Pak. J. Bot., 50(1): 381-387, 2018.

SEQUENCE ANALYSIS OF MICROBIAL COMMUNITY INTEGRATING METAGENOME SEQUENCE DATA OBTAINED FROM *POA ALPIGENA* GRASSLAND IN THE SANJIANGYUAN

PING FAN[1,2], YAOU ZHANG[2*] AND RUIQIANG ZHANG[1*]

[1]State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University;
[2]Graduate School at Shen Zhen, Tsinghua University;
[*]Corresponding author's email: 383970338@qq.com; Tel: 09715318556; Tel: 075526036884

Abstract

A metagenomic analysis was performed on the soil samples from *Poa alpigena* grassland in the Sanjiangyuan region of the Qinghai–Tibetan Plateau. The metagenomic results were studied to understand the effect of vegetation change on microbes in *Poa alpigena* grassland soil. A total of 62,724 ORFs with an average length of 188.9 bp were identified. The sequences were divided into different groups of NCBI taxonomy database and SSUrRNA database. Moreover, unigenes highly enriched in "translation, ribosomal structure and biogenesis functions" (115, 18.7 %), "Transcription" (61, 0.10 %), "Amino acid transport and metabolism (51, 0.08 %)" and "Signal transduction mechanisms" (34, 0.06 %). Using KEGG analysis, we found that a large amount of genes were involved in "Biosynthesis of secondary metabolites", "Microbial metabolism in diverse environments", "Bacterial secretion system", Carbon metabolism and Amino acid metabolism, indicating that annotated unigenes might involve in the metabolic pathways in the response in soil restoration.

Key words: *Poa alpigena* grassland; Metagenomic analysis; Microbes; Metabolic pathways.

Introduction

A large area of alpine meadow has been degraded into the "black beach" resulting in the reduction of grassland fractional coverage and soil fertility and flourishing rodent population in the Sanjiangyuan due to human disturbance and climatic change (Dong *et al*., 2015). In order to restore the degraded grasslands, the Chinese government has made massive capital and technological investments in the region (He *et al*., 2008; Shan *et al*., 2015). The main measures has been taken to recover the "black beach", such as fencing and reseeding (Richard *et al*., 2005). Obviously, good results have been achieved by breeding and planting pasture species, especially *Poa alpigena* (Xin *et al*., 2014).

Soil microbes are seen as important drivers of plant diversity and productivity in terrestrial ecosystems (Van *et al*., 2008; Mou *et al*., 2016). Recent researches have reported that measurements of the soil microbial community can evaluate the restoration of ecological environment after degradation (Harris, 2003; Nejidat *et al*., 2016). Land degradation results in the decline of soil microbial diversity (Singh, 2015). Similarly, the loss of soil microorganism cause not only a decrease dramatically in soil productivity but also the soil ecosystem functioning (Seneviratne & Kulasooriya, 2013; Khan *et al*., 2017). As the gradual recovery of the ecosystem of black beach, soil microbes in the areas have an inevitable change with the restoration of vegetation (Li *et al*., 2016; Shao *et al*., 2016). Therefore, to study soil microbial communities has an important meaning on understanding the directly effect on soil microorganisms by altering soil edaphic properties.

Metagenomics is an environmental and community genomic technologies and bioinformatics tools to insight into genetic content of microorganisms obtained from a common habitat (Huson *et al*., 2007; Thomas *et al*., 2012), which can provide the information of functional genes composition of microbial communities. The data of Metagenomics could be used to answer fundamental questions in microbial ecology (Riesenfeld *et al*., 2004). Using the next-generation sequencing, an in-depth study of the soil microbial diversity can be implemented by sequence-based metagenomics (Ercolini *et al*., 2013). Despite the significant role that microbes play in ecosystem recovery, little information is available about the effects of ecological recovery on microbial communities in soils of the Sanjiangyuan. Therefore, the metagenomic sequencing library has been constructed to understand the extent and role of microbial diversity of *Poa alpigena* grassland soil. This study aimed to provide the first metagenome sequence data obtained from *Poa alpigena* Grassland in the Sanjiangyuan used for further research.

Materials and Methods

Experimental site and sampling: This study was conducted at the Sanjiangyuan National Nature Reserve (SNNR), located at the Golog prefecture, Dawu, Qinghai, China, (34°27.8' N, 100°12.7' E) at an altitude of 3,742 m. The average annual temperature is −3.9°C, the annual sunshine is 2,260 hours, the average annual precipitation is 513.2–542.9 mm, and the average annual evaporation is 2471.6 mm. This is an old experimental site that has planted *Poa alpigena* for four consecutive years. The samplings were performed in May 2012 and August 2012 at black soil land using the five-point sampling method (Zhao *et al*., 2010). Each sample was taken from approximate 20 cm depth, then the collected five samples was evenly mixed to being transported to the laboratory and stored at -80°C.

