Bastida, F., **Jehmlich, N.**, Torres, I.F., García, C. (2018):
The extracellular metaproteome of soils under semiarid climate: A methodological comparison of extraction buffers
Sci. Total Environ. **619–620** , 707 – 711

http://dx.doi.org/10.1016/j.scitotenv.2017.11.134

TITLE PAGE

Title:

The extracellular metaproteome of soils under semiarid climate: a methodological comparison of extraction buffers

Authors:

F. Bastida[a,^,*], N. Jehmlich[b,^], I.F, Torres[a], C. García[a]

[^]equally contributed

Affiliations:

[a]CEBAS-CSIC, Department of Soil and Water Conservation, campus Universitario de Espinardo, 30100 Murcia, Spain

[b]Helmholtz-Centre for Environmental Research, UFZ, Department of Molecular Systems Biology, Permoserstr. 15, 04318 Leipzig, Germany

[^]equally contributed

*Corresponding author: Dr. Felipe Bastida (fbastida@cebas.csic.es); Phone: +34 968396106; Fax: +34 968396213

Present address: (F. Bastida): [a]CEBAS-CSIC, Department of Soil and Water Conservation, campus Universitario de Espinardo, 30100 Murcia, Spain

Article type: Short-Communication

The extracellular metaproteome of soils under semiarid climate: a methodological comparison of extraction buffers

F. Bastida, N. Jehmlich, I.F, Torres, C. García

Abstract

We compare the protein extraction efficiencies, as well as the phylogenetic and functional information provided, of two extraction protocols in soils that differ mainly in their organic matter and clay contents, the main factors limiting protein extraction in semiarid soils. These protocols utilise extractants commonly used for the assay of extracellular enzyme activities. The first method was based on the utilisation of the modified universal buffer (MUB). The second was based on the extraction of humic substances with sodium pyrophosphate. When compared to the total amount of proteins in soil, the results indicate a very-low extraction efficiency for both protocols. Analysis in an Orbitrap Fusion mass spectrometer and further searching against an "*ad hoc*" metagenome evidenced that the phylogenetic and functional information retrieved from the extracellular soil metaproteome can be biased by the extraction buffer.

Keywords: extraction; functionality; protein diversity; extracellular proteins; soil metaproteomics

The direct identification of soil proteins may augment the information available about the activity and function of microbial populations, and their connection to soil ecosystem services (C and N cycles, gas fluxes, plant growth, etc.). In this respect, the development and accuracy of new mass spectrometers coupled to nano-HPLC systems, together with improved extraction methods and the implementation of genomic databases, have fostered recent soil metaproteomics studies [1-6] and have provided unprecedented insights into soil microbial ecology [7]. However, despite this progress, a major gap in soil metaproteomics is still the identification of extracellular proteins that are directly or indirectly related to ecosystem processes such as organic matter mineralisation and CO_2 efflux [7-8]. These extracellular proteins remain cryptic and hard to extract due to their physico-chemical interaction with organic matter and soil mineral particles [9]. An additional flaw in soil metaproteomics is that the yield of protein extraction is frequently not assessed properly. It is known that colorimetric and fluorometric reactions do not provide a reliable indication of protein content [10], while the quantification of amino acids released by acid hydrolysis is a straightforward approach to the quantification of the proteins extracted [3].

We hypothesised that extracellular proteins should be able to be identified in extracts that are commonly utilised to assay the extracellular enzyme activity, once the cells have been eliminated by filtering. The removal of the cellular fraction reduces the complexity of the metaproteome and may enhance the probability of extracellular protein identification. Here, we critically compare the protein extraction efficiencies, as well as the phylogenetic and functional information provided, of two extraction protocols used in soils developed under semiarid climate. .

