ILLUMINA MISEQ-BASED SEQUENCING ANALYSIS OF BACTERIAL COMMUNITY IN VIETNAMESE GINSENG CULTIVATED SOIL IN THE NGOC LINH MOUNTAIN, VIETNAM

Ngoc Lan Nguyen[1] , Bao Tram Tran[2], Huong Son Pham[2], The Hai Pham[3]
[1]Institute of Genome Research,
Vietnam Academic of Science and Technology
[2]National Center for Technological Progress,
Ministry of Science and Technology
[3]VNU University of Science

Vietnamese ginseng (VNG) or Ngoc Linh ginseng (*Panax vietnamensis* Ha *et* Grushv.) is an endemic species in VietNam and has been considered as precious medicine. Since being discovered so far, there have been numerous studies on VNG that focused on chemical compositions and their pharmacological characteristics. However research on bacterial community in VNG soil is lacked. Bacteria play an important role in improving soil structure and soil aggregation, recycling soil nutrients and water in the soil as well as interacting with plants. Therefore, understanding the natural bacterial community in ginseng soil would be effective way to reflect the health status of soil and the productive. Culture dependent methods was used to detect bacterial population in VNG soil (Nguyen et al., 2015; Tran et al., 2015). However, just a few of bacteria could be detected by this method.

In the past decades, next generation sequencing has improved our understanding of bacterial diversity in soil (Simon and Daniel, 2011). Nguyen et al. (2016) used 454 pyrosequencing to detect bacterial community in Korean ginseng cultivated soil in Korea. But the high cost of 454 pyrosequencing tools have limited small laboratories' access. Due to that, in 2001 Illumina developed MiSeq which have enabled deep sequencing of microbial communities at a lower cost (Caporaso et al., 2012). Thus, the aim of the present study was to Miseq to investigate bacterial community and diversity in VNG cultivated soil in the Ngoc Linh Mountain. We can detect bacterial population in soil at relative speed, and the ability to detect uncultivable organisms. We further predicted functional profiles from obtained 16S rRNA data.

I. MATERIALS AND METHODS

1. Sample preparation and DNA Extraction

Soil samples were collected in June 2016 from a ginseng cultivated area in the Ngoc Linh Mountain, Nam Tra My district, Quang Nam province, Vietnam (15°01'54"N, 107°58'45"E) where Vietnamese ginseng is originally detected. Soil samples were collected near rhizosphere of 6-year-old ginseng roots. Five samples were pooled. Soil samples were kept in Ziploc bags, then transferred to the laboratory, where they were stored in -20°C within one week for isolation of DNA. Genomic DNA was extracted using the PowerSoil® DNA Isolation Kit (MO Bio, CA, USA) following the manufacturer's instructions. After that, DNA was purified using Powerclean® DNA Clean-Up Kit (MO Bio, CA, USA). Subsequently, DNA was assessed quantity and quality. The passed DNA was sequenced using Illumina Miseq platform.

2. Primer design and library preparation

The V3-V4 hyper-variable regions of the 16S rDNA gene were amplified from the DNA extracts using universal primer 341F and universal primer 805F, which are amplicon primers as described in Table 1.

Table 1

Amplicon primers target V3-V4 region of the 16S rRNA gene

Primer	Left (29mer)	i5 (8mer)	Right (14mer)	Sequencing adaptor (19mer)	Target sequence (17mer)
S511 (Reverse)	AATGATACGG CGACCACCGA GATCTACAC	TCTCTC CG	TCGTCG GCAGCG TC	AGATGTGTAT AAGAGACAG	CCTACGG GNGGCW GCAG
N720 (Forward)	CAAGCAGAAG ACGGCATACG AGAT	AGGCTC CG	GTCTCG TGGGCT CGG	AGATGTGTAT AAGAGACAG	GACTACH VGGGTAT CTAATCC

The first PCR was carried out with primers that contained right, sequencing adaptor, and target sequence. The amplifications were carried out using an initial denaturation at 95°C for 3 min, followed by 25 cycles of denaturation at 95°C for 30 sec, primer annealing at 55°C for 30 sec, and extension at 72°C for 30 sec, with a final elongation at 72°C for 5 min, and hold at 4°C. The PCR product was used as a template in the second PCR. The primers for second PCR were left, i5, and right. PCR conditions were performed as above. After normalization, PCR products were pooled. The amplicon library was quantified using the Library Quantification Kit for Illumina (Kapa Biosciences, Woburn, MA, USA), denatured, and then sequenced on 250PE Miseq run at the Chunlab, Inc. (Seoul, Korea).