Extraction of genomic DNA, PE library construction and Metagenome assembly: The soil DNA were extracted using an EZNA® Soil DNA kit (Omega Bio-tek, Norcross, GA, U.S.). DNA yields of samples ranged from 0.2 to 1.0 g. The quality of genomic DNA was determined

by 1% agarose gel electrophoresis. Sample of DNA can be used for PE library construction. The DNA was processed into fragments with a size of 300 bp using the Covaris M220 platform. The PE100 library was constructed using TruSeq™ DNA Sample Prep Kit according to the manufacturer instructions. The "Y" shaped adapter was ligated and magnetic beads were used to remove the fragments from the self-ligation of adapters. The enrichment of the library template was completed via the bridge polymerase chain reaction method using cBot Truseq PE Cluster Kit v3-cBot-HS. The amplified DNA clusters were denatured by sodium hydroxide to produce single-stranded DNA fragments. Sequencing was completed on the Illumina Hiseq2000 platform (Illumina, USA), and the sample processing before sequencing was performed using Truseq SBS Kit v3-HS (200 cycles).

Data analysis: The processing of sequencing results included filtering of sequencing data, quality trimming, sequence assembly and gene prediction, and functional annotation and taxonomic classification were performed on the predictive genes. The SeqPrep and Sickle were used for adapter removal and quality trimming. In order to ensure the reliability of subsequent data processing, the raw reads were trimmed using a minimum quality score 20 and a minimum read length 20 bp, according to Albertsen *et al*. (2012). Clean reads were then assembled by SOAP denovo Version 1.3, scaffolds length >500 bp would be retained. The assembly of the optimized sequence after processing was accomplished using SOAP denovo software Version 1.3 (http://soap.genomics.org.cn/). The Kmer value of the main parameter was set to 25–37, the scaffolds with a size of more than 300 bp were counted, and the optimum assembly result was chosen (Chikhi, 2014). The open reading frames (ORFs) were predicted using Contig (http://staden.sourceforge.net/ contig.html) and Singleton (Tchitchek, 2014) in the splicing results from MetaGene Annotator (http://metagene.cb.k.u-tokyo.ac.jp/). The ORF gene sequences or sequencing reads were compared with the NR database and the taxonomic information of the genes was obtained according to the National Center for Biotechnology Information (NCBI) taxonomy database by BLAST (BLAST Version 2.2.25). The sequencing reads were compared with the small-subunit ribosomal RNA database of the SILVA system using the Basic Local Alignment Search Tool (BLAST), and the expectation value (Evalue) was set to 1e−5, to obtain taxonomic information and conduct statistical analysis. The GO functional analysis and KEGG Pathway analysis were performed on Unigenes.

Results

Characteristics of the metagenome of the microbial community: The metagenomic library produced from the soil sample had high quality of sequencing such as evenly distributed and lower N % content, which could be used for subsequent analysis (Table 1). Macro genome sequencing generated 0.33 billion contigs with an average 121-180 bp of sequence length. A total of 62,724 ORFs were identified having an average length of 188.9 bp, among which a length of more than 100 bp were translated into amino acid sequences. In addition, the phylogenetic analysis were conducted on sequences from soil metagenomes, the different types of microorganism could clearly be classified into eight phylum of microbe. Especially for *Nitrososphaera*, there are not close relationship between other microbial genera and groups (Fig. 1).

Table 1. The detailed information of metagenome assembly.

Assembly content	Statistics
Scaffold number	333, 527, 28
Scaffold N50 length	407 bp
Scaffold GC content	61.51%
Contig number	333, 829, 97
Contig N50 length	393 bp
Contig GC ontent	61.51%

Distribution of functional genes in microbial species: To analyze the relative abundances of microbial species at each taxonomic level, the percentage of annotated genes from different species was calculated as shown in Fig. 2A. The amount number of unigenes in Proteobacteria, Actinobacteria, Firmicutes, Acidobacteria and Thaumarchaeota were identified, which were 32.51%, 13.62%, 11.50 %, 11.09 % and 10.25%, respectively. The percentage of unigenes related to crenarchaeota, euryarchaeota, streptophytaand and chordata were 2.21%, 1.87%, 1.79% and 0.51%. Additionally, archaea and bacteria only accounted for 0.49 % and 2.5 %, respectively.