Since the organic C and clay contents are factors that strongly influence the extraction and identification of proteins [9], we selected soil samples that varied greatly in these properties. Soil samples were collected in two natural semiarid areas in Southeastern Spain (Bastida et al. 2016 [6]; Table S1). The first is “Barranco de Gebas”, a pre-

desertic, badland area developed on a marsh substrate. The soil type is Gypsic xerosol (FAO, 1998). Here, two soil samples were taken, at locations GEB (37º54'38''N; 1º24'7''W) and GEP (37º54'15''N; 1º24'38''W): GEB had a low plant cover of 5%, composed mainly of *Rosmarinus officinalis*, and GEP had a plant cover of 40%, dominated by *Pinus halepensis* and *Stipa tenacissima*. The second area is "Sierra Espuña", a semiarid forest area. The soil of this area is classified as a Calcaric regosol (FAO, 1998). Two soil samples were taken in this area, both at locations with nearly 100% plant cover: SE1 (37º51'25''N; 1º28'13''W) was dominated by *P. halepensis* and SE5 (37º52'4''N; 1º33'56''W) was dominated by *Quercus rotundifolia*.

Soil sampling was performed in May 2015, with three replicates (n=3) per sample. For each replicate, six soil sub-samples were taken at a depth of 0-20 cm, after removal of litter, and were mixed to obtain one composite sample per replicate. The samples were sieved (2 mm) and kept at 4ºC for chemical and enzymatic analyses, and at -20ºC for proteomic analyses.

The GEB and GEP samples had lower organic C concentrations (3.3 and 2.3 g 100 g^{-1}, respectively) than the SE1 and SE5 samples (6.5 and 17.4 g 100 g^{-1}, respectively). The SE1 and SE5 samples contained lower percentages of clay (19.9 and 17.3 g 100 g^{-1}, respectively) than the GEB and GEP samples (44.7 and 41.5 g 100 g^{-1}, respectively) (Table S1). A biochemical and microbiological characterisation of the samples is available in the Supporting Information.

The total nitrogen (N) and total organic carbon (TOC) concentrations were determined using an Elemental Analyzer (C/N Flash EA 112 Series-Leco Truspec). Enzyme activities were analysed as described elsewhere [6] (Supporting Information; Table S1). The DNA from the soil samples taken in this study, together with that of other semiarid soils [6], was utilised for the preparation of a semiarid soil metagenome (Supporting Information).

111 Two protein extraction methods were selected. These methods utilise extractants
112 commonly used for the assay of extracellular enzyme activities. The first extraction
113 method was based on the utilisation of modified universal buffer (MUB), composed of
114 tris (hydroxymethyl) aminomethane, maleic acid, citric acid, boric acid, NaOH, HCl and
115 distilled water at pH 6.5 [11]. The second extraction protocol was based on the
116 extraction of humic substances with 0.1 M sodium pyrophosphate, at pH 7.1 [12-13].
117 Five grams of soil were mixed with 20 ml of extractant and shaken gently for 1 h. No
118 detergents or ultrasonication were used. Afterwards, the soil suspensions were
119 centrifuged for 15 min at 13000 rpm and 4ºC. Each supernatant was filtered through a
120 0.22-µm filter, to eliminate cells. In order to pellet the proteins, tri-chloro acetic acid
121 (TCA) was added to the cell-free suspension, at a final concentration of 25%. The
122 mixtures were incubated for 10 min at -18ºC and then overnight at 4ºC. Then, they
123 were centrifuged for 10 min at 13000 rpm and 4ºC. The protein-enriched pellets
124 obtained were subjected to further purification [1,3]. One millilitre of chilled 100%
125 acetone was added to the protein pellets, followed by sonication in a bath and
126 incubation for 30 min at -18ºC. Then, a purified protein pellet was obtained by
127 repeating the centrifugation step; the supernatant was discarded. The acetone washing
128 was then repeated. Finally, the protein pellets were dried in an oven at 25ºC for 15 min
129 and N-rich air was fluxed to the vials to avoid oxidation of amino acid residues.
130 The analysis of the amino acid contents in the protein pellets and bulk soils was
131 initiated by acid hydrolysis with 6 N HCl, for 22 h at 110ºC, using norleucine as
132 standard. The amino acids were analysed in a Biochrom 30 series Amino Acid
133 Analyser, based on continuous flow chromatography, following the manufacturer's
134 indications.
135 Tryptic digestion of protein pellets, chromatographic separation of peptides and mass
136 spectrometric analysis were performed as described in Bastida et al. (2016) [6] (see
137 Supporting Information). The "PROteomics results Pruning & Homology group
138 ANotation Engine" (PROPHANE) (http://www.prophane.de) was applied to assign

