Philippine Journal of Science
152 (5): 1737-1751, October 2023
ISSN 0031 - 7683
Date Received: 22 Mar 2023

Bacterial and Fungal Community Profiling of Karst Ecosystem in Basey, Samar, Philippines Using Shotgun Metagenomic Approach

John Daniel P. Ong[1], Edmerson B. Geronimo[1], Eula Francia M. Bosito[1], Kristiann Gabriele D.M. Villarama[1], Zachary B. Lara[2], Florabelle D. Querubin[1], and Marilen P. Balolong[2†]

[1]Microbiology Division, Institute of Biological Sciences, College of Arts and Sciences, University of the Philippines Los Baños, Los Baños, Laguna 4030 Philippines
[2]Department of Biology, College of Arts and Sciences, University of the Philippines Manila, Ermita, Manila Philippines

The Philippines has an abundance of karst or forest over limestone landscapes, but only a small percentage of them have been studied and protected. Although the flora and fauna of karst forests are diverse and unique, little is known about the microflora that inhabits this ecosystem. The taxonomic and functional composition of bacteria and fungi in soil samples from three locations within three karst forest zones in Basey, Samar, Philippines were analyzed using Illumina shotgun metagenome sequencing. Proteobacteria and Actinobacteria bacterial sequences were most prevalent in the karst soil, followed by those of the Firmicutes, Acidobacteria, Chloroflexi, Planctomycetes, Cyanobacteria, Verrumicrobia, Bacteriodetes, and Deinococcus-Thermus. The most abundant fungal sequences belonged to Ascomycota, followed by Basidiomycota. An average of 33 million predicted protein features was detected across all sites. Enzyme pathways for nitrogen and sulfur metabolism, and several carbon fixation pathways, appeared nearly complete. To our knowledge, this is the first report to provide baseline information on the microbial community and their possible roles in karst forest ecosystem health in the Philippines, which may lead to identifying new microbes with specialized metabolism and promoting biodiversity conservation of karst forests in the Philippines. Furthermore, correlation analysis with plant diversity will reveal plant-microbe interaction leading to the understanding of the adaptation, abundance, survival, and diversity of microorganisms and plants.

Keywords: functional diversity, karst, MG-RAST, microbial diversity, Philippines, shotgun metagenomic sequencing

INTRODUCTION

Forests over limestone, more popularly known as karst ecosystems, are one of the most ecologically fragile ecosystems in the world. These landscapes are threatened by anthropogenic activities and natural processes. Thin, gritty, easily eroded, and deteriorating are typical characteristics of karst soil (He *et al.* 2008). Almost one-quarter of the global population resides in karst regions (Fleury 2009; Drew 2017). It encompasses around 30,000 km^2 or 10% of the Philippines' total geographical area, although a substantial section of it has not yet been investigated and protected. Vegetation development is stunted in karst regions due to the high levels of

*Corresponding author: jpong2@up.edu.ph
†Deceased

exchangeable calcium in the soil. Symbiotic or associative microbes such as *Rhizobium* sp., *Bradyrhizobium* sp., *Frankia* sp., and *Azospirillum* sp. provide plants with the adaptive potential to grow in such an environment, either in part or entirely (Li *et al*. 2018). Since they have such a large impact on soil productivity and the distribution and retention of soil nutrients, soil microbial communities (SMCs) are essential to ecosystem recovery and maintenance, especially in karst environments (Chen *et al.* 2012). Furthermore, these microbes drive biogeochemical processes and mediate nutrient turnover, both of which have significant effects on biodiversity and ecosystem functioning in terrestrial ecosystems (Doran and Zeiss 2000; Bardgett and van der Putten 2014). Additional roles of soil microorganisms include the synthesis of phytohormones, the cycling of nutrients, the remediation of polluted soils, the biocontrol of soil-borne plant pathogens, and the organic matter decomposition. However, the processes mediated by SMCs – as well as the limiting resource for microbial growth in karst ecosystems – remain poorly understood (Chen *et al.* 2018).

One of the important soil microorganisms are actinomycetes. Actinomycetes have been identified as potential sources of a variety of secondary metabolites, antibiotics, and bioactive substances that influence the growth of microorganisms (Chaudhary *et al.* 2013). Antimicrobial compounds extracted from actinomycetes were even found to inhibit the growth of multi-drug resistant strains of *Escherichia coli*, *Staphylococcus aureus*, and Enterococci (Singh *et al.* 2012). Other applications of actinomycetes include the transformation of xenobiotic compounds, immunomodifiers, biosurfactants, and enzyme inhibitors (Chavan *et al.* 2013).

Another important group of soil microbes is fungi. Fungi are involved in various processes vital in the cycling of nutrients in karst ecosystems. These include organic matter decomposition, increase of availability of nutrients by mineralization, protection against leaching of nutrients by accumulating biomass (Fan *et al.* 2019), and moisture preservation by sealing microcracks in rocks, which boosts the effectivity of bacterial by-products (Lian *et al.* 2010).

The availability of resources frequently limits the growth and activity of soil microorganisms. Given the importance of SMCs to the soil's biogeochemical cycle, understanding the soil processes that are associated with these microorganisms is crucial for understanding the factors that regulate soil fertility and for predicting how an ecosystem will react to a changing climate. Moreover, baseline data generated from this study may provide valuable insight into this ecosystem and may pave the way for long-term monitoring of microbial diversity. Despite the critical role that SMC-plant interactions play in controlling the organization and function of karst ecosystems, the variety of SMC in this type of landscape has received little attention.

Karst landscapes have been researched extensively in China, Europe, and Mexico's Yucatán Peninsula (Santillán *et al.* 2021). In other countries, the metagenomic approach has been used to effectively report on the soil's microbial diversity and its functional potential as a plant growth promoter and source of antibiotic resistance genes (Cheng *et al.* 2021; Liu *et al*. 2021). For the first time, we use metagenomic analysis to report on the microbial community composition, taxonomic diversity, and metabolic activities of a forest over a limestone habitat in Basey, Samar, Philippines. The shotgun metagenomic approach allows for higher resolution in the creation of taxonomic and functional profiles because of the depth of information provided by this when compared to amplicon and culture-dependent techniques (Liu e*t al.* 2020). The taxonomic and functional profile of the karst ecosystem's SMC will potentially accelerate research on natural microbial communities, thereby promoting the adaptive capacity of host plants to abiotic stresses such as high calcium stress. Karst forests are more prone to rapid degradation processes as compared to non-karst forests such as soil loss and reduced water holding capacity, resulting in irreversible changes in vegetation cover (Peng e*t al.* 2013; Tang 2013). Therefore, understanding the microbial community structure is vital for planning methods of intervention that aim to effectively restore vegetation in karst areas. Data from this study will also be valuable for determining which microbial strains can be used for field application or for further culture-dependent study in karst topography.

MATERIALS AND METHODS

Study Area and Sampling

In October 2019, soil samples were collected and pooled from approximately 3–5 points within each sampling site and placed in individual sterile plastic bags (Figure 1; Table 1). The study sites were in three different areas of the Samar Island Natural Park, a known limestone forest in Basey, Samar (UNDP 2012; Tolentino *et al.* 2020).

Table 1. Coordinates and elevation of the sampling sites in Basey, Samar.

Location	Site	Coordinates
Basey	B1	11°21'11"N 125°10'04"E
	B2	11°21'11"N 125°10'01"E
	B3	11°21'16"N 125°09'58"E

Figure 1. Geographical location of sampling points in karst forest areas of Basey, Samar, Philippines, relative to each other.

Approximately 2 kg each of samples labeled as B1, B2, and B3 were sourced from the soil surface, with a depth ranging from 0–10 cm in contact with the limestone. The first subset of the samples was stored at –80 °C, whereas the second subset was set aside for analyses of the soil's physicochemical parameters.

Soil Physicochemical and Biological Parameters

Soil samples were sent to the Analytical Services Laboratory, Division of Soil Science, Agricultural Systems Institute, College of Agriculture and Food Science, University of the Philippines Los Baños, for analyses. The following parameters were measured: pH, soil organic matter (SOM), available phosphorus, total nitrogen, exchangeable sodium, calcium, magnesium, and phosphorus, trace elements (Fe, Zn, Cu, and Mn), and electrical conductivity (EC).

