus organic soil. *J Genet Eng Biotechnol* **17**: 6.

Siggins, A., Gunnigle, E., and Abram, F. (2012) Exploring mixed microbial community functioning: recent advances in metaproteomics. *FEMS Microbiol Ecol* **80**: 265–280.

Singh, J., Behal, A., Singla, N., Joshi, A., Birbian, N., Singh, S., *et al*. (2009) Metagenomics: concept, methodology, ecological inference and recent advances. *Biotechnol J* **4**: 480–494.

Soil Survey Division Staff. (1993) Soil texture triangle according to the USDA nomenclature. In *Soil Survey Manual USDA Handbook No 18*. Washington, DC: US Government Printing Office.

Starke, R., Jehmlich, N., and Bastida, F. (2019) Using proteins to study how microbes contribute to soil ecosystem services: the current state and future perspectives of soil metaproteomics. *J Proteomics* **198**: 50–58.

Storck, V., Nikolaki, S., Perruchon, C., Chabanis, C., Sacchi, A., Pertile, G., *et al*. (2018) Lab to field assessment of the ecotoxicological impact of Chlorpyrifos, Isoproturon, or Tebuconazole on the diversity and composition of the soil bacterial community. *Front Microbiol* **9**: 1412.

Sturn, A., Quackenbush, J., and Trajanoski, Z. (2002) Genesis: cluster analysis of microarray data. *Bioinformatics* **18**: 207–208.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., *et al*. (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**: D607–D613.

Tang, L., Yao, A., Yuan, M., Tang, Y., Liu, J., Liu, X., and Qiu, R. (2016) Transcriptional up-regulation of genes involved in photosynthesis of the Zn/cd hyperaccumulator *sedum alfredii* in response to zinc and cadmium. *Chemosphere* **164**: 190–200.

Taylor, E.B., and Williams, M.A. (2010) Microbial protein in soil: influence of extraction method and C amendment on extraction and recovery. *Microb Ecol* **59**: 390–399.

Techtmann, S.M., and Hazen, T.C. (2016) Metagenomic applications in environmental monitoring and bioremediation. *J Ind Microbiol Biotechnol* **43**: 1345–1354.

Thakur, M.P., Phillips, H.R.P., Brose, U., De Vries, F.T., Lavelle, P., Loreau, M., *et al*. (2020) Towards an integrative understanding of soil biodiversity. *Biol Rev Camb Philos Soc* **95**: 350–364.

Thomas, J.C., IV, Oladeinde, A., Kieran, T.J., Finger, J.W., Jr., Bayona-Vasquez, N.J., Cartee, J.C., *et al*. (2020) Co-occurrence of antibiotic, biocide, and heavy metal resistance genes in bacteria from metal and radionuclide contaminated soils at the Savannah River site. *J Microbial Biotechnol* **13**: 1179–1200.

Thorn, C.E., Bergesch, C., Joyce, A., Sambrano, G., McDonnell, K., Brennan, F., *et al*. (2018) A robust, cost-effective method for DNA, RNA and protein co-extraction from soil, other complex microbiomes and pure cultures. *Mol Ecol Resour* **19**: 439–455.

Thukral, A.K. (2017) A review of measurement of alpha diversity in biology. *Agric Res J* **54**: 1–10.

Tibbett, M., Fraser, T.D., and Duddigan, S. (2020) Identifying potential threats to soil biodiversity. *PeerJ* **8**: e9271.

Torsvik, V., and Ovreas, L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5**: 240–245.

United States Salinity Laboratory Staff. (1954) Diagnosis and improvement of saline and alkaline soils. In *Agriculture Handbook No 60*, Richards, L.A. (ed). Washington, DC: United States Department of Agriculture, pp. 105–106.

Ventorino, V., Pascale, A., Adamo, P., Rocco, C., Fiorentino, N., Mori, M., *et al*. (2018) Comparative assessment of autochthonous bacterial and fungal communities and microbial biomarkers of polluted agricultural soils of the terra dei Fuochi. *Sci Rep* **8**: 14281.

Vioque-Fernández, A., Alves de Almeida, E., and López-Barea, J. (2009) Assessment of Doñana National Park contamination in *Procambarus clarkii*: integration of conventional biomarkers and proteomic approaches. *Sci Total Environ* **407**: 1784–1797.

Vioque-Fernández, A., de Almeida, E.A., Ballesteros, J., García-Barrera, T., Gómez-Ariza, J.L., and López-Barea, J. (2007) Doñana National Park survey using crayfish (*Procambarus clarkii*) as bioindicator: esterase inhibition and pollutant levels. *Toxicol Lett* **168**: 260–268.

Wang, D.Z., Kong, L.F., Li, Y.Y., and Xie, Z.X. (2016) Environmental microbial community proteomics: status, challenges and perspectives. *Int J Mol Sci* **17**: 1275.

Weissbach, H., Etienne, F., Hoshi, T., Heinemann, S.H., Lowther, W.T., Matthews, B., *et al*. (2002) Peptide methionine sulfoxide reductase: structure, mechanism of action, and biological function. *Arch Biochem Biophys* **397**: 172–178.

Xie, Y., Fan, J., Zhu, W., Amombo, E., Lou, Y., Chen, L., and Fu, J. (2016) Effect of heavy metals pollution on soil microbial diversity and Bermudagrass genetic variation. *Front Plant Sci* **7**: 755.

Yin, H., Niu, J., Ren, Y., Cong, J., Zhang, X., Fan, F., *et al*. (2015) An integrated insight into the response of sedimentary microbial communities to heavy metal contamination. *Sci Rep* **5**: 14266.

Zampieri, E., Chiapello, M., Daghino, S., Bonfante, P., and Mello, A. (2016) Soil metaproteomics reveals an inter-kingdom stress response to the presence of black truffles. *Sci Rep* **6**: 25773.

Zarraonaindia, I., Smith, D.P., and Gilbert, J.A. (2013) Beyond the genome: community-level analysis of the microbial world. *Biol Philos* **28**: 261–282.

Zhao, X., Huang, J., Lu, J., and Sun, Y. (2019) Study on the influence of soil microbial community on the long-term heavy metal pollution of different land use types and depth layers in mine. *Ecotoxicol Environ Saf* **170**: 218–226.

Zhao, X., Huang, J., Zhu, X., Chai, J., and Ji, X. (2020) Ecological effects of heavy metal pollution on soil microbial community structure and diversity on both sides of a river around a mining area. *Int J Environ Res Public Health* **17**: 5680.

Zhou, J., He, Z., Yang, Y., Deng, Y., Tringe, S.G., and Alvarez-Cohen, L. (2015) High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *mBio* **6**: e02288-02214.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Supporting Information.