139 proteins to their phylogenetic and functional origins. The diversity of the active
140 microbial community was calculated as the Shannon-Wiener index [14], based on
141 NSAF values at the genus level. Proteins in the following categories were computed
142 from the extracellular metaproteome: i) *Transporters* involved in the transport of
143 carbohydrates, amino acids, peptides, phosphate and cyclic compounds; ii) proteins
144 involved in redox and *ROS* reactions (catalases and superoxide dismutase, involved in
145 the oxidation of soil organic matter); and iii) *Hydrolases* involved in the biodegradation
146 of carbohydrates (cellobiohydrolase, pectate lyase), proteases, peptidases, lipases,
147 etc. (Table 1; Table S2).
148 The normality and homogeneity of variance of the variables were tested by the
149 Kolmogorov-Smirnov and Levene tests, respectively. In order to determine pairwise
150 differences by *post-hoc* tests, the chemical indicators of the bulk soil were subjected to
151 one-way ANOVA. The protein extraction efficiency and metaproteomic indicators
152 (diversity, percentage of transporters, percentage of ROS and percentage of
153 hydrolases) were subjected to a two-way ANOVA. The two factors included in this
154 experimental design were soil sample and extraction buffer. Differences at $P<0.05$
155 were regarded as statistically significant. The Shannon-Wiener index of diversity was
156 calculated for the diversity of the microbial community based on the taxonomic analysis
157 of proteins at the genus level. The structure of the microbial community was analysed
158 by principal component analysis with the relative abundances at the order level.
159 PERMANOVA was applied to test the significance of the analysed factors with regard
160 to the structure of the microbial communities. SPSS v22.0 and R software v.3.1.3 were
161 used for the statistics.
162
163 The protein concentrations of the GEB and GEP bulk soil samples were similar (Table
164 S1), while that of SE1 (8.3 ± 1.3 mg protein g^{-1} soil) was higher than that of GEB (4.2 ±
165 0.7 mg protein g^{-1} soil) and GEP (4.4 ± 0.7 mg protein g^{-1} soil). The protein
166 concentration was highest in soil SE5 (16.7 ± 4.3 mg protein g^{-1} soil).

To estimate the yield of the protein extraction, the amino acids in the extracted pellets were expressed as a percentage (%) of the total amino acid content of the bulk soil samples. The extraction efficiency was influenced by both soil sample and extraction buffer (Table 1). The protein extraction yield and protein content in the pellets were 10-times higher when proteins were extracted with sodium pyrophosphate, relative to MUB. The extraction yield was always highest for SE1 and reached a maximum of 0.28% of the total protein content in bulk soil when sodium pyrophosphate was used as the extractant. The total protein content of the pellets was highest for SE5, in agreement with the higher protein content of this soil sample.

Differences in the relative abundance of microbial orders were observed (Fig. 1A, Supporting Information). The PERMANOVA revealed significant effects of both soil sample and extraction buffer on the diversity and structure of the microbial community (P<0.001) (Fig. 1B; Table 1). The diversity of proteins extracted with MUB was higher than that extracted with sodium pyrophosphate, for each sample. The protein diversity was lowest in GEB, for both extraction methods (Table 1).

Table 1. Protein extraction efficiency, protein diversity and functional protein groups