DNA Extraction and Purification

DNA extraction was performed using MP BIO FastDNA™ Spin Kit for Soil (Irvine CA, USA), following the manufacturer's protocol with several modifications. The modifications made in the protocol were as follows: [a] 400 mg of soil sample was used to prevent overflow upon addition of buffers; [b] cell lysis: bead beating was employed for 40 seconds at 2,400 rpm or 5,645 x g with incubation on ice every 10 s, in order to avoid overheating of the sample and tube, as well as the degradation of genetic material; [c] protein denaturation: after addition of protein precipitating solution (PPS), samples were placed on an orbital shaker to facilitate the binding of PPS to protein contaminants and effective removal of the contaminants after decanting. The quality and quantity of the extracted DNA were evaluated using electrophoresis (Mupid ® - One, Chuo-ku, Tokyo, Japan) on a 1% agarose gel and Multiskan™ SkyHigh Microplate spectrophotometer (ThermoFisher Scientific, USA). DNA samples were kept at –20 °C.

Next-Generation Sequencing and Metagenomic Analysis

The DNA samples were submitted to Apical Scientific for next-generation shotgun metagenomic sequencing. Library preparation and metagenomic shotgun sequencing were performed by First Base Laboratories-Apical Scientific (Selangor, Malaysia) using an Illumina MiSeq instrument (Illumina, California, USA) which generated reads with an average length of 246 bp (Supplementary Table S1). The whole genome sequences are available from NCBI with the following BioSample accession numbers: SAMN26879029, SAMN26879030, and SAMN26879031.

MG-RAST version 4.0.3 (http://metagenomics.anl.gov/) was used to process and analyze raw sequences from Illumina Miseq. The data was uploaded as FASTAQ

files and the paired ends were joined. The joined paired-end reads were then subjected to the MG-RAST pipeline analysis (Supplementary Figure S2). To trim low-quality regions, from the FASTQ data, the uploaded data were preprocessed using SolexaQA. To remove artifactual duplicate reads (ADRs), de-replication was then conducted using a simple k-mer approach. These ADRs were then analyzed using duplicate read inferred sequencing error estimation (DRISEE) to determine the degree of variation among prefix-identical sequences in the template. The sequences were screened using Bowtie 2.4.2 so that only reads that do not match the model organisms would proceed to the annotation pipeline. Gene calling was performed using FragGeneScan to predict proteins or protein fragments from *de novo* sequence data. In order to preserve the relative abundances, QIIME was used to build clusters of proteins at the 90% identity level. A representative for each cluster was then subjected to similarity analysis. Instead of using BLAST, functional identification of representative species was done using sBLAT. Sequence similarity searches were then computed against a protein database derived from the MD5-based non-redundant protein database M5NR (Wilke *et al.* 2012). By doing so, it's possible to use different databases such as COG (Clusters of Orthologous Genes), KO (KEGG Orthology), NOG (eggNOG functionally annotated orthology), and SEED (subsystems annotation) subsystems without recomputation (Keegan *et al.* 2016). The taxonomic assignment was accomplished based on the RefSeq database, whereas the functional assignment was performed based on COG, KO, NOG, and SEED. The maximum E-value was $1e^{-5}$, the minimum sequence identity was 60%, and the minimum alignment length was 15 bases. MG-RAST was also used to perform an analysis of species richness *via* alpha-diversity and rarefaction. Visualization of the data was done using R-studio, Microsoft Excel, and the visualization tools in MG-RAST.

RESULTS

Soil Physicochemical and Biological Parameters

The soil physicochemical parameters are shown in Table 2. The soil pH values for all areas in Basey were not different from each other (7.1, 6.9, and 7.6, for samples B1, B2, and B3, respectively). In terms of salinity, B1 had the highest EC (77 µs/cm), followed by B2 (12 µs/cm). Sample B3 had the lowest EC (0.174 µs/cm). Of the three sites, B3 had the highest SOM (9.66%), followed by B1 (8.83%) and B2 (8.03%). The concentration of nitrogen is low in all sites (0.63, 0.54, and 0.41 for B3, B1, and B2, respectively). The amount of phosphorus is highest in B2 (13 mg/kg), followed by B3 (12.3 mg/kg) and B1 (11.5 mg/kg). Calcium concentration is high for all areas, with B3 having the highest Ca concentration (23.42 $cmol_c$/kg), followed by B1 (19.68 $cmol_c$/kg) and B2 (15.65 $cmol_c$/kg). The concentration of Mg is also high for all areas, with B3 having the highest Mg concentration (25.21 $cmol_c$/kg), followed by B2 (23.49 $cmol_c$/kg) and B1 (19.21 $cmol_c$/kg). Potassium concentration is highest for B2 (0.31 $cmol_c$/kg), followed by B1 (0.27 $cmol_c$/kg) and B3 (0.22 $cmol_c$/kg). The concentration of various micronutrients (Fe, Zn, Cu, and Mn) is also indicated in Table 2.

Table 2. Physicochemical and biological factors of the tested karst soil samples from Basey, Samar, Philippines.

Physicochemical and biological factors	Karst soil samples		
	B1	B2	B3
pH	7.10	6.90	7.60
SOM (%)	8.83	8.09	9.66
EC (µS/cm)	77.00	12.00	0.17
N (%)	0.54	0.41	0.63
P (mg/kg)	11.50	13.00	12.30
K ($cmol_c$/kg)	0.27	0.31	0.22
Na ($cmol_c$/kg)	1.46	1.40	1.69
Ca ($cmol_c$/kg)	19.68	15.65	23.42
Mg ($cmol_c$/kg)	19.21	23.49	25.21
Fe (ppm)	6.74	3.41	4.72
Zn (ppm)	1.30	0.76	0.72
Cu (ppm)	11.74	9.82	5.99
Mn (ppm)	11.75	9.52	5.07

Sequencing Read Statistics and Protein Prediction

In B1, of the 34,982,454 sequences (totaling 8,688,711,024 bp) that passed quality control, 7,103,551 sequences (20.3% of the total) were identified as ADRs. Of the sequences without rRNA genes, 33,121,839 contained predicted protein features, 14,674,548 (44.3%) of which were assigned an annotation using at least one protein database, and 18,447,291 (55.7%) contained predicted proteins with unknown function.

In B2, of the 28,632,312 sequences (totaling 7,128,702,401 bp) that passed quality control, 5,783,510 (20.2% of the total) were identified as ADRs. Of the sequences without rRNA genes, 27,381,495 contained predicted protein features, 12,367,805 (45.2%) of which were assigned an annotation using at least one protein database, and 15,013,690 (54.8%) contained predicted proteins with unknown function.

In B3, of the 41,417,390 sequences (totaling 10,223,783,602 bp) that passed quality control, 7,579,424 (18.3%) were identified as ADRs. Of the sequences without rRNA genes, 39,908,629 contained predicted protein features, 17,296,006 (43.3%) of which were assigned an annotation using at least one protein database, and 22,612,623 (56.7%) contained predicted protein proteins with unknown function.

Each sample yielded over 3.4608×10^7 sequence reads. After quality filtering, the minimum number of reads per sample was 2.8623×10^7. The reads for all sites had the same average GC content (65 ± 8%). The minimum identified protein features per sample were 1.2367×10^7 (Supplementary Table S1).

Taxonomic Profiles

Alpha diversity of the metagenomes summarized the diversity of all organisms. The α-diversity of B1, B2, and B3 were 462, 471, and 508, respectively. The rarefaction curve plateau to the right, with a maximum E-value cut-off of $1e^{-5}$, showed the species richness in the soil from the karst ecosystem (Supplementary Figure S3).

The numbers of sequences affiliated with each bacterial taxon in karst soil were similar across samples. In all areas, Bacteria is the most dominant domain (98%), followed by Eukarya (0.8–1%). Archaeal (0.9–1%) and unclassified sequences (0.001–0.05%) were also present. After analysis of Domain Bacteria, a total of 28 (1 unclassified) phyla, 51 (9 unclassified) classes, 110 (18 unclassified) orders, 244 (36 unclassified) families, and 596 (34 unclassified) genera were detected.

Based on the RefSeq database, the number of sequences affiliated with each bacterial taxon in the karst soil is almost similar across all karst soil samples from Basey, Samar. The most dominant taxa are *Actinobacteria* (B1: 38.4%, B2: 37.2%, and B3: 32.5%) and *Proteobacteria* (B1: 34.5%, B2: 35.4%, and B3: 37.8%). In all karst soil samples, there is an abundance of Firmicutes (5.8–6.9%), Acidobacteria (3.6–4.3%), Chloroflexi (3.2–3.4%), Planctomycetes (2.7–2.9%), Cyanobacteria (2.6–2.8%), Verrumicrobia (2.1–2.2%), and Bacteriodetes (2.1–2.4%). Minor groups represented at the phylum level included Chlorobi, Nitrospirae, Aquificae, Candidatus Poribacteria, Chrysiogenetes, Chlamydiae, Deferribacteres, Deinococcus-Thermus, Dictyoglomi, Elusimicrobia, Fusobacteria, Gemmatimonadetes, Lentsphaerae, Spirochaetes, Synergistetes, Tenericutes, and Thermotogae. A small percentage (0.3%) of the sequences were unclassified (Figure 2; Supplementary Table S2).

