Title
Environmental proteomics: A long march in the pedosphere

Authors
Renella, Giancarlo
Ogunseitan, Oladele
Giagnoni, Laura
et al.

Publication Date
2014-02-01

DOI
10.1016/j.soilbio.2013.10.035

Peer reviewed

Soil Biology & Biochemistry 69 (2014) 34–37

Contents lists available at ScienceDirect

Soil Biology & Biochemistry

journal homepage: www.elsevier.com/locate/soilbio

News and views

Environmental proteomics: A long march in the pedosphere

Giancarlo Renella [a,*], Oladele Ogunseitan [b], Laura Giagnoni [a], Mariarita Arenella [a]

[a] Department of Agrifood Production and Environmental Sciences, University of Florence, Piazzale delle Cascine 28, 50144 Florence, Italy
[b] Program in Public Health, University of California, Irvine, CA 92697-3957, USA

ARTICLE INFO

Article history:
Received 1 July 2013
Received in revised form
21 October 2013
Accepted 22 October 2013
Available online 1 November 2013

Keywords:
Soil
Proteomics
Genomics
Microbial diversity
Soil functions
Environmental change

ABSTRACT

Environmental proteomics, the study of the expression profile of proteins extracted directly from living organisms and some stabilized extracellular proteins present in environmental samples, is a developing branch of soil science since the seminal papers appeared twenty years ago. Soil microbial communities hold the key to understanding terrestrial biodiversity; they are extremely complex and their physiological responses to dynamic environmental parameters are under-characterized. Therefore, the slow development of environment-related proteomic databases, and the high chemical reactivity of environmental matrices hamper the extraction, quantification, and characterization of proteins; and soil proteomics remains still in its infancy. We underscore the main achievements of environmental proteomics focusing on soil ecosystems, and we identify technical gaps that need to be bridged in the context of relevant ecological concepts that have received little attention in the development of proteomics methods. This analysis offers a new framework of research of soil proteomics toward improved understanding of the causal linkages between the structure and function of the soil microbiome, and a broader grasp of the sensitivity of terrestrial ecosystems to environmental change.

1. Advances in environmental proteomics

In 2013, analysis of soil protein diversity and functions using the tools of molecular biology (soil proteomics) reached the age of twenty since the pioneering work by Ogunseitan (1993) that initiated the environmental proteomics era in microbiology. This milestone publication stood alone for several years while most investigators continued the long-standing tradition of studying selected proteins, except for additional contributions to refine methodology and applications (Ogunseitan, 1996, 1997, 1998). Subsequently, Craig and Collins (2002) initiated a research program to detect proteins markers of human settlements in archaeological surveys, for archaeological site interpretation. The 'lag phase' of environmental proteomics ended in mid-2000s with a series of studies beginning with the soil metaproteomic study of soil microbial communities in Cd contaminated soils published by Singleton et al. (2003). Since then, the potential of research on proteins in the environment as an unprecedented approach for monitoring past and current biochemical processes in the environment became clear, also supported by the development of the mass spectrometry (MS) based proteomics (Aebersold and Mann, 2003).

In the last decade environmental proteomics has broadened its analytical spectrum to various natural and polluted soil types, plant phyllosphere and rhizosphere, waters and other environmental matrices, such as organic wastes and mine spoils (Wilmes and Bond, 2004; Tyson et al., 2004; Ram et al., 2005; Schulze et al., 2005; Benndorf et al., 2007; Chourey et al., 2010; Williams et al., 2010; Wang et al., 2011; Wu et al., 2011). Early environmental proteomics aimed at developing efficient methods for increasing the yield of protein extraction in the direction 'the more the proteins the better', leading mostly to inventories of bulk soil proteins than to discovery of relevant protein expression profile of soil microbial communities. While the progress of analytical methods allowed wider application of the proteomic approach to the most recalcitrant environmental matrices, it also led to a more 'technology driven' than 'hypothesis-driven' research for soil proteomics, particularly with the revolutionary application MS. However, a peculiar aspect of research in soil proteomics is that some of the major concepts developed and tested in decades of research in soil microbiology and biochemistry were not adequately exploited. Here, we focus on four key concepts related to soil microbial activity and ecology that, if considered will lead to a more rigorous framework of soil proteomics and its powerful

* Corresponding author. Tel.: +39 055 3288381; fax: +39 055 333273.
E-mail address: giancarlo.renella@unifi.it (G. Renella).