3. Bioinformatics analysis

After we obtained the raw sequences, the index sequences contained in the first 8 bp of each paired-end read were extracted. For metagenomics profiling, reads containing ambiguous bases (more than 2 Ns) or low quality bases (defined as average scores of <25) were filtered using Trimmomatic 0.32. Paired-end reads were overlapped using PANDAseq v.2.9 with a required overlap length of >300 bp. Primers were trimmed using pairwise alignment and the Hidden Markov Model. Clustering was performed using CD-HIT tool for de-noising. Taxonomic assignment was carried out by comparing the sequence reads against the EzTaxon-e database, using a combination of the initial BLAST-based searches and additional pairwise similarity comparisons. Then the UCHIME algorithm was used to detect chimeric sequences (Edgar et al., 2011). Taxonomic assignment was carried out by comparing the sequence reads against the EzTaxon-e database, using a combination of the initial BLAST-based searches and additional pairwise similarity comparisons. The following criteria were applied for the taxonomic assignment of each read (x = distance values): species ($x \leq 0.03$), genus ($0.03 < x \leq 0.05$), family ($0.05 < x \leq 0.1$), order ($0.1 < x \leq 0.15$), class ($0.15 < x \leq 0.2$), and phylum ($0.2 < x \leq 0.25$). If the distance was greater than the cutoff value, the read was assigned to an unclassified group. If the sequence cluster could not be identified with a valid name, the accession number of the GenBank sequence entry sharing the highest sequence similarity with the sequence cluster was used as a provisional name. Then the UCHIME algorithm was used to detect chimeric sequences (Edgar et al., 2011). The obtained high-quality reads were subjected to analyze diversity index in the software CLcommunity (Chunlab, Inc., Seoul, South Korea).

4. Inferred metagenomics by PICRUST

First, a collection of closed-reference OTUs was obtained from the filtered reads by using QIIME v 1.0.0 and by querying the data against a reference collection (GreenGenes database, May 2013 version; http://greengenes.lbl.gov) and OTUs were assigned at 97% identity. The

resulting OTU table was then used for microbial community metagenome prediction with PICRUSt on the online Galaxy interface (http://huttenhower.sph.harvard.edu/galaxy/). PICRUSt was used to derive relative Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway abundance. The nearest sequenced taxon index (NSTI) value show the level of uncertainty of the metagenome prediction, increasing the accuracy of the prediction the smaller are its values.

II. RESULTS AND DISCUSSION

1. Sequencing analysis

A total of 50,913 raw sequences were obtained. After removing of low quality reads, pair-end merging, trimming of primer and length, de-noising, and discarding chimera reads, 16,231 valid reads remained for further analyses (Table 2). The lowest length of read is 315 bp, the largest one is 461 bp, and the average length is 421 bp.

2. α-diversity

Rarefaction curves (Fig. 1) indicated that the number of detected OTUs increased with the number of sequences sampled in soil sample. The curve did not reached an asymptote. It interpreted to mean that more species present could been detected. The richness (OTUs) and diversity estimators (Ace, Chao1, and Shannon) are summarized in Table 3. Among 1795 of OTUs, 134 OTUs were singletons. Compared to Korean ginseng cultivated soil (Nguyen et al., 2016), number of OTUs as well as estimated richness Ace and Chao1 of Vietnamese ginseng cultivated is lower. But the Shannon index is similar with one of Korean ginseng soil.

3. Bacterial community

Mi-sequencing data revealed great bacterial diversity in the soil sample examined. The bacteria were from 32 phyla, 81 classes, 151 orders, 310 families, 652 genera, and 1833 species. Relative abundances of members of bacterial phyla comprising at least 1% of the community were shown in Fig. 2. The bacterial community was dominated by the phyla Acidobacteria (relative abundance 49.07%) and Proteobacteria (25.57%), followed by the phyla Verrucomicrobia (4.64%); Chloroflexi (3.54%), Bacteroidetes (3.51%), Actinobacteria (3.44%); Gemmatimonadetes (2.41%); Planctomycetes (2.22%); Cyanobacteria (1.15%). All remaining phyla were found at <1% in relative abundance in the sample.