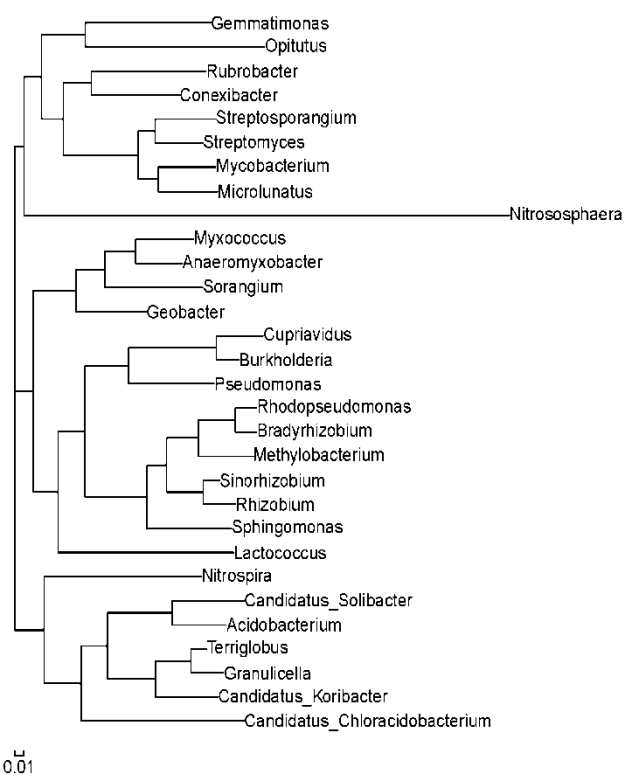

Fig.1. Phylogentic relationship analysis among sequences from soil metagenomes.

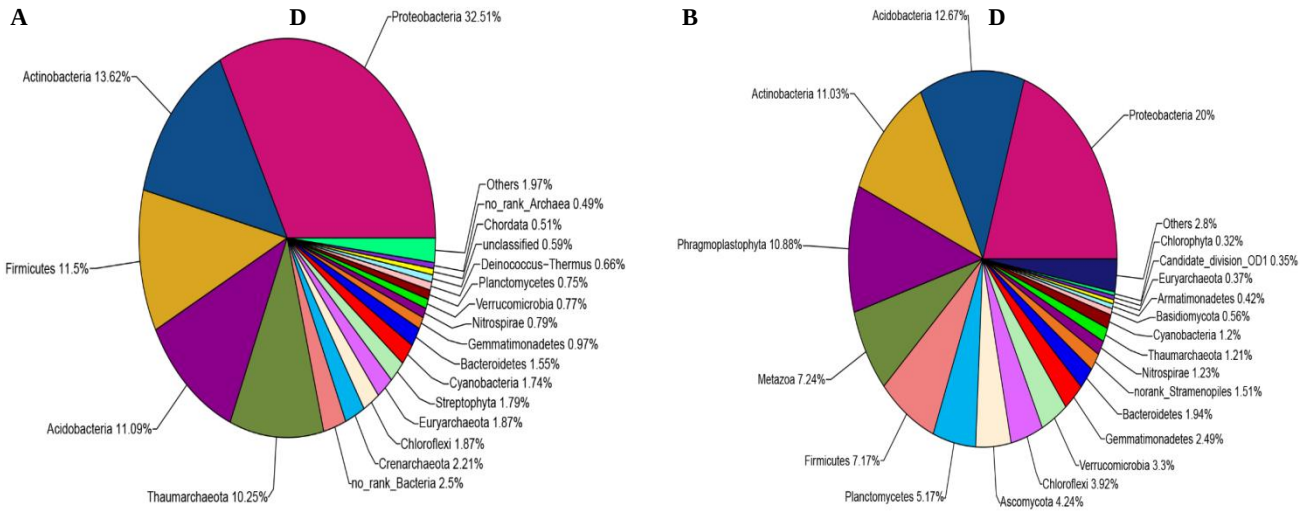

Fig. 2. The percentage of annotated sequences classified into different groups of NCBI taxonomy database (A) and SSUrRNA database (B).

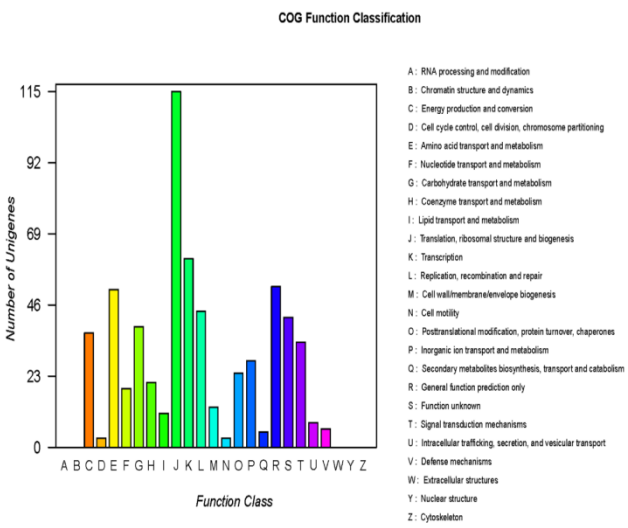

Fig. 3. The number of unigenes were classified into different COG categories.