	GEB		**GEP**		**SE1**		**SE5**		**ANOVA**		
	Mean	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*		F	*P* value
% Protein extraction									S	11.05	0.003
MUB	**0.013**	0.004	**0.013**	0.001	**0.020**	0.001	**0.008**	0.001	E	223.50	*
PYR	**0.10**	0.015	**0.17**	0.007	**0.28**	0.060	**0.18**	0.017	SxE	9.34	0.005
µg prot extr g^{-1} soil									S	40.68	*
MUB	**0.53**	0.18	**0.58**	0.03	**1.66**	0.11	**1.36**	0.19	E	227.94	*
PYR	**4.22**	0.64	**7.64**	0.29	**22.92**	4.92	**30.48**	2.86	SxE	35.03	*
Diversity									S	56.31	*
MUB	**2.37**	0.07	**3.48**	0.26	**3.53**	0.06	**3.74**	0.04	E	162.38	*
PYR	**1.87**	0.03	**2.54**	0.11	**2.82**	0.11	**2.52**	0.33	SxE	5.28	0.01
% Transporters									S	15.81	*
MUB	**7.77**	1.07	**20.74**	3.51	**16.57**	4.75	**19.41**	1.64	E	93.45	*
PYR	**5.22**	2.19	**9.16**	2.69	**1.15**	0.33	**9.01**	0.65	SxE	6.83	0.044
% ROS[1]									S	29.31	*
MUB	**5.06**	1.35	**1.22**	0.48	**2.02**	0.70	**1.20**	0.42	E	2.26	**0.15**
PYR	**4.44**	1.03	**2.82**	0.07	**4.18**	1.11	**0.00**	0.00	SxE	6.47	0.04
% Hydrolases									S	4.67	0.016
MUB	**2.45**	0.18	**4.76**	0.94	**3.02**	0.89	**2.00**	0.57	E	31.74	*
PYR	**5.86**	1.36	**4.69**	1.31	**9.60**	2.70	**4.85**	0.97	SxE	5.79	0.007

S (soil sample), E (extractant). MUB (Modified Universal Buffer); PYR (sodium pyrophosphate pH 7, 0.1 M); F (F ratio); Sd (standard deviation of the mean); * indicates $P<0.001$

The relative abundance of transporters and hydrolases was significantly influenced by both soil sample and extraction buffer (Table 1). The relative abundance of transporters was higher in GEP and SE5, the samples with the highest organic C concentrations in each area (Table 1; Table S2). Indeed, the relative abundance of transporters reached 20% of the proteins in the exo-metaproteome when MUB was used as the extractant. In contrast, the abundance of transporters peaked at 9% when proteins were extracted with sodium pyrophosphate, but again the abundance was higher in GEP and SE5.