0038-0717/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.soilbio.2013.10.035

inclusion in predictive quantitative models of environmental change.

2. Challenges posed by diversity and dynamics of the soil microbiome

The pedosphere hosts the largest and most complex diversity of microorganisms among terrestrial ecosystems (Torsvik et al., 2002), and currently there are no laboratory or field techniques for resolving the theoretical complexity of soil microbial diversity and its metaproteome (Keller and Hettich, 2009). Detection and characterization of key microbial species in soil with annotated proteomes pinned to specific physiological processes is the "holy grail" of soil metaproteomics. Inconsistencies between the occurrence of well-defined taxonomic groups and results of soil proteomic analysis (Wang et al., 2011) are due largely to insufficient annotation of proteomic information as compared to the genomic information (Table 1), particularly for strains of naturally occurring organisms (Fig. 1). Integrating proteomic–genomic approaches can reveal the spatio-temporal pattern of phenotypic expression, subject to the limitations imposed by the labile nature of protein molecules (Tyers and Mann, 2003). Such approaches have been successfully implemented in low diversity biotopes such as acid mine drainage, where genomic data pointed to proteins from dominant community members microbial communities and the proteomic analysis for the characterization of the active metabolic pathways responsible for Fe oxidation by the dominant *Leptospirillum* and *Ferrobacillum* strains (Tyson et al., 2004; Ram et al., 2005). This strain resolution approach, focusing on the proteomic profiling of specific members of the soil microbiome with sufficiently well annotated proteome (Fig. 2) may greatly advance our understanding of biodiversity-activity patterns in soils.

Most soil microorganisms are presumed to be in a viable but non-culturable state (Schloss and Handelsman, 2005) because they are resting cells or otherwise dormant, and/or they cannot be cultivated with the currently available growth media and conditions. Although we do not fully understand the soil conditions leading to changes in microbial viability *in situ* and culturability under laboratory conditions, some metabolic pathways involved in the physiological transition phases of bacteria have been elucidated from *in vitro* studies, and purported stress-related proteins (e.g. phasins) are produced by resting cells (Nystrom, 2005). Surprisingly such biomarker proteins have not been reported in soil proteomic studies, questioning this presumption in soil microbial ecology.

Species dominance within the soil microbiome also poses a challenge for proteomics (Janssen, 2006). The patchy distribution of limiting nutrients and moisture in natural soils often translates to local hot spots of rapid growth or local extinction for species. This phenomenon poses a difficulty for sampling schemes used in proteomic assessments of the structure and function of microbial communities, especially for interdisciplinary studies and the interpretation of results based on observations where the dominant species are not annotated.

3. Challenges posed by local and global environmental change

The urgent need to predict anthropogenic impacts on the local and global environment warrant the integration of models accounting for functions associated with agents of change and mediators of responses. Patterns of abundance and diversity of soil proteins is a potentially important indicator of ecosystem functioning in part because their synthesis reflects the high microbial

Table 1
Example of genomic and proteomic information available in NCBI (www.ncbi.nlm.nih.gov) and Expasy SwissProt (www.expasy.org) database for organisms belonging to various kingdoms (September 2013), of specific interest for soil proteomics.