Other species classification map was constructed based on SSUrRNA database (Fig. 2B), the microbial community were classified into twenty-two different groups. The percentage of unigenes mapped into proteobacteria, acidobacteria, actinobacteria, phragmoplastophyta, fungi ascomycota and basidiomycota were 20%, 12.67%, 11.03%, 10.88%, 4.24% and 0.56%, respectively. For archaea groups, the smaller proportions of euryarchaeota (0.37%) and thaumarchaeota (1.21%) were found, but the crenarchaeota and archaea was not observed. Similarly, some of unigenes related to for bacteria groups were identified, such as armatimonadetes (0.42%), candidate (0.35%) and chlorophyta (0.32%), except deinococcus.

Functional annotation and characterization of unigenes: A large number of unigenes were classified into four COG categories, including "Information storage and processing", "Cellular processes and signaling", "Metabolism" and "Poorly characterized". Unigenes highly enriched in "translation, ribosomal structure and biogenesis functions" (115, 18.7%), "Transcription" (61, 0.10%), "Amino acid transport and metabolism (51, 0.08%)" and "Signal transduction mechanisms" (34, 0.06%). On the contrary, none of unigenes were mapped to the categories relating to "RNA processing and modification", "chromatin structure and dynamics", "extracellular structures", "nuclear structure" and "cytoskeleton" (Fig. 3). Moreover, the functional characteristics of DEGs related to eukaryotes in the constructed metagenomic library were not clearly observed. Only a small number of unigenes were associated with "energy production and conversion", "coenzyme transport and metabolism" and "inorganic ion transport and metabolism". The unigenes that are known to be involved in biochemical process, and metabolites of microorganism were differentially expressed under ecological recovery.

A KEGG pathway analysis was conducted to reveal that microbial genes were involved in different metabolic pathways (Fig. 4). The large numbers of unigenes were found in "Biosynthesis of secondary metabolites (ko01110)", "Microbial metabolism in diverse environments (ko01120)" and "Bacterial secretion system" (ko03070) in the microbial community from *Poa alpigena* grassland soil. Furthermore, large amounts of carbon metabolism-associated genes were observed such as fructose and mannose metabolism (ko00051), amino sugar and nucleotide sugar metabolism (ko00520), glycolysis / gluconeogenesis (ko00010) and galactose metabolism (ko00052) (Fig. 5**)**. Similarly, a number of amino acid-related genes encoding transcriptional factors and enzymes were identified (Fig. 6), for example, glycine, serine and threonine metabolism (ko00260), lysine biosynthesis (ko00300), lysine degradation (ko00310), propanoate metabolism (ko00640) and tryptophan metabolism (ko00380). Genes encoding proteins showed up-regulation in the metabolic pathways, the results indicated that the activities of metabolism-associated genes may play an important role in the response of soil microbe to restore ecological environment after degradation.