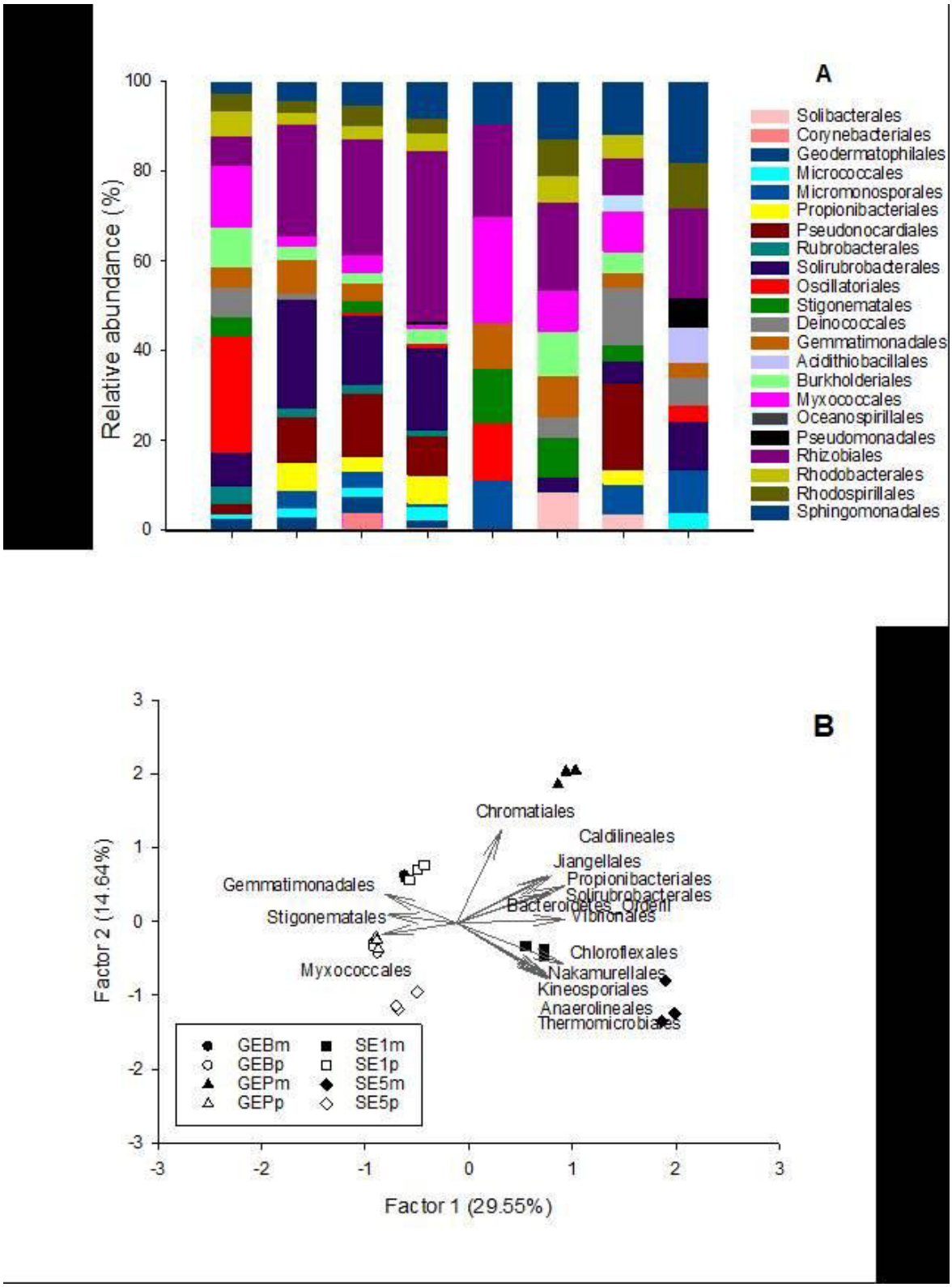

Figure 1. The relative abundance of bacterial orders (A) and the principal component analysis representing the structure of the microbial community (B), estimated by protein-based phylogeny.

206 The abundance of proteins involved in ROS transformation was higher in the
207 pyrophosphate extracts. For both MUB and pyrophosphate, GEB always contained the
208 highest abundance of proteins involved in ROS transformation. The abundance of
209 hydrolases was doubled when pyrophosphate was used as the extractant, in
210 comparison to MUB. The abundance of hydrolases was highest when SE1 was
211 extracted with pyrophosphate, reaching 9% of the total extracellular metaproteome
212 (Table 1).
213
214 The percentage of the total soil N present in the microbial biomass averages 4% [15];
215 hence, 96% of the total N in soil should be extracellular. Considering that a great
216 majority (95%) of N occurs in the form of organic N compounds and only about 5% is
217 mineral N [16], and even though N-forms other than proteinaceous N can occur in the
218 soil extracellular environment (i.e. nucleic acids, amino sugars) [17], our results
219 indicate a very-low efficiency of the two protein extraction methods, based on common
220 soil enzyme buffers, used here for extracellular proteins. Indeed, the methods assayed
221 did not extract even 0.3% of the total protein content. The protein extraction gave a low
222 yield, both in soil samples with a high organic carbon content (SE1 and SE5) and in
223 samples with a high clay content but a low organic C content (GEB and GEP),
224 suggesting that both factors limit the extraction of extracellular proteins from soil. In this
225 respect, Arenella et al. (2014) [18] indicated that the interaction between proteins and
226 soil humic substances affects protein identification through decreases in the protein
227 sequence coverage and in the number of proteins identified after tryptic digestion in
228 soil-protein-based models. Moreover, clay minerals can interact with proteins released
229 by cells and this interaction limits also the potential for soil proteomics [19-20].
230 The extraction with sodium pyrophosphate yielded a higher (at least 10-times) amount
231 of extracted proteins than that with MUB. Sodium pyrophosphate has been used
232 frequently to evaluate the activity of enzymes immobilised in soil humic substances.
233 Indeed, several studies have demonstrated the activity of hydrolytic enzymes involved