As for the fungal groups, a total of 6 (1 unclassified) phyla, 18 (1 unclassified) classes, 42 (1 unclassified) orders, 92 (7 unclassified) families, and 132 genera of fungi were detected. Based on the RefSeq database, the number of sequences affiliated with each fungal taxon in the karst soil was similar across all samples, with a dominance of Ascomycota (86.2–91.6%) and an abundance of Basidiomycota (8.3–13.6%). Minor groups represented at the phylum level included Blastocladiomycota (0.02–0.03%), Chytridiomycota (0.02–0.03%), and Glomeromycota (0.004–0.01%). The most dominant fungal classes are Sordariomycetes (B1: 39.9%, B2: 27.9%, and B3: 28.0%), and Eurotiomycetes (B1: 33.6%, B2: 37.5%, and B3: 36.9%). In all areas, there is an abundance of Saccharomycetes (4.7–9.4%), Agaricomycetes (4.9–7.4%), Dothideomycetes (3.7–6.0%), Leotiomycetes (3.6–4.6%), Schizosaccharomycetes (1.6–3.1%), Ustilaginomycetes (1.6–2.9%), and Tremellomycetes (1.5–2.8%) (Supplementary Table S3). Minor groups represented at the class level included Blastocladiomycetes, Chytridiomycetes, Exobasidiomycetes, Glomeromycetes, Monoblepharidomycetes, Pneumocustidomycetes, and Taphrinomycetes. A small percentage of the sequences were unclassified (0.02–0.05%) (Figure 3; Supplementary Table S3).

Functional Categories

Metabolic profiles were constructed using COG, NOG, KO, and SEED subsystem database that allows the comparison of the homology of functional genes with the database and show a group of annotated genes with the metagenomic samples (Figure 4).

The predicted protein classification showed that metabolism (51%) and poorly characterized proteins (47–48%) were predominant as per COG and NOG databases, respectively. Based on KO, metabolism (60–61%) was most abundant, followed by environmental information processing (17–18%), genetic information processing (16%), cellular processes (4%), human diseases (2%), and organismal systems (0.3%). Carbohydrate metabolism (13%) and

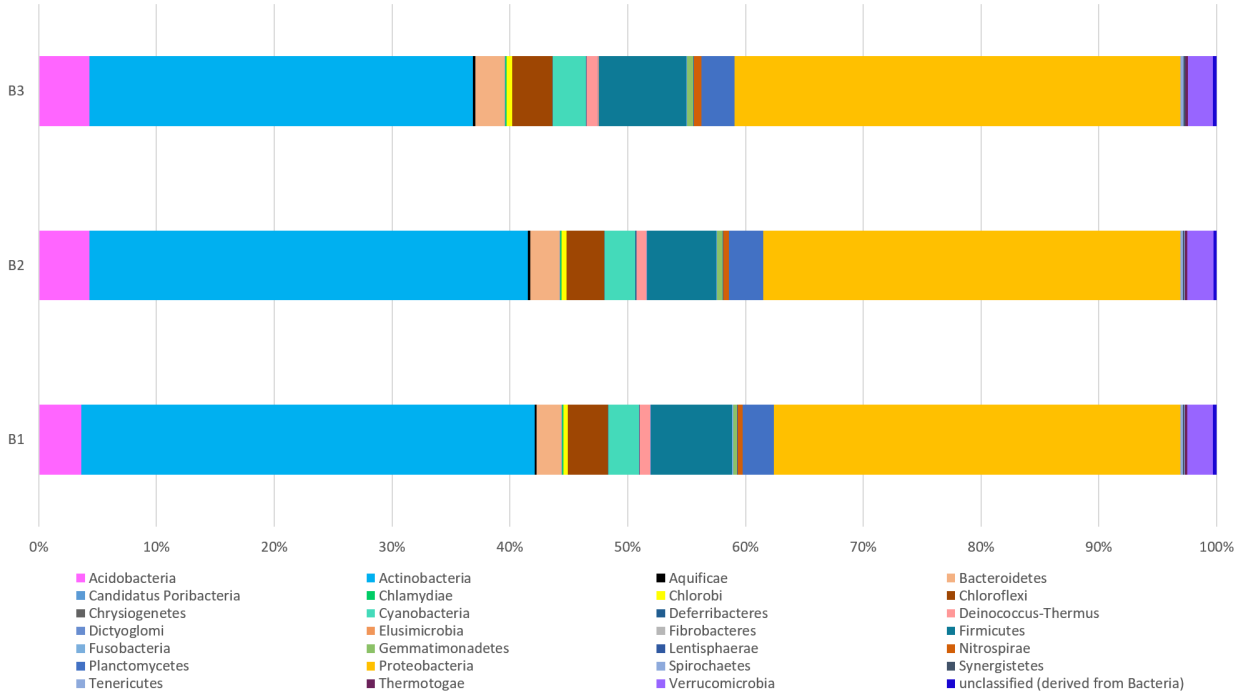


Figure 2. Structure of bacterial phyla detected in the soil samples obtained from the karst forest areas in Basey, Samar, Philippines.

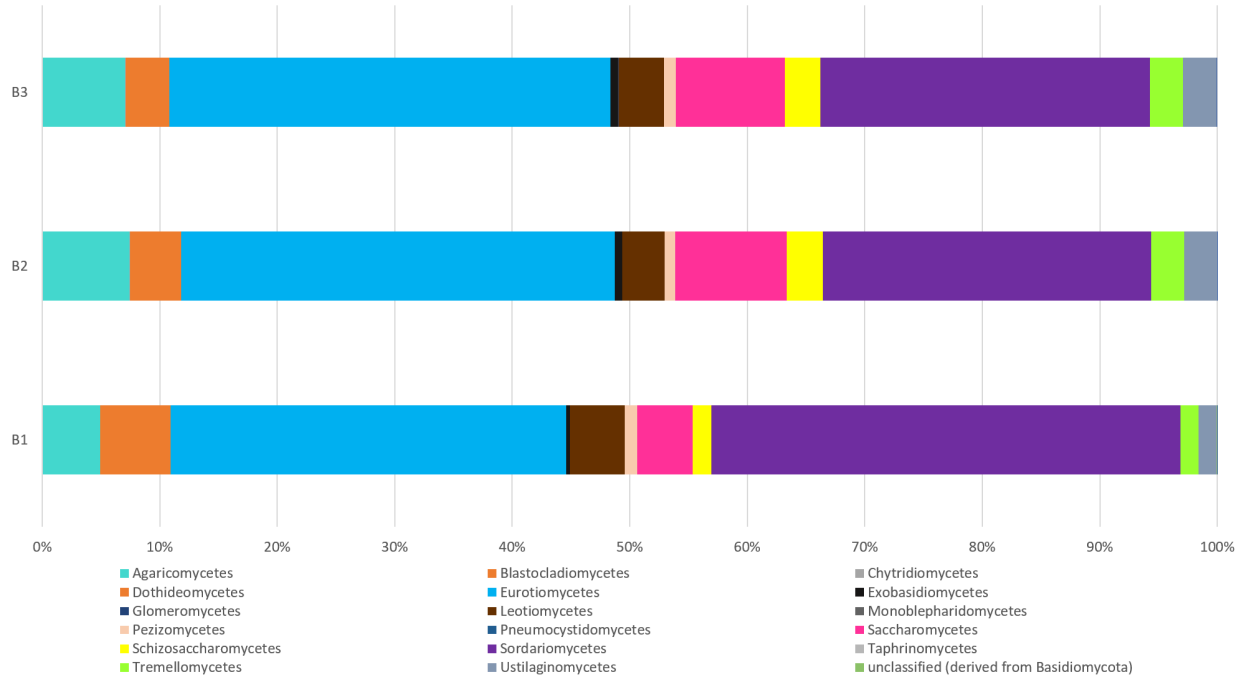


Figure 3. Structure of fungal classes detected in the soil samples obtained from the karst forest areas in Basey, Samar, Philippines.