Table 2

Illumina MiSeq sequencing and assembly metric

Metrics	Number
Total reads	50,913
Quality trimmed reads	48,012
Merged reads	47,397
Primer trimmed reads	47,053
Length trimmed reads	47,024
Sampled reads	20,000
Chimera reads	3614
Non-target reads	155
Valid reads	16,231
Min length	363
Max length	454
Average length	411

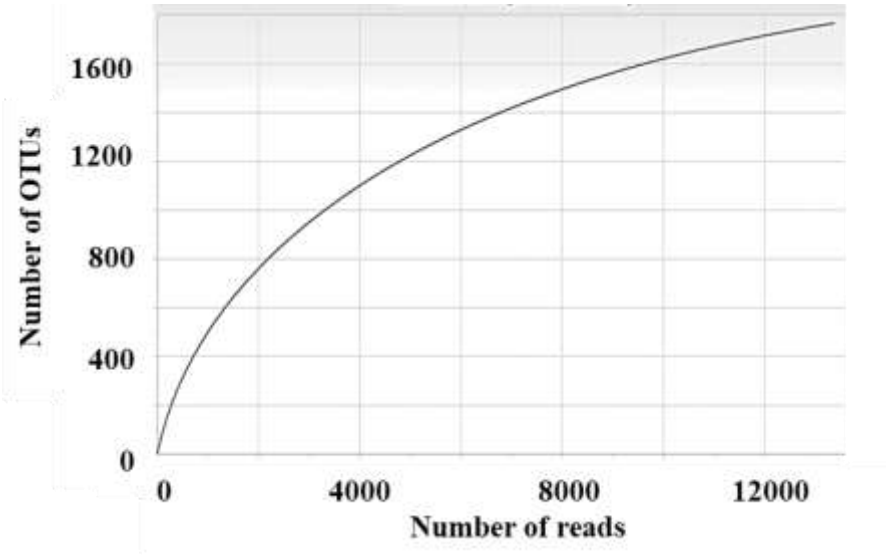

Figure 1: **Rarefaction curve.** OTUs are shown at the 3% genetic distance levels.

Table 3

Diversity indices obtained at a genetic distances of 3%

Target reads	Valid reads	OTUs	Ace	Chao1	Shannon	Goods Lib. Coverage
Value	16231	1795	2140	1996	6.48	99.69

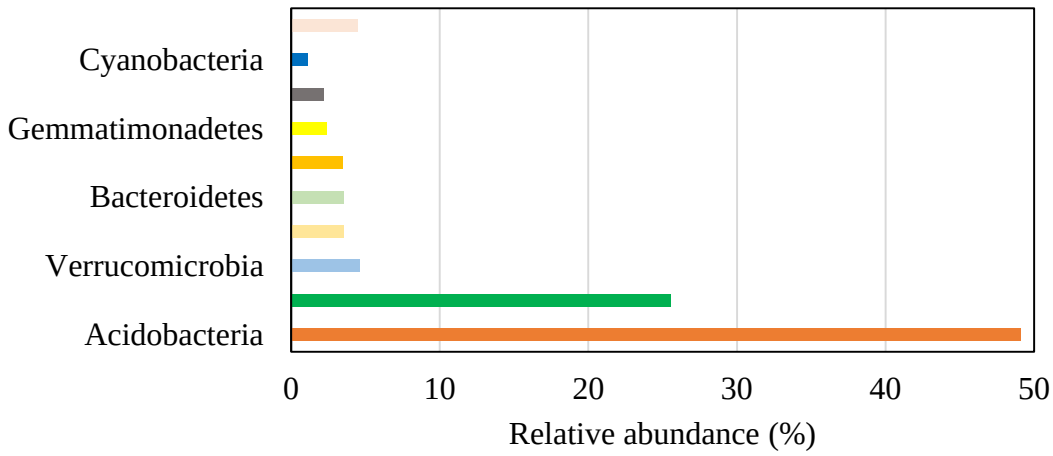

Figure 2: **Relative abundance of bacterial phyla in Vietnamese ginseng cultivated soil.** Relative abundances are reported as percent of total bacterial sequences observed per samples. The other category includes phyla showing a percentage of reads <1% of the total reads in all of the soil sample.