234 in the cycles of C, N and P in extracellular humic-complexes [13, 21-22]. These
235 proteins are stabilised in humic substances, protected against denaturing agents (i.e.
236 other proteases, temperature and moisture shocks), and may persist in soil even after
237 cell death. Precisely, the amount of proteins per g of soil was higher in SE1 and SE5,
238 which also contained higher amounts of soil organic matter and humic substances [6].
239 Interestingly, the diversity of proteins extracted with sodium pyrophosphate was lower
240 than with MUB. However, sodium pyrophosphate extraction indicated a higher amount
241 of extracellular hydrolases than MUB extraction. In contrast, the abundance of proteins
242 involved in the transport of sugars, phosphate and peptides (Table 1, Table S2) was
243 higher in MUB extracts, which seems logical considering that this buffer probably
244 extracts a more-soluble and dynamic fraction. Similarly, Johnson-Rollings et al. (2014)
245 [23] found a high amount of transport proteins involved in phosphate and amino acid
246 uptake. These authors utilised an extraction procedure based on K_2SO_4 but with an
247 amount of soil much greater than that utilised here.
248 We have evidenced that the phylogenetic and functional information retrieved from the
249 extracellular metaproteome can be biased by the extraction buffer. Indeed, the
250 structure of the microbial community determined according to the extracellular proteins
251 obtained with sodium pyrophosphate was different from the one based on the MUB
252 extracts. The extraction of extracellular soil proteins was very limited. The soil sample,
253 probably through its clay, organic matter and microbial biomass contents, and the
254 extraction buffer seem to play a key role in determining the protein extraction yield of
255 semiarid soils. Regardless of the soil characteristics, sodium pyrophosphate extracted
256 a higher amount of extracellular soil proteins than MUB and, consequently, it could be
257 potentially utilized for the study of the extracellular soil metaproteome. Nevertheless,
258 further studies are needed to enhance the yield of extraction and/or combine different
259 extraction methods. This will improve our understanding of microbial functionality and
260 its connections to soil functions.
261

Conflict of interest

The authors declare no conflict of interest.

References

[1] Chourey K, Jansson J, VerBerkmoes N, Shah M, Chavarria KL, Tom LM, Brodie EL, Hettich RL. Direct cellular lysis/protein extraction protocol for soil metaproteomics. J. Proteome Res. 2010;9:6615-2.

[2] Keiblinger K, Wilhartitz IC, Schneider T, Roschitzki B, Schmid E, Eberl L, Riedel K, Zechmeister-Boltenstern, S. Soil metaproteomics – comparative evaluation of protein extraction protocols. Soil Biol. Biochem. 2012;54 :14–24.

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

[4] Bastida F, Selevsek N, Torres IF, Hernández T, García C. Soil restoration with organic amendments: linking cellular functionality and ecosystem processes. Sci. Rep. 2015;5:15550.

[5] Hultman J, Waldrop MP, Mackelprang R, David MM, McFarland J, Blazewicz SJ, Harden J, Turetsky MR, McGuire AD, Shah MB, VerBerkmoes NC, Lee LH, Mavrommatis K, Jansson JK. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. Nature 2015;521:208–2.

[6] Bastida F, Torres IF, Moreno JL, Baldrian P, Ondoño S, Ruiz-Navarro A, Hernández T, Richnow HH, Starke R, García C, Jehmlich N. The active microbial diversity drives ecosystem multifunctionality and is physiologically related to carbon availability in Mediterranean semi-arid soils. Mol. Ecol. 2016;25:4660-73.

285 [7] Keiblinger KM, Fuchs S, Zechmeister-Boltenstern S, Riedel K. Soil and leaf litter
286 metaprotoemics – a brief guideline from sampling to understanding. FEMS Microbiol
287 Ecol. 2016;92:fiw180

288 [8] Burns RG, DeForest JL, Marxen J, Sinsabaugh RL, Stromberger ME, Wallenstein
289 MD, Weintraub MN, Zoppini A. Soil enzymes in a changing environment: Current
290 knowledge and future directions. Soil. Biol. Biochem. 2013;58:216-34.

291 [9] Bastida F, Moreno JL, Nicolás C, Hernández T, García C. Soil metaproteomics: a
292 review of an emerging environmental science. Significance, methodology and
293 perspectives. Eur. J. Soil Sci. 2009;60:845-59.