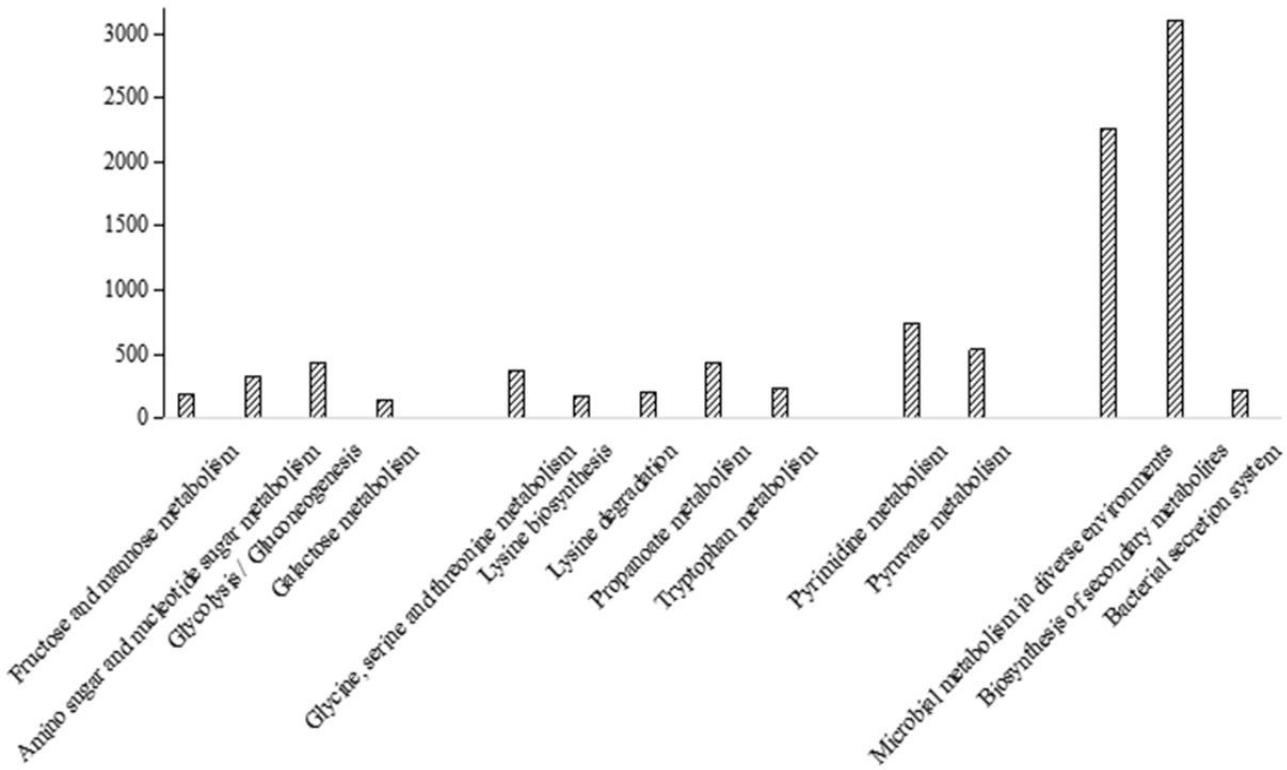

Fig. 4. The microbial genes involved in different metabolic pathways.

Discussion

A high diversity of soil microorganisms are important indicator to evaluate whether the soil is polluted or not (Daniel, 2004). Actually, 1 gram of soil includes more than 10 billion microorganisms which are possibly thousands of various species (Roselló-Mora & Amann, 2001). A wide range of microorganisms such as bacteria, actinomycetes, fungi and cyanobacteria, are considered as the important soil microbes (Steele *et al*., 2008). Metagenomics provides a new technical innovation that promotes study of environmental microbiology and microbial physiology. Moreover, it is an assessment system to discover the largely untapped genetic information of soil microorganism (Delmont *et al*., 2011). In recent years, a large number of metagenomic surveys for new genes concentrated on extreme environments, due to the fact that the environments consisted of a range of microbes having highly genetic diversity (Leresche & Meyer, 2006). Sales & Lee (2015) reported that meta-omics approaches were used to insight into the complex microbial communities in large-scale phosphorus recovery from domestic wastewater. The results of functional metagenomics showed the diverse of β-lactamases in a remote Alaskan soil (Allen *et al*., 2009). In our study, a large number of unigenes in Proteobacteria, Actinobacteria, Firmicutes, Acidobacteria and Thaumarchaeota were identified from metagenomic libraries, it is evident that the expression of unigenes might play an important role in restoring ecological environment after degradation in Sanjiangyuan.

The microbial genes distribution showed depth-variable community trends in carbon- and energy-related metabolic pathways (DeLong *et al*., 2006). Core pathways were often implemented by annotated genes within one or more niches of microbiome, such as carbon metabolism and amino acid metabolism (Thompson *et al*., 2011). Glycolysis is a key carbohydrate metabolic pathway, and stress results in transfer into amino acid and sucrose contents (Broeckling *et al*., 2005). Moreover, stress condition induce the synthesis of amino acids which contribute to turgor maintenance by osmotic adjustment in living cell (Arbona *et al*., 2008). Amino acids can be a sensitive measure of microbial stress in soil (Nannipieri *et al*., 2003). Campbell *et al*. (1997) found that maximum utilization of carbohydrates and amino acids were faster than amides, phenolic and long chain aliphatic acids, which were obtained from different soil samples with different types of grassland vegetation. In our study, several unigenes related to carbon metabolism and amino acid metabolism were found as shown in Fig. 5, indicating these genes were involved in the carbon metabolism and amino acid metabolism-associated with pathways in the response in soil restoration.

Acknowledgements

This study was funded by the Sanjiangyuan Special Fund from the Science and Technology Department of Qinghai (2011-1), the High-level Personnel Project of Qinghai University (2008-QGC-7), and the Foundation for Young to Middle-aged Scientists of Qinghai University (2012-30). We would also like to thank Ya-Ou Zhang from the Shenzhen Graduate School of Tsinghua University, who provided tremendous support in the follow-up work.