The predominant bacterial phyla Acidobacteria and Proteobacteria in this study agrees with the previous study in Korean ginseng soil (Nguyen et al., 2016), as well as in other agricultural soil (Kuramae et al., 2012; Lopes et al., 2013). Our soil sample was collected from bulk of 6-year-old ginseng roots, so it is not surprising when the ratio between Proteobacteria and Acidobacteria (approximately 0.52) in the soil sample is similar with those (approximately 0.5) of 6-year-old soil samples at the second round of cultivation of Korean ginseng soil (Nguyen et al, 2016). Members of Acidobacteria have been suggested to be adapted to nutrient-poor soils (Philippot et al., 2010; Chaudhry et al., 2012) and acidic tolerance (Hartman et al., 2008). Long time monoculture of ginseng cause low nutritional status of soil leading to abundance of Acidobacteria. Unlike Korean ginseng soil were originally paddy soil samples, our soil sample reside from forest soil, therefore, the phyla Chloroflexi only accounted small proportion in studied soil.

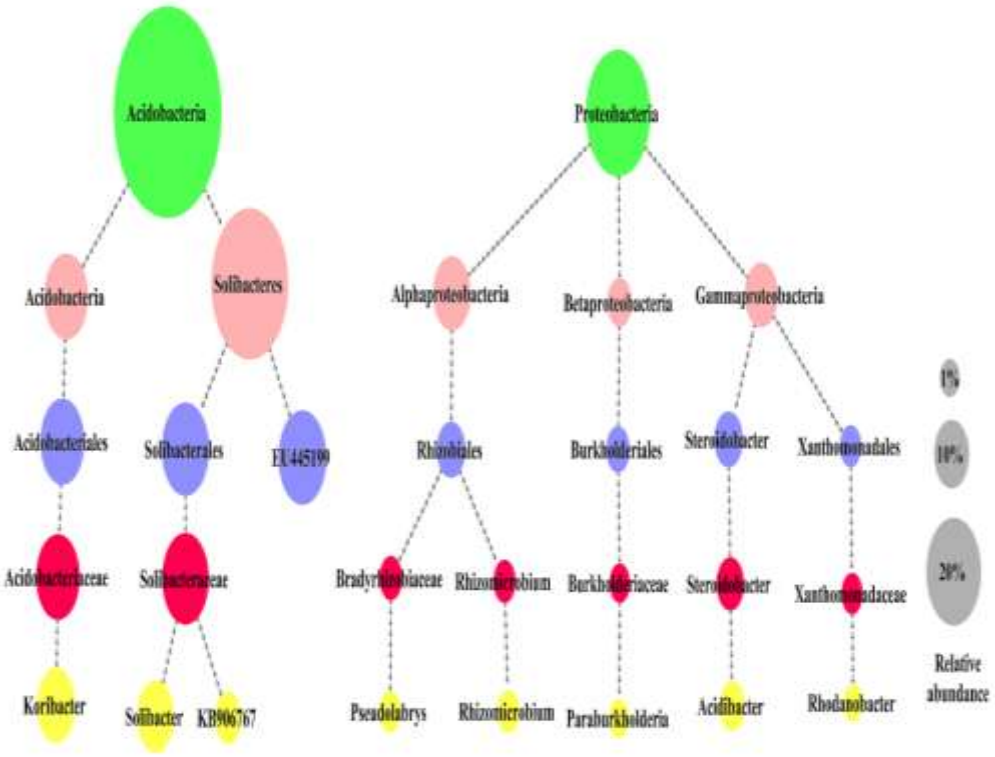

Figure 3: **Major compositions of the phyla Acidobacteria and Proteobacteria.**

Green, light pink, light cobalt blue, red, and pink circles indicate taxon at the level of phylum, class, order, family and genus, respectively. The size of circles represent to relative abundance of the taxon.