294 [10] Roberts P, Jones DL. Critical evaluation of methods for determining total protein in
295 soil solution. Soil Biol. Biochem. 2008;40:1485-95.

296 [11] Tabatabai MA. Soil Enzymes. In: Page AL, Miller RH, Keeney DR, editors.
297 Chemical and Microbiological Properties, 2nd, Methods of Soil Analysis. Madison, WI:
298 American Society of Agronomy-Soil Science Society of America; 1982, p. 903-947.

299 [12] Nannipieri P, Sequi P, Fusi P. Humus and enzyme activity. In: Piccolo A, editor.
300 Humic substances in terrestrial ecosystems, Amsterdam: Elsevier; 1996, p. 293-328.

301 [13] Bastida F, Jindo K, Moreno JL, Hernández T, García C. Effects of organic
302 amendments on soil carbon fractions, enzyme activity and humus-enzyme complexes
303 under semi-arid conditions. Eur. J. Soil Biol. 2012;53:94-2.

304 [14] Shannon CE, Weaver W. The mathematical theory of communication. University of
305 Illinois Press, Urbana, Illinois.

306 [15] Jenkinson DS. The determination of microbial biomass carbon and nitrogen in soil.
307 In: Wilson JE, editor. Advances in Nitrogen Cycling in Agricultural Ecosystems.
308 Wallingford: Commonwealth Agricultural Bureau. International; 1988, p.368-386.

309 [16] Bremener JM. Total nitrogen. In: Black CA, editor. Methods of Soil Analysis, vol 2.
310 Madison: American Society of Agronomy; 1965, p.1149-1178.

311 [17] Kögel-Knabner, I. Chemical structure of organic N and organic P in soil. In:
312 Nannipieri P, Smalla K, editors. Nucleic acids and proteins in soil. Berlin Heidelberg:
313 Springer-Verlag; 2006, p. 23-43

314 [18] Arenella M, Giagnoni L, Masciandaro G, Ceccanti B, Nannipieri P, Renella G.
315 Interactions between proteins and humic substances affect protein identification by
316 mass spectrometry. Biol Fertil Soils 2014;50:447-4.

317 [19] Giagnoni L, Magherini F, Landi L, Taghavi S, Modesti A, Bini L, Nannipieri P, van
318 der Lelie D, Renella G. Extraction of microbial proteome from soil: potential and
319 limitations assessed through a model study. Eur. J. Soil Sci. 2011;62:74-1.

320 [20] Giagnoni L, Migliaccio A, Nannipieri P, Renella G. High montmorillonite content
321 may affect soil microbial protoemic analysis. Appl. Soil Ecol. 2013;72:203-6.

322 [21] Bonmatí M, Ceccanti B, Nannipieri P. Protease extraction from soil by sodium
323 pyrophosphate and chemical characterization of the extracts. Soil Biol. Biochem.
324 1998;30:2113-25.

325 [22] Masciandaro G, Macci C, Doni S, Maserti BE, Calvo-Bado L, Ceccanti B,
326 Wellington E. Comparison of extraction methods for recovery of extracellular β-
327 glucosidase in two different forest soils. Soil Biol. Biochem. 2008;40:2156-1.

328 [23] Johnson-Rollings AS, Wright H, Masciandaro G, Macci C, Doni S, Calvo-Bado LA,
329 Slade SE, Vallin Plou C, Wellington EMH. ISME J. 2014;8:2148-50.

330

331 **Acknowledgements**

332 Felipe Bastida thanks the Spanish Government for his “Ramón y Cajal” contract (RYC-
333 2012-10666) and FEDER funds. The authors thank the Spanish Ministry for the CICYT

334 projects (AGL2014-54636-R, AGL2014-55269-R) and the Fundación Séneca
335 (19896/GERM/15). Kathleen Eismann (Department of Molecular Systems Biology) is
336 acknowledged for her technical support in proteome sample preparation.
337

338 **Figure Captions**

339 **Figure 1.** The relative abundance of bacterial orders (A) and the principal component
340 analysis representing the structure of the microbial community (B), estimated by
341 protein-based phylogeny.

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: Supporting Information soil_extracel_unmarked.doc