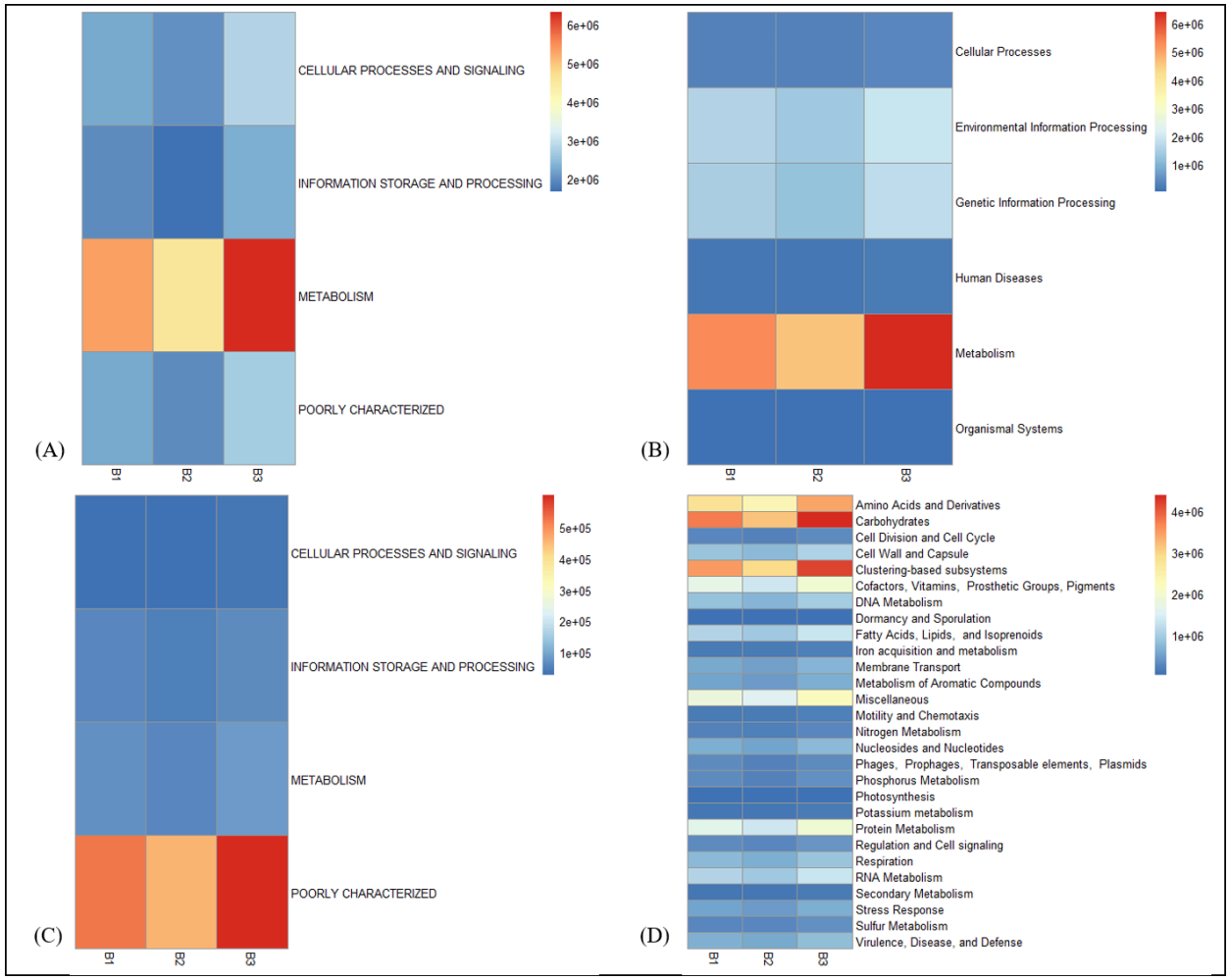


Figure 4. Heat map of the functional profiles for the microbial metagenomes of the soil from the karst forest in Basey, Samar. The figure displays the num sequence counts found for each metagenome for [a] COG, [b] KO, [c] NOG, and [d] SEED subsystem.

clustering-based system (13%) dominated functional classification by SEED subsystem database, followed by amino acids and derivatives (10%), protein metabolism (8%), miscellaneous (7%), cofactors, vitamins, prosthetic groups, pigments (6%), respiration (5%), membrane transport (4%), DNA metabolism (4%), RNA metabolism (3%), cell wall and capsule (3%), nucleosides and nucleotides (3%), virulence, disease, and defense (3%), fatty acids, lipids, and isoprenoids (3%), stress response (3%), metabolism of aromatic compounds (2%), nitrogen metabolism (1%), phosphorous metabolism (1%), phages, prophages, transposable elements, plasmids (1%), sulfur metabolism (1%), regulation and cell signaling (1%), potassium metabolism (0.9%), cell division and cell cycle (0.9%), iron acquisition and metabolism (0.7%), motility and chemotaxis (0.6%), secondary metabolism (0.3%), dormancy and sporulation (0.2%), and photosynthesis (0.1%).

The genes encoding for carbon and nitrogen metabolism pathways in bacteria and fungi were elucidated (Supplementary Table S4). It showed that carbon metabolism pathways – including carbon fixation pathways (Figure 5), glycolysis pathway, pentose phosphate pathway, and methanogenesis pathway – possibly exist in the karst forest of Basey, Samar. Enzymes for nitrogen and sulfur metabolism pathways were also detected (Figures 6 and 7). Functional prediction revealed enzymes involved in nitrogen fixation, denitrification, nitrate reduction, and nitrification (Supplementary Table S4).

DISCUSSION

Soil Physicochemical and Biological Parameters

Soil physicochemical parameters were usually regarded as indicators of soil quality. The maintenance of soil quality is essential to ensure the viability of the ecosystem. Soils are generally considered saline if the EC is 2–4 dS/m or

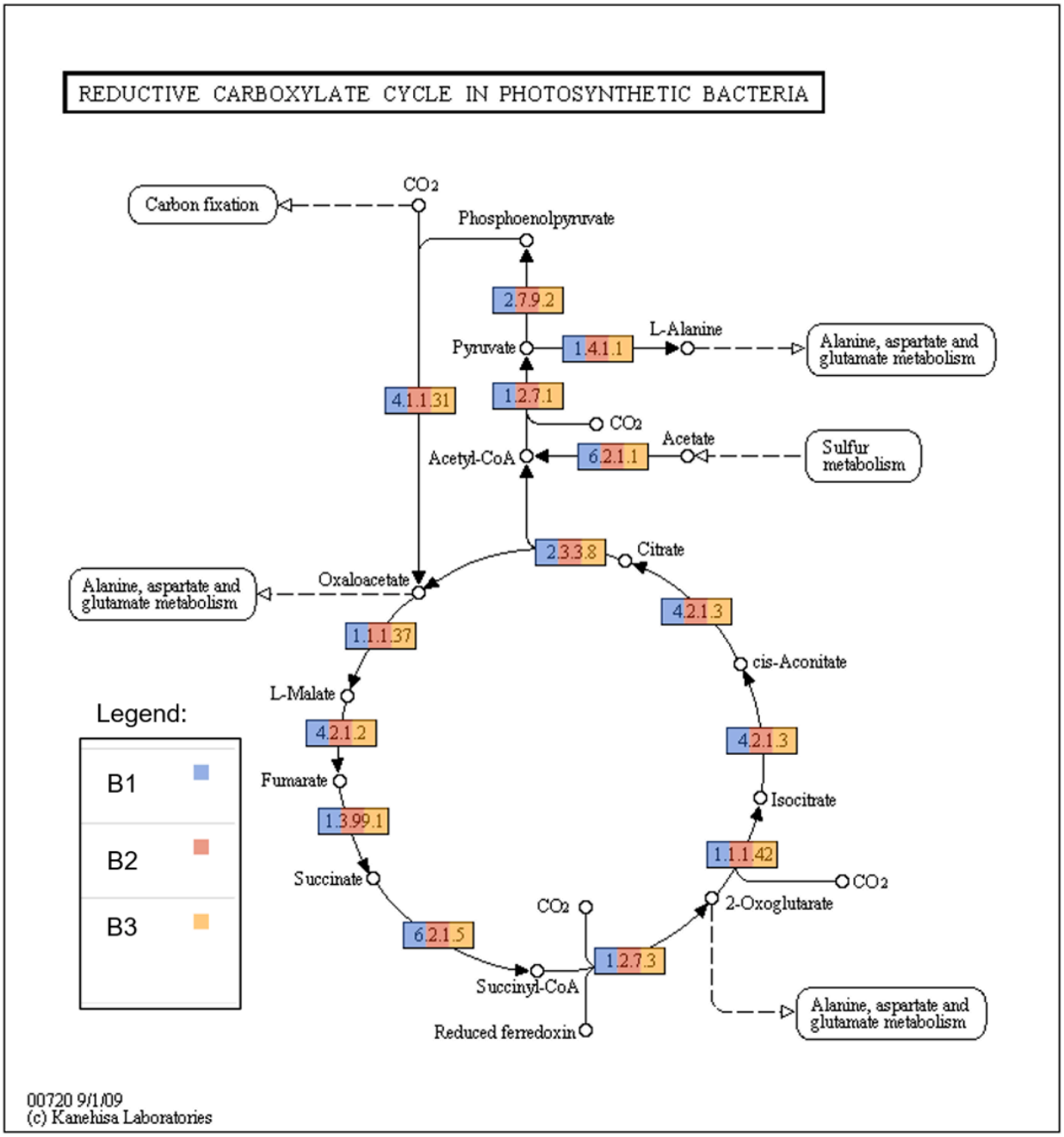


Figure 5. Reductive carboxylate cycle in photosynthetic bacteria performed by KEGG mapper from MG-RAST. Colored EC numbers are the enzymes identified in the three karst areas.