Detail taxonomic distribution of the phyla Acidobacteria and Proteobacteria is shown in Fig 3. The two most abundant classes of Acidobacteria (49.07%) were Acidobacteria (14.25%) and Solibacteres (32.27%) which dominant by three orders Acidobacteriales (14.25%), Solibacterales (16.21%), and unassigned-Solibacteres_EU44199 (16.50%), two families *Acidobacteriaceae* (14.21%), and *Solibacteraceae* (15.86%); and three genera *Koribacte*r (11.27%), *Solibacter* (10.41%), unassigned *Solibacteraceae*_KB906767 (4.84%). The majority of the sequences from Proteobacteria (25.57%) were affiliated to the three class

Alphaproteobacteia (11.00%), Betaprotebacteria (3.90%), and Gammaproteobacteria (8.42%). These three class comprised major amount of four orders Rhizobiales (5.95%), Burkholderiales (3.24%), Steroidobacter (5.27%), and Xanthomonadales (2.00%). These three orders divided to five major families *Bradyrhizobiaceae* (2.76%), *Rhizomicrobium* (2.45%), *Burkholderiaceae* (1.47%), *Steroidobacter* (5.27%), and *Xanthomonadaceae* (1.92%) and five major genera *Pseudolabrys* (1.89%), *Rhizomicrobium* (2.30%), *Paraburkholderia* (1.14%), *Acidibacter* (4.24%), and *Rhodanobacter* (0.89%).

Although *Koribacter* and *Solibacter* accounted major proportion in soil, however, these genera have not yet fully characterized. Until now, only *Koribacter versatilis* and *Solibacter usitatus* were sequenced whole-genome (Ward et al., 2009). The genome of the two strains encode the ability to degrade a variety of sugars, amino acids, alcohol, and metabolic intermediates. They also encoded the ability to use complex substrate such as chitin, starch, xylan, and cellulose. Therefore, acidobacteria can outcompete other bacterial species unable to use the complex substrates at low concentrations (Joseph et al., 2003).

Pseudolabrys and *Rhizomicrobium* represent the most abundant genera in the order Rhizobiales which are well-known beneficial partners in plant-microbe interactions such as plants hormones, auxin bioxynthesis, plant alkaloids, plant octadecanoids, nitrogen metabolism… (Erlacher et al., 2015). Changes in root exudates may be a major reason for high relative abundance of *Pseudolabrys* and *Rhizomicrobium* in soil.

Members of the *Burkholderiales* have versatile catabolic traits enabling them to degrade recalcitrant and aromatic compounds and survive in environments with limited nutrient availability (Li et al., 2012; Suárez-Moreno et al., 2012). Bacterial orders of *Xanthomonadales* and *Steroidobacter* that are known to be active under oxic conditions and can survive nutrient-poor conditions. Similarly members of the order *Xanthomonadales* were proposed to survive in niches where nutrients are limited by decomposing recalcitrant carbon sources such as hemicellulose (Déjean et al., 2013). That is reason this bacteria exist with high amount in 6-year-old Vietnamese ginseng soil.

Steroidobacter presents high proportion in the rhizosphere of soybean (Sugiyama et al., 2014) or rhizosphere of Jerusalem artichoke (Yang et al., 2016), so it can be understanding when it occupies high amount in our soil which was collected near rhizosphere of ginseng roots. Moreover, *Steroidobacter* is recognised to produce brassinosteroids which have been shown to control seed germination, stem and root elongation, vascular differentiation, leaf expansion and stress protection in plants (Fahrbach et al., 2008; Zarraonaindia et al., 2015). At the level of genus, *Acidibacter* spp. that adapt to low pH and low concentration of sugar (Falagán et al., 2014).The proportion of unclassified reads increased from 0.07% of all phyla to 4.21% of all classes, 24.74% of all orders, 34.55% of all families, 52.34% of all genera, and 95.32% of all species.

4. Metagenome prediction

Metagenome was predicted from the 16S rRNA gene sequences with an accuracy based on a NTSI average value 0.1894, which is typical for soil samples analyzed in other studies (Jiang et al., 2016)

The relative abundances of KEGG pathways at level 2 encoded in the microbiota present in the Vietnamese ginseng soil showed that carbohydrate metabolism, amino acid metabolism, membrane transport; transcription, replication and repair, energy metabolism were the most predominant microbiota activities (Fig. 4).