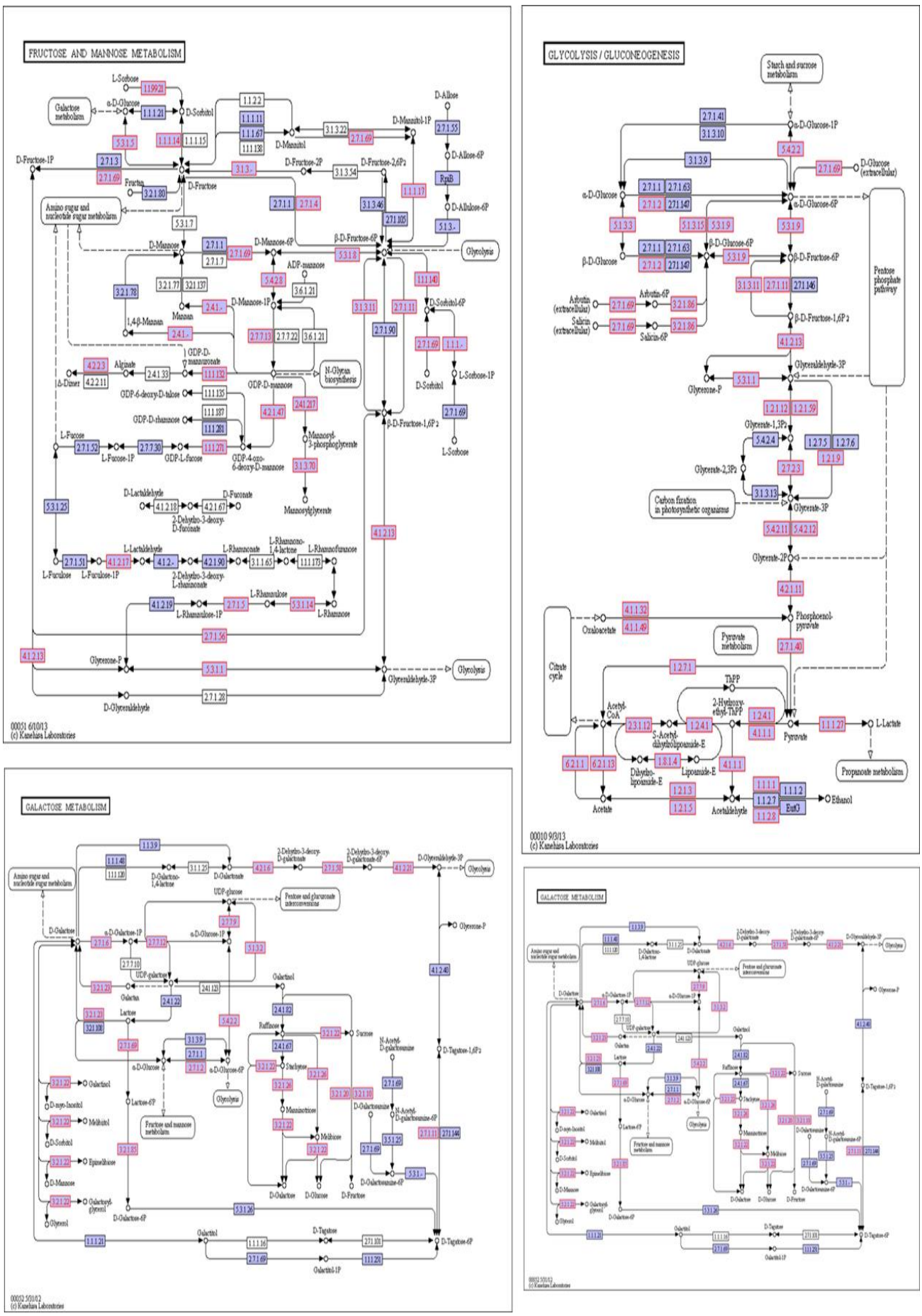

Fig. 5. The identified genes of microbial communities were related to Carbon metabolism response to ecosystem recovery.