2000–4000 µS/cm (Ong *et al.* 2018). Soil pH of all three areas falls within the neutral range. This corroborates the study conducted by Chen *et al.* (2018), wherein they compared the soil parameters of the forests, both karst and non-karst. The soil pH of the karst forest (6.6 ± 0.1) was higher in the former than in the latter by 12% (5.9 ± 0.1). In another study, the soil pH ranged from a relatively acidic pH of 5.44 under the shrubland stage to a neutral pH of 7.46 under the natural forest stage (Zhao *et al*. 2019). Soil microorganisms' activities are at their highest when the pH is near neutral; also, a near-neutral soil pH makes most minerals readily available to plants (McCauley *et al.* 2009). The SOM content in all three areas is high. In karst forests, plant residues, and litter were less impacted, which led to the easier decomposition of these materials, thus leading to high SOM content. Nitrogen concentration is low in all three areas. The high $CaCO_3$ in the soil in a karst ecosystem has a profound impact on stabilizing

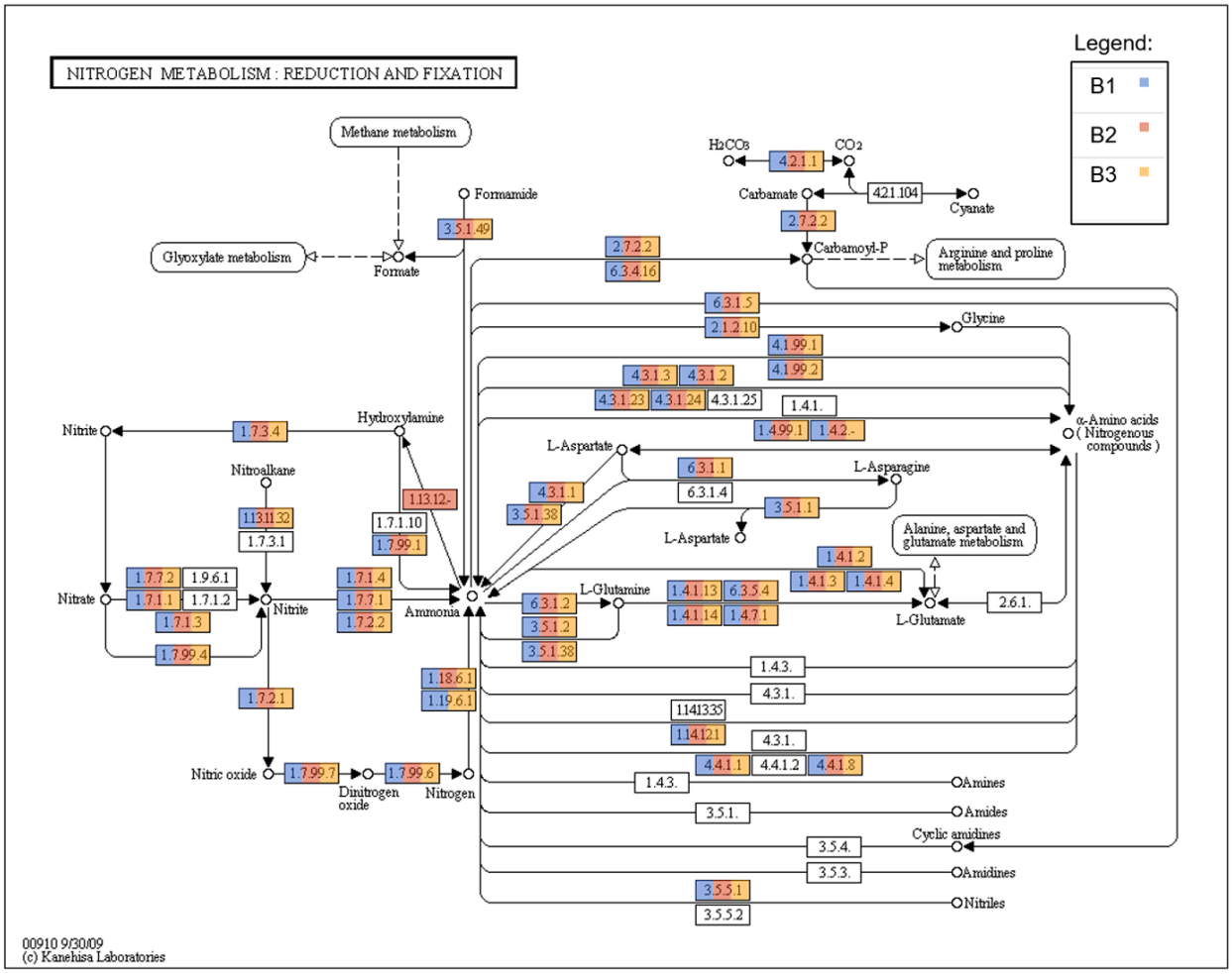


Figure 6. Nitrogen metabolic pathway performed by KEGG mapper from MG-RAST. Colored EC numbers are the enzymes identified in the three karst areas.

SOM, thus resulting in the SOM releasing low amounts of phosphorus, carbon, and nitrogen (Pan *et al.* 2016). The findings on the value of the E-Ca^{2+} are consistent with the study of Li *et al.* (2021) on the comparative analysis of soil bacteria in different karst landscapes in southwest China, in which the measured E-Ca^{2+} was 5.16 mg/g. The available phosphorus was lower (0.00115–0.0013%) than the total nitrogen content of karst soils. This supports the study of Chen *et al*. (2018) revealing through assays that P limitation rather than N limitation is more common in karst soils. The most abundant exchangeable ions in calcareous soils were calcium and magnesium. If these ions were reduced during soil erosion induced by anthropogenic disturbances, soil acidification would occur, followed by ecological deterioration (Wang and Dai 2012).

Sequencing Read Statistics and Protein Prediction

Pre-processing of raw metagenomic sequence data shows that about 19.6% of the filtered reads from all three sites were ADRs (Supplementary Table 1). Such abundances may be expected from shotgun metagenomic platforms such as Illumina, since this is a known limitation of the combined amplification and sequencing-by-synthesis design (Kircher *et al.* 2011). Thus, DRISEE (Keegan *et al.* 2016) is used to accurately estimate sequencing errors compared to score-based methods such as Phred. This step is crucial to ensure the reliability of any downstream analyses such as for annotation.

MG-RAST's implementation of protein identification used DIAMOND to perform similarity searches against the M5nr database, the latter of which is described as a non-redundant database combining SEED, COG, KO, NOG, *etc.* Approximately 33.5 million sequences containing possible protein features were detected across all sampling sites. This means that there are specific segments within the genetic sequences that are believed to