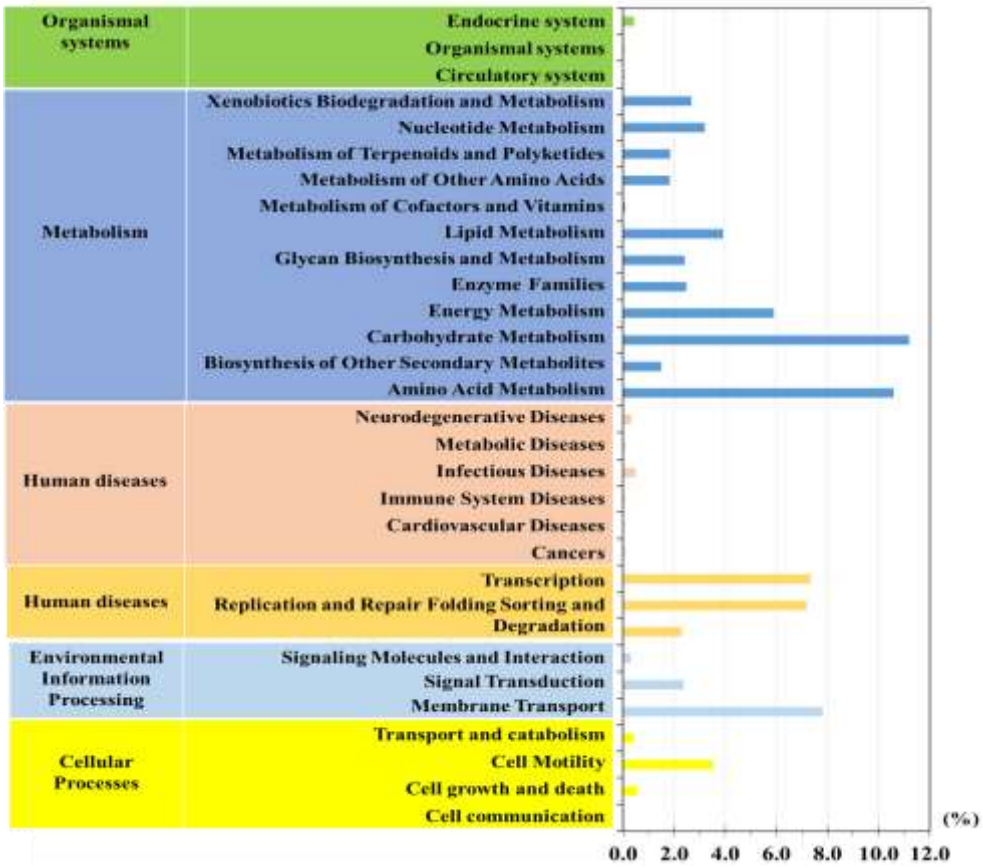

Figure 4: **Relative abundances of KEGG pathways at level 2 encoded in the bacterial community present in Vietnamese ginseng soil.**

III. CONCLUSIONS

This study presents the first comprehensive analysis of the structure of bacterial community in Vietnamese ginseng soil which includes 32 phyla, 81 classes, 151 orders, 310 families, 652 genera, and 1833 species. The bacterial population was predominant by Acidobacteria and Protebacteria. Understanding the patterns of microbial composition and diversity is a necessary first step in going on to assess the systemic effects of specific microbiota on their respective ginseng roots and ginseng soil.

Acknowledgements: *We would like to thank for the assistance of High-tech Biomedical Lab of National Center for Technological Progress in providing the facilities.*

REFERENCES

1. **Caporaso J. G., Lauber C. L., Walters W. A., Berg-Lyons D., Huntley J., Fierer N., Owens S. M., Betley J., Fraser L., Bauer M., Gormley N., Gilbert J. A., Smith G. & Knight R.,** 2012: Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal*, 6(8): 1621-1624.
2. **Chaudhry V., Rehman A., Mishra A., Chauhan P. S., Nautiyal C. S.,** 2012: Changes in bacterial community structure of agricultural land due to long-term organic and chemical amendments. *Microbial ecology*, 64(2): 450-460.