	Genomes (NBCI)	Proteins (NBCI)	Proteomes (SwissProt)
	Total	Entries	Total
Eukaryotes	$6.03 \cdot 10^3$	$1.49 \cdot 10^7$	$2.63 \cdot 10^2$
Opisthokonta	$4.82 \cdot 10^3$	$1.06 \cdot 10^7$	$2.81 \cdot 10^2$
Metazoa	$4.28 \cdot 10^3$	$7.37 \cdot 10^6$	$1.45 \cdot 10^2$
• Mesozoa	None	$8.70 \cdot 10^1$	None
• Eumetazoa	$4.08 \cdot 10^3$	$7.3 \cdot 10^6$	$1.41 \cdot 10^2$
○ Bilateria	$4.00 \cdot 10^3$	$7.25 \cdot 10^6$	$1.39 \cdot 10^2$
• Nematoda	$1.16 \cdot 10^2$	$3.43 \cdot 10^5$	9
• Arthropoda	$4.85 \cdot 10^2$	$1.94 \cdot 10^6$	$1.3 \cdot 10^1$
• Anellida	$2.7 \cdot 10^1$	$5.28 \cdot 10^3$	None
Fungi	$6.72 \cdot 10^2$	$3.14 \cdot 10^6$	$1.32 \cdot 10^2$
• Dikarya	$5.92 \cdot 10^2$	$3.02 \cdot 10^6$	$1.23 \cdot 10^2$
○ Ascomycota	$4.32 \cdot 10^2$	$2.37 \cdot 10^6$	$1.06 \cdot 10^2$
○ Basidiomycota	$1.60 \cdot 10^2$	$6.48 \cdot 10^5$	$1.70 \cdot 10^1$
Green plants (Viridiplantae)	$8.99 \cdot 10^2$	$2.78 \cdot 10^6$	$3.70 \cdot 10^1$
Green algae (Chlorophyta)	4.90	$1.79 \cdot 10^5$	8
Amoebozoa	1.90	$2.24 \cdot 10^5$	$1.1 \cdot 10^1$
Bacteria (Eubacteria)	$4.50 \cdot 10^3$	$6.58 \cdot 10^7$	$1.67 \cdot 10^3$
Proteobacteria	$1.8 \cdot 10^3$	$3.43 \cdot 10^7$	$7.64 \cdot 10^2$
• α-Proteobacteria	5.39	$5.23 \cdot 10^6$	$1.75 \cdot 10^2$
• β-Proteobacteria	2.65	$3.78 \cdot 10^6$	$1.17 \cdot 10^2$
• γ-Proteobacteria	$8.07 \cdot 10^2$	$2.30 \cdot 10^7$	$3.62 \cdot 10^2$
Firmicutes	$9.74 \cdot 10^2$	$1.61 \cdot 10^7$	$3.61 \cdot 10^2$
Cyanobacteria (blue–green algae)	9.6	$1.49 \cdot 10^6$	$4.9 \cdot 10^1$
Fusobacteria	2.1	$2.04 \cdot 10^5$	5
Actinobacteria	$7.46 \cdot 10^2$	$6.88 \cdot 10^6$	$1.82 \cdot 10^2$
Nitrospirae	10	$4.69 \cdot 10^4$	3
Archaea	$2.70 \cdot 10^2$	$1.8 \cdot 10^6$	$1.33 \cdot 10^2$
Viruses	$3.70 \cdot 10^3$	$2.23 \cdot 10^6$	$1.13 \cdot 10^3$
Plastids	$1.05 \cdot 10^5$ (Nucleotides)	$4.83 \cdot 10^3$	None
Mitochondria	9	$9.32 \cdot 10^5$	None

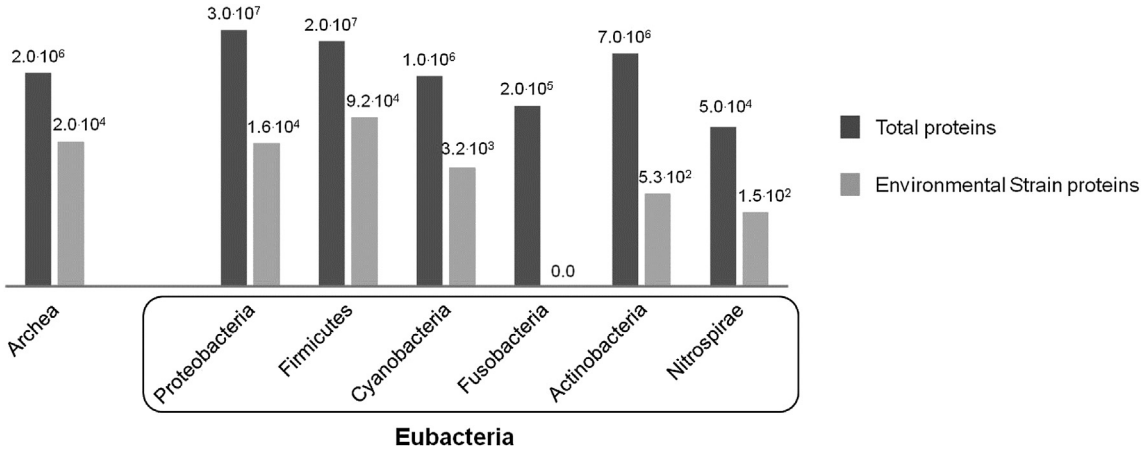

Fig. 1. Comparison of total bacterial and archaeal protein numbers as compared to proteins of naturally occurring strains in Expasy SwissProt (September 2013), a bioinformatic database portal operated by the Swiss Institute of Bioinformatics (www.expasy.org).