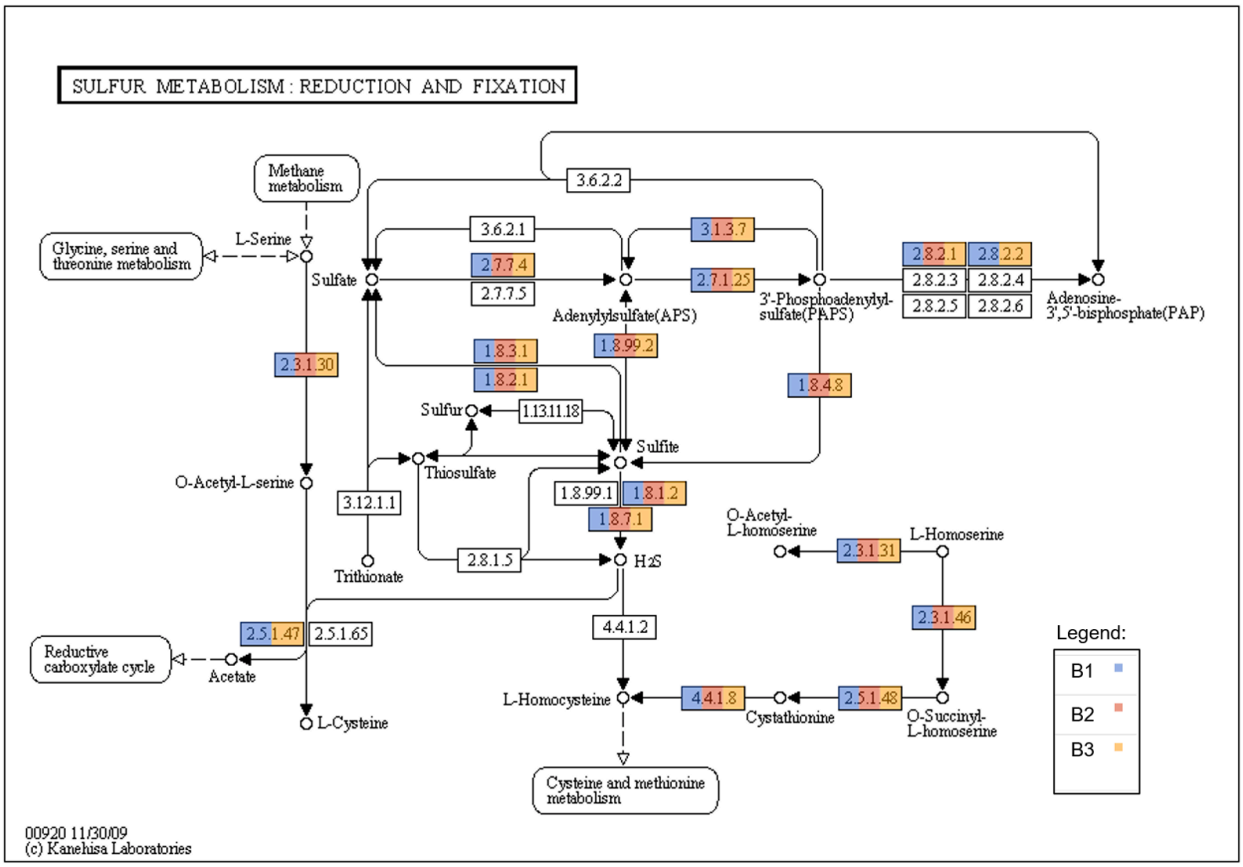


Figure 7. Sulfur metabolic pathway performed by KEGG mapper from MG-RAST. Colored EC numbers are the enzymes identified in the three karst areas.

contain information for producing certain characteristics or functional elements within proteins. These features could include binding sites, structural motifs, enzymatic domains, or other functional regions. Of the 33.5 million predicted protein features, an average of 14.7 million sequences (55.7%) were classified as proteins with unknown functions. Such values may be expected from genomic datasets since protein databases have been observed with a rise in cataloged structures with unknown functions. Currently, it is not possible to elucidate the biological function of such unknown proteins since these have not yet been characterized through assays. It is also possible that the uncharacterized protein features have no closely homologous features to proteins that have known structures or functions, as reported by Nadzirin and Firdaus-Raih (2012).

Metagenomic Analysis: Taxonomic Profiles and Functional Categories

The comparison of species richness between the three sampling sites was described by an alpha diversity index. These values are contextual, meaning that they are understood only relative to each other. In this case, B3 presented the highest abundance of species relative to B2 and B1. A rarefaction curve was also generated which supports this (Supplementary Figure S3).

Taxonomic profiles reveal the abundance of functional microbial communities linked to the karstification process. Most bacterial sequences in all three sites belonged to two dominant bacterial phyla – Actinobacteria and Proteobacteria. Alphaproteobacteria were the most abundant among the Proteobacteria (Yun *et al.* 2016). Proteobacteria, a group of microbes capable of responding to unstable carbons sources, was also found in other karst ecosystems. Such observation is similar to a study by Yun *et al.* (2016), which reported that at the phylum level, Actinobacteria and Proteobacteria were dominant in soil samples collected from weathered rock areas. Among the Alphaproteobacteria, various genera – including *Bradyrhizobium, Rhizobium*, and *Nitrobacter*, responsible for nitrogen cycling – were abundant across all three sampling sites (Supplementary Table S2). The relationships between legumes and *Rhizobium* and *Bradyrhizobium* bacteria are considered essential nitrogen-fixing symbiotic associations (Vance 2001).

On the other hand, *Nitrobacter*, the most studied nitrite oxidizer under α-proteobacteria, utilizes nitrite oxidoreductase [EC:1.7.2.1; EC:1.7.1.4] to catalyze nitrite oxidation. Analysis of KO reveals the presence of functional proteins and enzymes involved in nitrogen cycling. Different genera of interest under Gammaproteobacteria found to be abundant in the sampling sites include *Escherichia*, *Haemophilus*, *Salmonella*, and *Klebsiella* (Supplementary Table S2). These bacteria produce the enzyme carbonic anhydrase (CAs, EC 4.2.1.1), a metalloenzyme that catalyzes the hydration of carbon dioxide to bicarbonate and protons (Supuran 2011). In addition, the said enzyme participates in two chemical reactions, corrosion, and precipitation, which are the critical processes in karstification (Lian *et al.* 2011).

Santillán *et al*. (2021) reported that the bacteria belonging to the phyla Actinobacteria, Proteobacteria, and Acidobacteria were abundant in the tropical karst soils in the northern Yucatan Peninsula of Mexico. Actinomycetes are one of the most abundant organisms in the soil and can produce bioactive compounds that can inhibit specific pathogens. Acidobacteria were widely distributed in soils and recalcitrant to cultivation methods. Molecular studies indicated that the genome of Acidobacteria has genes involved in the carbon, nitrogen, and sulfur cycles, as well as those required for the degradation of complex polysaccharides (Kalam *et al.* 2020). Firmicutes were also abundant, possibly due to their resistance to nutrient stress and ability to survive in extreme habitats. The group of autotrophic green non-sulfur bacteria responsible for CO_2 fixation, the Chloroflexi, was also detected. Planctomycetes, a phylum of widely distributed bacteria found in terrestrial and aquatic habitats, was detected. This phylum plays an essential role in carbon and nitrogen cycles. Many species belonging to this phylum are capable of anaerobic ammonium oxidation or annamox. Some autotrophic bacteria belonging to the groups Nitrospirae, Chloroflexi, and Chlorobi might be involved in transforming the N element into the nitrogen cycle by the nitrogenase NifH and utilize inorganic compounds (*e.g.* ammonia) by nitrification.

Santillán *et al*. (2021) reported that the bacteria belonging to the phyla Actinobacteria, Proteobacteria, and Acidobacteria were abundant in the tropical karst soils in the northern Yucatan Peninsula of Mexico. Actinomycetes are one of the most abundant organisms in the soil and can produce bioactive compounds that inhibit specific pathogens. Acidobacteria were widely distributed in soils and recalcitrant to cultivation methods. Molecular studies indicated that the genome of Acidobacteria has genes involved in the carbon, nitrogen, and sulfur cycles, as well as those required for the degradation of complex polysaccharides (Kalam *et al.* 2020). Firmicutes were also abundant, possibly due to their resistance to nutrient stress and ability to survive in extreme habitats. The group of autotrophic green non-sulfur bacteria responsible for CO_2 fixation, the Chloroflexi, was also detected. Planctomycetes, a phylum of widely distributed bacteria found in terrestrial and aquatic habitats, was detected. This phylum plays an essential role in carbon and nitrogen cycles. Many species belonging to this phylum are capable of anaerobic ammonium oxidation or annamox. Some autotrophic bacteria belonging to the groups Nitrospirae, Chloroflexi, and Chlorobi might be involved in transforming the N element into the nitrogen cycle by the nitrogenase NifH gene and utilize inorganic compounds (*e.g.* ammonia) by nitrification.

The two most dominant genera under *Actinobacteria* were *Streptomyces* and *Mycobacterium* (Supplementary Table S2). Aside from being known as the most significant source of biologically active compounds, *Streptomyces* are also ammonifying bacteria thriving in soils and aid in the decomposition of organic materials. The identified *Streptomyces* can induce calcification by undergoing at least a single biomineralization pathway (Maciejewska *et al*. 2017). Moreover, *Mycobacterium* contains the key enzymes (citrate lyase [EC:4.1.3.6] and 2-oxoglutarate synthase [EC:1.2.7.3]) for the complete reductive TCA cycle (Srinivasan and Morowitz 2006). In addition, the large and small subunits of glutamate synthase – which play an essential role in ammonia assimilation pathways – are primarily conserved throughout mycobacterial genomes (Amon *et al.* 2010).