3. **Déjean G., Blanvillain-Baufumé S., Boulanger A., Darrasse A., Bernonville T.G., Girard A. L., Carrere S., Jamet S., Zischek C., Lautier M., Sole M., Buttner D., Jacques M. A., Lauber E. & Arlat M.,** 2013: The xylan utilization system of the plant pathogen Xanthomonas campestris pv campestris controls epiphytic life and reveals common features with oligotrophic bacteria and animal gut symbionts. *New Phytologist*, 198(3): 899-915.
4. **Edgar R. C., Haas B. J., Clemente J. C., Quince C., Knight R.,** 2011: UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16): 2194-2200.
5. **Erlacher A., Cernava T., Cardinale M., et al.,** 2015: Rhizobiales as functional and endosymbiontic members in the lichen symbiosis of Lobaria pulmonaria L. *Frontiers in microbiology*, 6: 53.
6. **Fahrbach M., Kuever J., Remesch M., Soh J., Sensen C.W., Grube M. & Berg G.,** 2008: Steroidobacter denitrificans gen. nov., sp. nov., a steroidal hormone-degrading gammaproteobacterium. *International Journal of Systematic and Evolutionary Microbiology*, 58(9): 2215-2223.
7. **Falagán C. & Johnson, D. B.,** 2014: Acidibacter ferrireducens gen. nov., sp. nov.: an acidophilic ferric iron-reducing gammaproteobacterium. *Extremophiles*, 18(6): 1067-1073.
8. **Hartman W.H., Richardson C. J., Vilgalys R. & Bruland G. L.,** 2008: Environmental and anthropogenic controls over bacterial communities in wetland soils. *Proceedings of the National Academy of Sciences*, 105(46): 17842-17847.
9. **Jiang L., Song M., Yang L., Zhang D., Sun Y., Shen Z., Luo C. & Zhang G.,** 2016. Exploring the Influence of Environmental Factors on Bacterial Communities within the Rhizosphere of the Cu-tolerant plant, Elsholtzia splendens. *Scientific Reports, 6*.
10. **Joseph S.J., Hugenholtz P., Sangwan P., Osborne C. A. & Jassen P. H.,** 2003: Laboratory cultivation of widespread and previously uncultured soil bacteria. *Applied and environmental microbiology*, 69(12): 7210-7215.
11. **Kuramae E. E., Yergeau E., Wong L. C., Piji A. S., van Veen J. A. & Kowalchuk G. A.,** 2012: Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS Microbiology Ecology*, 79(1): 12-24.
12. **Li R., Khafipour E., Krause D. O., Entz M. H, Kievit T. R & Fernando W. G. D.,** 2012: Pyrosequencing reveals the influence of organic and conventional farming systems on bacterial communities. *PloS one*, 7(12): e51897.
13. **Lopes A. R., Manaia C. M. & Nunes, O. C.,** 2014: Bacterial community variations in an alfalfa-rice rotation system revealed by 16S rRNA gene 454-pyrosequencing. *FEMS microbiology ecology*, 87(3): 650-663.
14. **Nguyen N. L., Kim Y. J., Hoang V. A., et al.,** 2015: *Paracoccus panacisoli* sp. nov., isolated from a forest soil cultivated with Vietnamese ginseng. *International journal of systematic and evolutionary microbiology*, 65(5): 1491-1497.
15. **Nguyen N. L., Kim Y. J., Hoang V.A., Tran B. T., Pham H. S., Yang D. C.,** 2016: Bacterial Diversity and Community Structure in Korean Ginseng Field Soil Are Shifted by Cultivation Time. *PloS one*, 11(5): e0155055.
16. **Philippot L., Andersson S.G., Battin T.J., Prosser J.I., Schimel J.P., Whitman W.B & Hallin S.,** 2010: The ecological coherence of high bacterial taxonomic ranks. *Nature Reviews Microbiology*, 8(7): 523-529.
17. **Simon C. & Daniel, R.,** 2011: Metagenomic analyses: past and future trends. *Applied and*

Environmental Microiology, 77: 1153–1161.