sensitivity to changes in soil and environmental conditions. For example, attempts to predict the responses of agricultural productivity and biogeochemical cycling to global climate change must necessarily include the role of soil proteins involved in carbon and nitrogen dynamics. However, the potential of soil proteomics is currently limited by the difficulties of distinguishing between intracellular and extracellular proteins, for example, in the well-studied case of extracellular enzymes that persist in soils but are not synthesized by viable organisms at the time of extraction (Burns, 1982). Soil microorganisms typically contain approximately 5% of the total soil nitrogen. Therefore, a large amount of soil proteins are sorbed onto reactive soil solid phases or trapped into organo-mineral complexes (Tomaszewski et al., 2011). Extracellular proteins constitute a dominant background against which newly expressed protein profiles of the soil microbiome must be resolved. Microbial-derived N in soil has been mainly recovered as associated to the acid-hydrolyzable humic fraction, mostly as proteinaceous materials and amino acids, less as amino-sugars, whereas there is still uncertainty about the heterocyclic N compounds. Alkaline-SDS solutions commonly used for soil protein extraction solubilize a minor proportion of total soil N. The paucity of data on soil N/protein N is compounded by the challenges facing precise quantification of whole soil proteins because interferences from co-extracted polyphenolics (Roberts and Jones, 2008) either prevent quantitative in-gel or off-gel proteomics (e.g. 2-DiGE, iTRAQ, ICAT), and make it difficult to produce quantitative data on soil proteomics suitable for inclusion in the general numeric models of environmental change.

The relative rarity of key function protein molecules against a background of abundant and diverse structural proteins may be responsible for the low rate of unique discoveries in soil metaproteomic studies, despite the rapid progress in protein extraction and detection methodologies. Even the most exhaustive studies are only able to identify 10–300 proteins (Benndorf et al., 2007; Chourey et al., 2010; Williams et al., 2010; Wang et al., 2011; Wu et al., 2011). Soil proteolytic activity (Renella et al., 2002) and geochemical denaturation are partially responsible for the low yield of protein extraction and identification. Post-expression modifications also pose challenges for protein identification by MS analysis because of the methodological dependence on existing content of proteomic databases. There is a need, therefore, for developing a database of post-translational protein modifications