The most dominant fungal phylum is Ascomycota, followed by an abundance of Basidiomycota. Fan *et al.* (2019) and Santillán *et al.* (2021) reported similar results from previous studies on karst landscapes. In karst ecosystems, fungi play an essential role in many critical environmental processes such as organic matter decomposition, protection against leaching by element storage in biomass, and element released by mineralization (Hagn *et al.* 2003). Lian *et al.* (2010) described the fungal mycelia's role in the acceleration of carbonate rock breakdown. Fungal mycelia were inserted into crevices and absorbed moisture from microcracks in the rock, which led to the complete dissolution of acid secretions of bacteria within the rock. The $Ca(NO_3)_2$ serves as a nitrogen source for some fungal strains. These strains reduce NO_3 to NH^{4+} by nitrate reductase [EC:1.7.1.1; EC:1.7.99.4] and nitrite reductase [EC:1.7.2.1; EC:1.7.1.4], which then lead to an increase in the pH. The CO_2 or HCO_3^- released by fungal respiration reacts with the Ca^{2+} ions present in the medium, to produce calcium carbonate [Hou *et al.* (2011), as cited by Lian *et al.* (2011)].

CONCLUSION

The karst forest in Basey, Samar supports a diverse ecosystem, as evidenced by the presence of various bacterial and fungal phyla, which perform specialized functions in nutrient cycling. This pioneering study on the taxonomic and functional diversity of microbial communities was made possible using Illumina sequencing of the karst soil metagenome. The results highlight the significant link between the taxa present and functional potential. For example, the abundance of *Streptomyces* in the karst soils can be attributed to their capability of inducing calcification through at least one biomineralization pathway, as suggested by the presence of enzymes revealed through sequence comparison with the KO database. Furthermore, analysis of the metagenome reveals potential metabolic pathways for the metabolism and cycling of carbon, nitrogen, and sulfur.

Future research may focus on elucidating the interaction of these microorganisms with other organisms such as plants in the karst area. Studies on how they could affect the survival of other species in the ecosystem may be done. As the current study provides a baseline for the karst metagenome, comparison with other ecosystems may be worth exploring.

ACKNOWLEDGMENTS

We would like to thank the DOST-SEI (Department of Science and Technology–Science Education Institute) for allowing the third author to take part in this research program through the CIP (Career Incentive Program).

This work is supported by the DOST-PCAARRD (Department of Science and Technology–Philippine Council for Agriculture, Aquatic, and Natural Resources Research and Development) (Fund Code: N9A652A). It is part of the project titled *"*Project 3: Taxonomic and Functional Diversity Profiling of Soil Rhizosphere Microbial Communities in Forest over Limestone Ecosystem*"* under the program "Assessment and conservation of forest over limestone ecosystem biodiversity in selected municipalities of Samar, Philippines (CONserve-KAIGANGAN)."

This study is dedicated to the late Dr. Marilen P. Balolong, who spearheaded the establishment of the CON-KAIGANGAN Project 3.

STATEMENT ON CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

NOTES ON APPENDICES

The complete appendices section of the study is accessible at the following link: https://drive.google.com/drive/folders/1OegABLybvm5VN9swY74e_FSHdIHCXIcO

REFERENCES

AMON J *et al.* 2010. Common patterns – unique features: nitrogen metabolism and regulation in Gram-positive bacteria. FEMS Microbiology Reviews 34(4): 588–605. https://doi.org/10.1111/j.1574-6976.2010.00216.x

BARDGETT RD, VAN DER PUTTEN WH. 2014. Below ground biodiversity and ecosystem functioning. Nature 515: 505–511.

CHAUDHARY H, GOPALAN N, SHRIVASTAVA A, SINGH S, SINGH A, YADAV, J. 2013. Antibacterial activity of actinomycetes isolated from different soil samples of Sheopur (a city of central India). J Adv Pharm Technol Res 4(2): 118–123.

CHAVAN DILIP V, MULAJE SS, MOHALKAR RY. 2013. A review on actinomycetes and their biotechnological application. Int J Pharm Sci Res 4: 1730–1742.

CHEN XB, SU YR, HE XY, WEI YW, WEI WX, WU JS. 2012. Soil bacterial community composition diversity responds to cultivation in Karst ecosystems. World J Microbiol Biotechnol 28: 205–213.

CHEN H, LI D, XIAO K, WANG K. 2018. Soil microbial processes and resource limitation in karst and non-karst forests. Funct Ecol 32: 1400–1409.

CHENG X, YUN Y, WANG H, MA L, TIAN W, MAN B, LIU C. 2021. Contrasting bacterial communities and their assembly processes in karst soils under different land use. Sci Total Environ 751: 142263

DORAN JW, ZEISS MR. 2000. Soil health and sustainability: managing the biotic component of soil quality. Appl Soil Ecol 15: 3.

DREW D. 2017. Karst hydrogeology and human activities: impacts, consequences, and implications: IAH International Contributions to Hydrogeology 20. Routledge. UK: CRC Press, Taylor and Francis Group.

FAN Z, LU S, LIU S. 2019. Changes in plant rhizosphere microbial communities under different vegetation restoration patterns in karst and non-karst ecosystems. SciRep 9: 8761.

FLEURY S. 2009. Karst processes, landforms, and issues. In: Land use policy and practice on karst terrains. Dodrercht: Springer. doi.org/10.1007/978-1-4020-9670-9_1

GOOGLE. n/d. Google Earth View of Three Locations in Basey, Samar, Philippines. Retrieved on 10 Jun 2021 from https://earth.google.com/

HAGN A, PRITSCH K, SCHLOTER M, MUNCH JC. 2003. Fungal diversity in agricultural soil under different farming management systems, with special reference to biocontrol strains of *Trichoderma* spp. Biol Fertil Soils 38: 236–244.

HE XY, WANG KL, ZHANG W, CHEN ZH, ZHU YG, CHEN HS. 2008. Positive correlation between soil bacterial metabolic and plant species diversity and bacterial and fungal diversity in a vegetation succession on karst. Plant Soil 307: 123–134.

KALAM S, BASU A, AHMAD I, SAYYED R Z, EL ENSHASY HA, DAILIN DJ, SURIANI N. 2020. Recent understanding of soil Acidobacteria and their ecological significance: a critical review. Front Microbiol 11: 2712.

KEEGAN KP, GLASS EM, MEYER F. 2016. MG-RAST, a metagenomics service for analysis of microbial community structure and function. Methods Mol Biol 1399: 207–233.

KIRCHER M, HEYN P, KELSO J. 2011. Addressing challenges in the production and analysis of Illumina sequencing data. BMC Genom 12(1).

LI F, ZHANG X, GONG J, LIU L, YI Y. 2018. Specialized core bacteria associated with plants adapted to adverse environments with high calcium contents. PLoS ONE 13: 3.

LI Q, YANG H, SONG A. 2021. Comparative analysis of soil bacteria assemblages across land-use types in a given karst landscape in southwest China. Pol J Environ Stud 30(4): 3699–3711.

LIU Y, QIN Y, CHEN T, LU M, QIAN X, GUO X, BAI Y. 2020. A practical guide to amplicon and metagenomic analysis of microbiome data. Protein Cell 12(5): 315–330.

LIU J, HE X, SUN J, MA Y. 2021. A Degeneration Gradient of Poplar Trees Contributes to the Taxonomic, Functional, and Resistome Diversity of Bacterial Communities in Rhizosphere Soils. Int J Mol Sci 22(7).

LIAN B, CHEN Y, TANG Y. 2010. Microbes on carbonate rocks and pedogenesis in karst regions. J Earth Sci 21: 293–296.

LIAN B, YUAN DX, LIU ZH. 2011. Effect of microbes on karstification in karst ecosystems. Sci Bull 56(35): 3743–3747.

MACIEJEWSKA M, ADAM D, NAOME A, MARTINET L, TENCONI E, CALUSINSKA M, DELFOSSE P, HANIKENNE M, BAURAIN D, COMPERE P, CARNOL M, BARTON HA, RIGALI S. 2017. Assessment of the potential role of streptomyces in cave moonmilk formation. Front Microbiol 8: 1181.

MCCAULEY A, JONES C, JACOBSEN J. 2009. Soil pH and organic matter. Nutrient Management Module 8(2): 1–12.

NADZIRIN N, FIRDAUS-RAIH M. 2012. Proteins of unknown function in the protein data bank (pdb): an inventory of true uncharacterized proteins and computational tools for their analysis. Int J Mol Sci 13(10): 12761–12772.