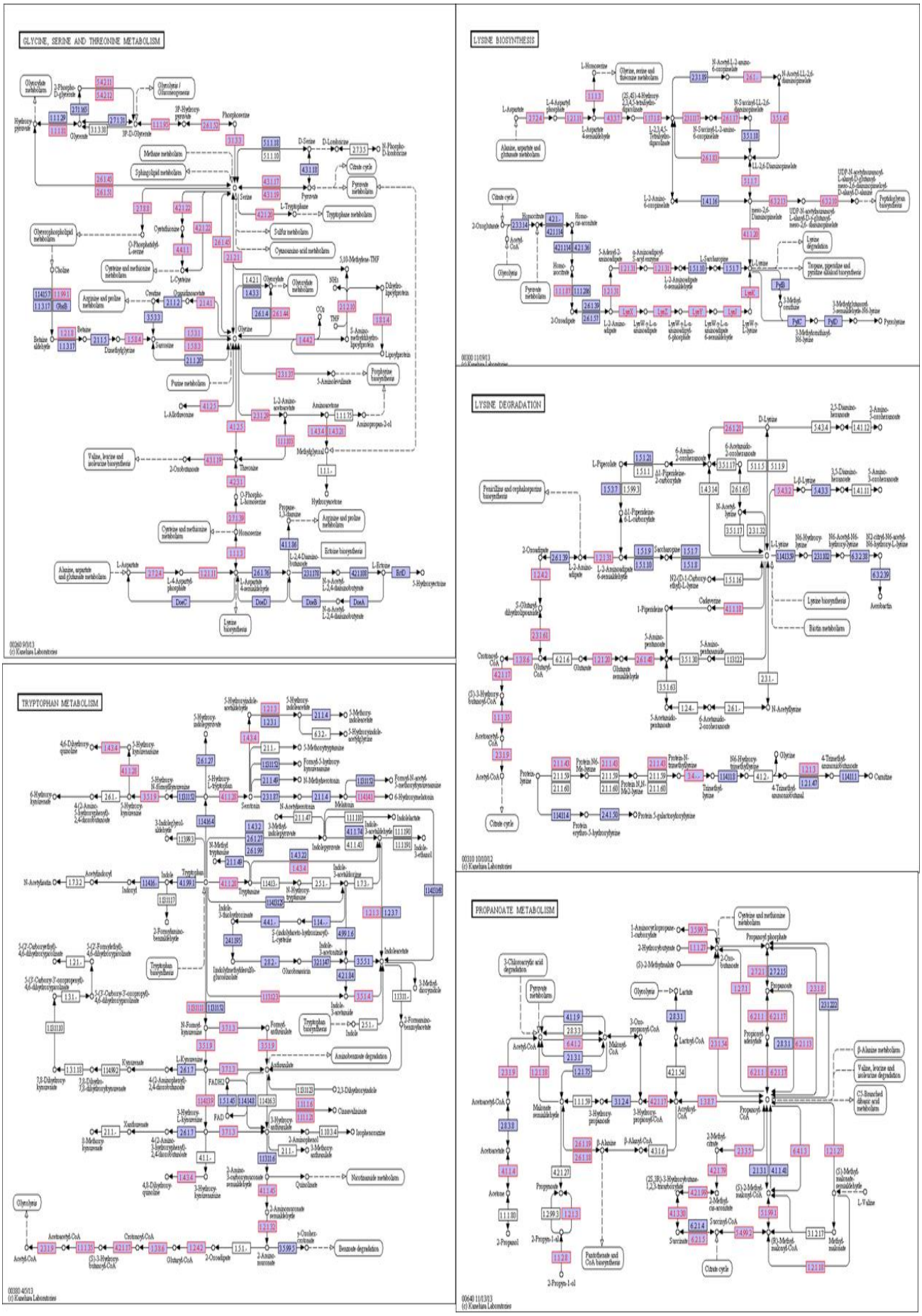

Fig. 6. Effects on the expression of genes associated with Amino acid metabolism.

References

Albertsen, M., L.B.S. Hansen, A.M. Saunders, P.H. Nielsen and K.L. Nielsen. 2012. A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *ISME J*., 6(6): 1094-1106.

Allen, H.K., L.A. Moe, J. Rodbumrer, A. Gaarder and J. Handelsman. 2009. Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. *ISME J*., 3(2): 243-251.

Arbona, V., Z. Hossain, M.F. López-Climent, R.M. Pérez-Clemente and A. Gómez-Cadenas. 2008. Antioxidant enzymatic activity is linked to waterlogging stress tolerance in citrus. *Physiol Plant*, 132: 452-466.

Broeckling, C.D., D.V. Huhman, M.A. Farag, J.T. Smith, G.D. May, P. Mendes, R.A. Dixon and L.W. Sumner. 2005. Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. *J. Exp. Bot*.. 56: 323-336.

Campbell, C.D., S.J. Grayston and D.J. Hirst. 1997. Use of rhizosphere carbon sources in sole carbon source tests to discriminate soil microbial communities. *J. Microbiol. Methods*, 30(1): 33-41.

Chikhi, R. and P. Medvedev. 2014. Informed and automated k-Mer size selection for genome assembly. *Bioinformatics*, 30(1): 31-7.

Daniel, R. 2004. The soil metagenome–a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol*., 15(3): 199-204.

Delmont, T.O., P. Robe, S. Cecillon, I.M. Clark, F. Constancias, P. Simonet, P.R. Hirsch and T.M. Vogel. (2011). Accessing the soil metagenome for studies of microbial diversity. *Appl. Environ. Microbiol*., 77(4): 1315-1324.

Delong, E.F., C.M. Preston, T. Mincer, V. Rich, S.J. Hallam, N.U. Frigaard, A. Martinez, M.B. Sullivan, R. Edwards, B.R. Brito, S.W. Chisholm and D.M. Karl. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, 311(5760): 496-503.

Dong, Q.M., Y.S. Ma, C.J. Xu, J.J. Shi, Y. Wang, Y.L. Wang, L. Sheng and S.X. Li. 2015. Study of classification and gradation, restoration of black-soil beach degraded in the headwaters of three rivers. *Acta Agrestia Sinica*., 23(3): 441-447.

Ercolini, D. 2013. High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Appl. Environ. Microbiol*., 79(10): 3148-3155.

Harris, J. 2003. Measurements of the soil microbial community for estimating the success of restoration. *Eur. J. Soil Sci*., 54(4): 801-808.

He Y.L., H.K. Zhou, X.Q. Zhao, D.Z. Lai and J.Z. Zhao. 2008. Alpine grassland degradation and its restoration on qinghai-tibet plateau. *Pra. & Am. Hus.*, 11: 1-9.