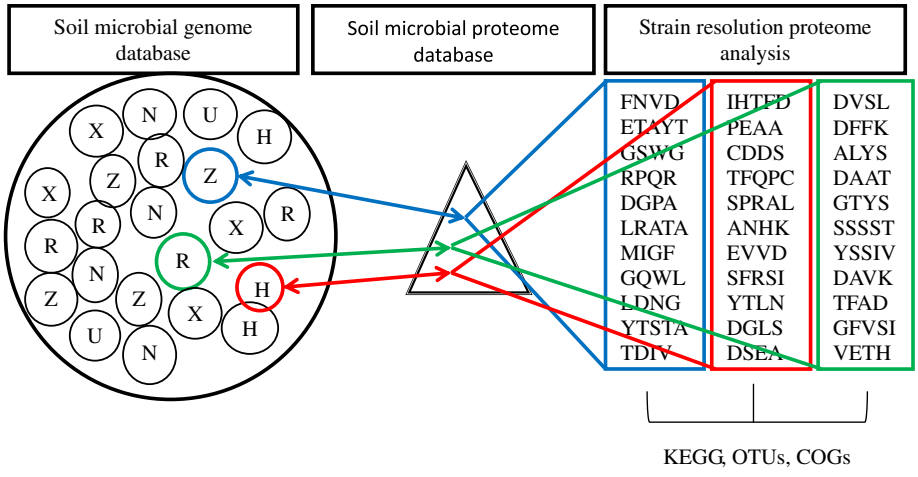

Fig. 2. Strain resolution approach for improving soil and environmental proteomics. The Kyoto Encyclopedia of Genes and Genomes (KEGG), is a database allowing the description of multiple level functions expressed by biological systems using molecular information (http://www.genome.jp/kegg). Operational Taxonomic Units (OTUs) are obtained by grouping DNA sequences defined with bioinformatics software, forming nodes in the phylogenetic analysis of biological communities. Clusters of Orthologous Groups (COGs) of proteins, are phylogenetic classifications of proteins encoded by complete genomes, with each COG consisting of three or more ortholog proteins sharing conserved domains shared by with homologs with distantly related species, thus allowing to infer protein functions information from less characterized organisms. The letters in the colored boxes represent amino acids that can be obtained by the protein mass spectrometry analysis of hypothetical soil microbial strains Z, H and R.

(e.g. phosphorylation, glycosylation) to enhance the interpretation of results in environmental proteomics studies. Further, developing a database of "wildtype" strains recovered from field sampling experiments will complement existing studies, especially if such strains are used as 'internal indicator species' in subsequent microcosm experiments to provide estimates of the efficiency of protein recovery and fingerprinting (Luo et al., 2007; Giagnoni et al., 2011, 2012, 2013).

Although much progress has been made on the identification and quantification of specific proteins within heterogenous mixtures through multi-dimensional chromatography coupled to MALDI or ESI MS, these approaches remain inherently qualitative (Cox and Mann, 2008). The situation can be improved for environmental proteomics through enhanced targeting of specific precursor ions, their fragmentation and the detection of ions in a triple or quadruple MS. Selected multiple reaction monitoring (SRM) or multiple reaction monitoring (MRM) are now capable of targeting up to 100 proteins simultaneously (Picotti et al., 2010). However, targeted proteomics is not yet a routine method because it requires sophisticated instrumentation set up and customized expertize, but its further development should provide new insights on the factors regulating the expression of key proteins (e.g. intracellular enzymes) in the environment.

4. Future prospects of soil proteomics

This vicennial provides a chance for a critical evaluation of the achievements in environmental proteomics, with applications particular to soil science and environmental change. We identify gaps and needs for future applications of proteomics for the analysis the biochemical pathways active in complex environmental matrices. It is clear that the potential number of distinct proteins, and differences in protein expression levels in natural environments exceed the range of electrophoresis-assisted proteomics. Emphasis on developing more sophisticated off-gel proteomics and bioinformatics is increasingly warranted. Progress in soil and environmental proteomics will be accelerated by soil 'microbiomics', the integration of metagenomics, proteomics, transcriptomics, and metabolomics, to generate datasets from different soils, providing rigorous cross-laboratory verification of data among the increasing number of environmental biologists that will take the challenge in the coming years.

Acknowledgments

The Department of Agrifood Production and Environmental Sciences receives grant-aided support from the Ente Cassa di Risparmio di Firenze for research in soil proteomics.