18. **Smit B., Pilifosova O., Burton I., et al.,** 2001: Adaptation to climate change in the context of sustainable development and equity. Climate Change 2001: Impacts, Adaptation and Vulnerability. Contribution of the Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change, J.J. McCarthy, O. Canziani, N.A. Leary, D.J. Dokken and K.S. White, Eds., Cambridge University Press, Cambridge, pp. 877-912.
19. **Suárez-Moreno Z.R., Caballero-Mellado J., Coutinho B.G., Mendoca-Previato L., James E.K. & Venturi V.,** 2012: Common features of environmental and potentially beneficial plant-associated Burkholderia. *Microbial Ecology*, 63(2): 249-266.
20. **Sugiyama A., Ueda Y., Zushi T., Takase H. & Yazaki K.,** 2014: Changes in the bacterial community of soybean rhizospheres during growth in the field. *PLoS One*, 9(6): e100709.
21. **Tran Bao Tram, Nguyen Thi Thanh Mai, Nguyen Thi Hien,** 2015: Investigation of the Fluctuation of Microorganic Populations in Soil Cultivated wwith Ngoc Linh Ginsseng. *Journal of Vietnam Science Technology*, No.24 (691): 47-50.
22. **Ward N. L., Challacombe J. F., Janssen P. H., et al.,** (2009). Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Applied and Environmental Microbiology*, 75(7): 2046-2056.
23. **Yan H., Hu J., Long X., Liu Z. & Rengel Z.,** 2016: Salinity altered root distribution and increased diversity of bacterial communities in the rhizosphere soil of Jerusalem artichoke. *Scientific reports*, 6.
24. **Zarraonaindia I., Owens S.M., Weisenhorn P., West K., Hampton-Marcell J., Lax S., Bokulich N.A., Mills D.A., Martin G., Taghavi S., Lelie D. & Gilbert J.A.,** 2015: The soil microbiome influences grapevine-associated microbiota. *MBio*, 6(2): e02527-14.

PHÂN TÍCH ĐA DẠNG QUẦN XÃ VI KHUẨN ĐẤT TRỒNG SÂM VIỆT NAM BẰNG KỸ THUẬT ILLUMINA MISEQ-BASED SEQUENCING

Nguyễn Ngọc Lan, Trần Bảo Trâm,
Phạm Hương Sơn, Phạm Thế Hải

TÓM TẮT

Sự tiến bộ về kĩ thuật giải trình tự gen thế hệ mới đã cải thiện đáng kể việc đánh giá trực tiếp toàn bộ hệ gen của quần xã vi sinh vật trong đất. Trong đó, Miseq được đánh giá là kĩ thuật có mức chi phí vừa phải nhưng khá hiệu quả trong phân tích hệ gen vi sinh vật trong đất. Do đó trong nghiên cứu này chúng tôi sử dụng kĩ thuật Illumina Miseq để xác định quần xã vi khuẩn trong đất trồng sâm Việt Nam ở núi Ngọc Linh thuộc tỉnh Quảng Nam bằng việc giải trình tự khu V3-V4 của gen 16S rRNA. Phân tích số liệu cho thấy sự đa dạng vi khuẩn trong đất trồng sâm với sự có mặt của 32 ngành, 81 lớp, 151 bộ, 310 họ, 652 chi, và 1833 loài. Trong đó, hai ngành ưu thế là Acidobacteria và Proteobacteria, tương ứng chiếm 49,07% và 25,57% tổng số loài. Có hai họ đã được định danh và công nhận *Solibacteraceae* và *Acidobacteriaceae* cùng với 1 họ không xác định unassigned-Acidobacteria-EU44199 là 3 họ lớn. Trong khi đó, *Koribacter* và *Solibacter* là hai chi đã biết có số lượng lớn nhất trong đất trồng sâm Việt Nam. Chúng tôi cũng dự đoán chức năng của hệ vi sinh vật trong đất trồng sâm thông qua phần mềm PICRUSt. Các gen được chú giải trong các con đường chuyển hóa khác nhau của KEGG và phân bố chủ yếu trong các chức năng chuyển hóa carbon, chuyển hóa amino acid, vận chuyển màng, phiên mã, sửa chữa và sao chép, và chuyển hóa năng lượng.