ONG JDP, PATERNO ES, LANTICAN NB, DIAZ MGQ. CRUZ WT. 2018. Characterization of plant growth-promoting diazotrophs from salt-affected areas in the Philippines. Philipp J Crop Sci 43: 56–68.

PAN F, LIANG Y, ZHANG W, ZHAO J, WANG K. 2016. Enhanced nitrogen availability in karst ecosystems by oxalic acid release in the rhizosphere. Front Plant Sci 7: 687.

PENG J, XU YQ, ZHANG R, XIONG K N. LAN AJ. 2013. Soil erosion monitoring and its implication in a limestone land suffering from a rocky destination in the Huajiang Canyon, Guizhou, Southwest China. Environ Earth Sci 69: 831–841.

R CORE TEAM. 2021. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.rstudio.com/

SANTILLÁN J, LÓPEZ-MARTÍNEZ R, AGUILAR-RANGEL E, HERNÁNDEZ-GARCÍA K, CRAM S, ALCÁNTARA-HERNÁNDEZ R, VASQUEZ MURRIETA M. 2021. Microbial diversity and physicochemical characteristics of tropical karst soils in the northeastern Yucatan peninsula, Mexico. Appl Soil Ecol 165.

SINGH S, KUMAR P, GOPALAN N, SHRIVASTAVA B, KUHAD RC, CHAUDHARY HS. 2012. Isolation and partial characterization of actinomycetes with antimicrobial activity against multidrug resistant bacteria. Asian Pac J Trop Biomed 2: S1147–S1150.

SRINIVASAN V, MOROWITZ H J. 2006. Ancient genes in contemporary persistent microbial pathogens. Biol Bull 210: 1–9.

SUPURAN C. 2011. Bacterial carbonic anhydrases as drug targets: toward novel antibiotics? Front Pharmacol 2: 34.

TANG YQ. 2013. Fractal characteristics and stability of soil aggregates in karst rocky desertification areas. Nat Hazards 65: 563–579.

TOLENTINO PJS, NAVIDAD JRL, DELOS ANGELES M, FERNANDEZ DAP, VILLANUEVA ELC. OBEÑA RDR, BUOT IE. 2020. Biodiversity of forests

over limestone in Southeast Asia with emphasis on the Philippines. Biodiversitas 21(4): 1597–1613.

[UNDP] United Nations Development Programme. 2012. Samar Island Biodiversity Project [Terminal Report] Retrieved from https://faspselib.denr.gov.ph/sites/default/files//Publication Files/Samar Island Biodiversity Project (SIBP) Project Completion Report_June 2012.PDF

VANCE C. 2001. Symbiotic Nitrogen Fixation and Phosphorus Acquisition: Plant Nutrition in a World of Declining Renewable Resources. Plant Physiology 127(2): 390–397. https://doi.org/10.1104/pp.010331

WANG W-J, DAI W-H. 2012. Study on soil pH and acidic buffering properties in Anhui Province. Chinese Agricultural Science Bulletin 28(15): 67–72.

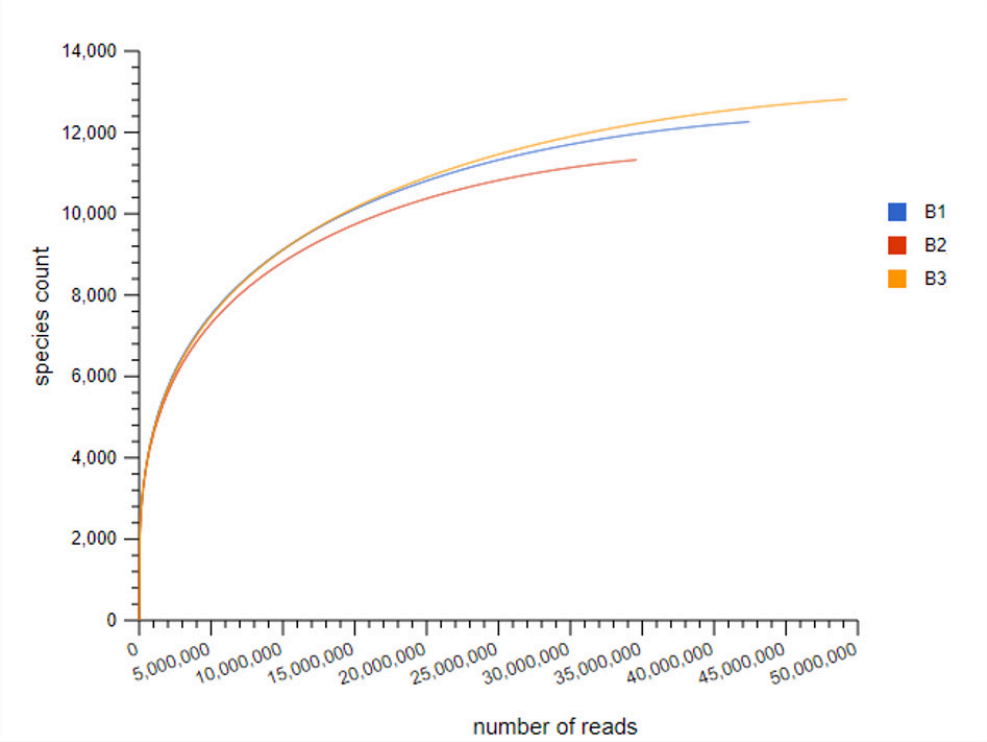
WILKE A, HARRISON T, WILKENING J, FIELD D, GLASS EM, KYRPIDES N, MAVROMMATIS K, MEYER F. 2012. The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. BMC Bioinform 13(1): 141.

YUN Y, WANG H, MAN B, XIANG X, ZHOU J, QIU X, DUAN Y, ENGEL AS. 2016. The relationship between pH and bacterial communities in a single karst ecosystem and its implication for soil acidification. Front Microbiol 7: 1955.

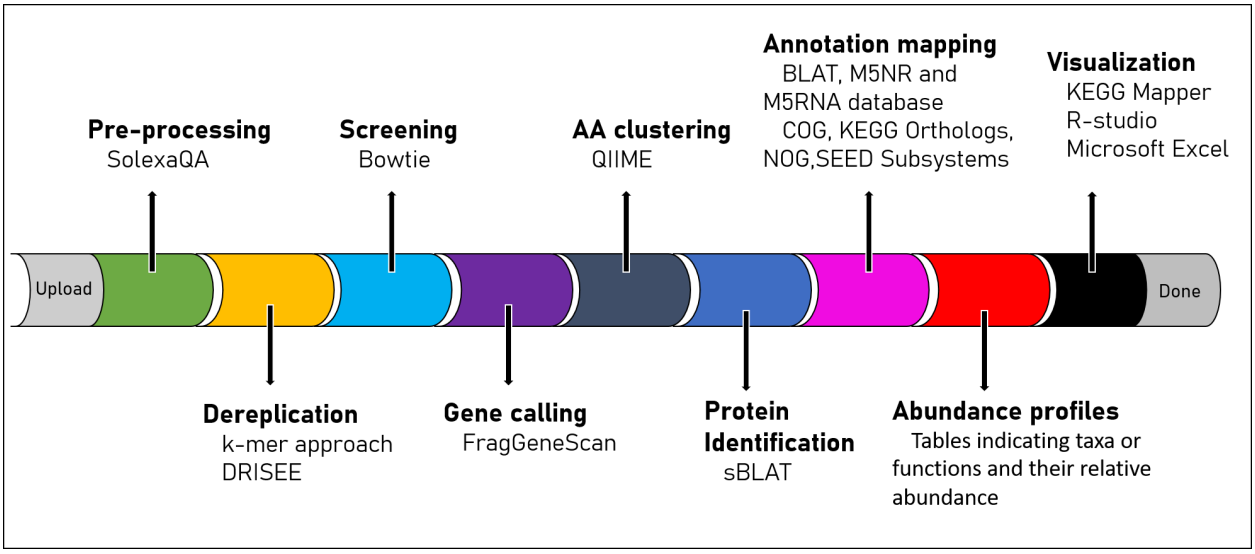
ZHAO C, LONG J, LIAO H. 2019. Dynamics of soil microbial communities following vegetation succession in a karst mountain ecosystem, Southwest China. Sci Rep 9: 2160.

SUPPLEMENTAL FIGURES

Supplementary Figure S1. Collection of soil samples from karst forest areas conducted in_Basey, Samar, Philippines



Supplementary Figure S3. Rarefaction curves of B1, B2, and B3, showing the species_richness in the soil from the karst forest areas in Basey, Samar, Philippines



Supplementary Figure S2. Metagenomics pipeline used in analyzing the taxonomic diversityand metabolic functions of bacterial and fungal communities present in forest over limestoneecosystems