References

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. Nature 422, 198–207.
Benndorf, D., Balcke, G.U., Harms, H., von Bergen, M., 2007. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. ISME J. 1, 224–234.
Burns, R.G., 1982. Enzyme activity in soil: location and a possible role in microbial ecology. Soil Biol. Biochem. 14, 423–427.
Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K.L., Tom, L.M., Brodie, E.L., Hettich, R.L., 2010. Direct cellular lysis/protein extraction protocol for soil metaproteomics. J. Proteome Res. 9, 6615–6622.
Cox, J., Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372.
Craig, O.E., Collins, M.J., 2002. The removal of protein from mineral surfaces: implications for residue analysis of archaeological materials. J. Archaeol. Sci. 29, 1077–1082.
Giagnoni, L., Magherini, F., Landi, L., Taghavi, S., Modesti, A., Bini, L., Nannipieri, P., van der Lelie, D., Renella, G., 2011. Extraction of microbial proteome from soil: potential and limitations assessed through a model study. Eur. J. Soil Sci. 62, 74–81.
Giagnoni, L., Magherini, F., Landi, L., Taghavi, S., van der Lelie, D., Puglia, M., Bianchi, L., Bini, L., Nannipieri, P., Renella, G., Modesti, A., 2012. Soil solid phases effects on the proteomic analysis of *Cupriavidus metallidurans* CH34. Biol. Fertil. Soils 48, 425–433.
Giagnoni, L., Migliaccio, A., Nannipieri, P., Renella, G., 2013. High montmorillonite content may affect soil microbial proteomic analysis. Appl. Soil Ecol. 72, 203–206.
Janssen, P.H., 2006. Identifying the dominant soil bacterial taxa genes in libraries of 16S rRNA and 16S rRNA. Appl. Environ. Microbiol. 72, 1719–1728.
Keller, M., Hettich, R., 2009. Environmental proteomics: a paradigm shift in characterizing microbial activities at the molecular level. Microbiol. Mol. Biol. Rev. 73, 62–70.
Luo, Y., Vilain, S., Voigt, B., Albrecht, D., Hecker, M., Brözel, V.S., 2007. Proteomic analysis of *Bacillus cereus* growing in liquid soil organic matter. FEMS Microbiol. Lett. 271, 40–47.
Nystrom, T., 2005. Bacterial senescence, programmed death and premeditated sterility. Am. Soc. Microbiol. News 71, 363–369.
Ogunseitan, O.A., 1993. Direct extraction of proteins from environmental samples. J. Microbiol. Methods 17, 273–281.
Ogunseitan, O.A., 1996. Protein profile variation in cultivated and native freshwater microorganisms exposed to chemical environmental pollutants. Microb. Ecol. 31, 291–304.
Ogunseitan, O.A., 1997. Direct extraction of catalytic proteins from natural microbial communities. J. Microbiol. Methods 28, 55–63.
Ogunseitan, O.A., 1998. Protein method for investigating mercuric reductase gene expression in aquatic environments. Appl. Environ. Microbiol. 64, 695–702.
Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., Wenschuh, H., Aebersold, R., 2010. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. Nat. Methods 7, 43–46.
Ram, R.J., VerBerkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C., Shah, M., Hettich, R.,L., Banfield, J.L., 2005. Community proteomics of a natural microbial biofilm. Science 308, 1915–1920.
Renella, G., Landi, L., Nannipieri, P., 2002. Hydrolase activity during and after the chloroform-fumigation of soils as affected by protease activity. Soil Biol. Biochem. 34, 51–60.
Roberts, P., Jones, D.L., 2008. Critical evaluation of methods for determining total protein in soil solution. Soil Biol. Biochem. 40, 1485–1495.
Schloss, P.D., Handelsman, J., 2005. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome Biol. 6, 229.
Schulze, W.X., Gleixner, G., Kaiser, K., Guggenberger, G., Mann, M., Schulze, E.D., 2005. A proteomic fingerprint of dissolved organic carbon and of soil particles. Oecologia 142, 335–343.
Singleton, I., Merrington, G., Colvan, S., Delahunty, J.S., 2003. The potential of soil protein-based methods to indicate metal contamination. Appl. Soil Ecol. 23, 25–32.
Tomaszewski, J.E., Schwarzenbach, R.P., Sander, M., 2011. Protein encapsulation by humic substances. Environ. Sci. Technol. 45, 6003–6010.
Torsvik, V., Øvreås, L., Thingstad, T.F., 2002. Prokaryotic diversity – magnitude, dynamics and controlling factors. Science 296, 1064–1066.
Tyers, M., Mann, M., 2003. From genomics to proteomics. Nature 422, 193–197.
Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37–43.
Wang, H.B., Zhang, Z.X., Li, H., He, H.B., Fang, C.X., Zhang, A.J., Li, Q.S., Chen, R.S., Guo, X.K., Lin, H.F., Wu, L.K., Lin, S., Chen, T., Lin, R.Y., Peng, X.X., Lin, W.X., 2011. Characterization of metaproteomics in crop rhizospheric soil. J. Proteome Res. 10, 932–940.
Williams, M.A., Taylor, E.B., Mula, H.P., 2010. Metaproteomic characterization of a soil microbial community following carbon amendment. Soil Biol. Biochem. 42, 1148–1156.
Wilmes, P., Bond, P.L., 2004. Metaproteomics: studying functional gene expression in microbial ecosystems. Trends Microbiol. 14, 92–97.
Wu, L., Wang, H., Zhang, Z., Lin, R., Zhang, Z., Lin, W., 2011. Comparative metaproteomic analysis on consecutively *Rehmannia glutinosa*-monocultured rhizosphere soil. PLoS One 6, e20611.