Huson, D.H., A.F. Auch, J. Qi and S.C. Schuster. 2007. MEGAN analysis of metagenomic data. *Genome Res*., 17(3): 377-386.

Khan, M.A., M.U. Shirazi, A. Shereen, S.M. Mujtaba and M.A. Khan. 2017. Identification of some wheat (*Triticum aestivum* L.) lines for salt tolerance on the basis of growth and physiological characters. *Pak. J. Bot*., 49(2): 397-403.

Leresche, J.E. and H.P. Meyer. 2006. Chemocatalysis and biocatalysis (biotransformation): Some thoughts of a chemist and of a biotechnologist. *Org. Process. Res. Dev*., 10: 572-580.

Li, G.Y., C.H. Jiang, T. Cheng, H.R. Zhang and Z.T. Chen. 2016. Spatial-temporal variation of vegetation phenology and their relationships with vegetation degradation in a Qinghai lake watershed. *Acta Prataculturae Sinica*., 25(1): 22-32.

Mou, X.J., X.Y. Zhao, S. Rao, Q. Huang and H.X. Chai. 2016. Changes of ecosystem structure in qinghai-tibet plateau ecological barrier area during recent ten years. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 52(2): 279-286.

Nannipieri, P., J. Ascher, M. Ceccherini, L. Landi, G. Pietramellara and G. Renella. 2003. Microbial diversity and soil functions. *Eur. J. Soil Sci*., 54(4): 655-670.

Nejidat, A., R.M. Potrafka and E. Zaady. 2016. Successional biocrust stages on dead shrub soil mounds after severe drought: Effect of micro-geomorphology on microbial community structure and ecosystem recovery. *Soil Biol. & Biochem.*, 103: 213-220.

Richard, C. 2005. Developing alternatives to resettlement for pastoralists on the Tibetan Plateau. *Nomadic Peoples*, 9(1-2): 103-106.

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

Rosselló-Mora, R. and R. Amann. 2001. The species concept for prokaryotes. *FEMS Microbiol. Rev*., 25: 39-67.

Sales, C.M. and P.K.H. Lee. 2015. Resource recovery from wastewater: Application of meta-omics to phosphorus and carbon management. *Curr. Opin. Biotechnol*., 33: 260-267.

Seneviratne, G. and S.A. Kulasooriya. 2013. Reinstating soil microbial diversity in agroecosystems: The need of the hour for sustainability and health. *Agri. Ecosys. & Environ.*, 164: 181-182.

Shan, J.J. 2015. Achievements, problems and Countermeasures of ecological protection in Sanjiangyuan. *Res. on Dev*., 5: 21-24.

Shao, Q.Q., J.W. Fan, J.Y. Liu, L. Huang and W. Cao. 2016. Assessment on the effects of the first-stage ecological conservation and restoration project in Sanjiangyuan region. *Acta Geographica Sinica*., 71(1): 3-20 (Chinese).

Singh, J.S. 2015. Microbes: the chief ecological engineers in reinstating equilibrium in degraded ecosystems. *Agri. Ecosys. & Environ.*, 203: 80-82.

Steele, H.L., K.E. Jaeger, R. Daniel and W.R. Streit. 2008. Advances in recovery of novel biocatalysts from metagenomes. *J. Mol. Microbiol. Biotechnol*., 16(1-2): 25-37.

Tchitchek, N., D. Safronetz, A.L. Rasmussen, C. Martens, K. Virtaneva, S.F. Porcella, H. Feldmann, H. Ebihara and M.G. Katze. 2014. Sequencing, annotation and analysis of the Syrian hamster (*Mesocricetus auratus*) transcriptome. *PLoS One*. 9(11): e112617.

Thomas, T., J. Gilbert and F. Meyer. 2012. Metagenomics-a guide from sampling to data analysis. *Microb. Inform Exp*., 2(1): 1.

Thompson, L.R., Q. Zeng, L. Kelly, K.H. Huang, A.U. Singer, J. Stubbe and S.W. Chisholm. 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc. Natl. Acad. Sci. U S A*. 108(39): E757-E764.

Van, Der Heijden, M.G., R.D. Bardgett and N.M. Van Straalen. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.*, 11(3): 296-310.

Xin, Y.C. 2014. The degradation situation of natural grassland and control measures in Qinghai province. *Qinghai Prataculture*, 23(4): 44-49.

Zhao, Z.X., K. Luo, G.H. Chen, Y.H. Yang, Z.C. Mao, E.M. Liu and B.Y. Xie. 2010. Analysis of bacterial diversity in rhizosphere of cucumber in greenhouse by the methods of metagenomicend-random sequencing and 16S rDNA technology. *Acta Ecologica Sinica*, 30(14): 3849-3857.

(Received for publication 10 January 2017)