Environmental and Experimental Biology (2022) 20: 205–218
http://doi.org/10.22364/eeb.20.19

Original Paper

Insights into the core bacterial consortia of root endophytes in two cultivated varieties of rice in West Bengal

ISSN 2255-9582

Subhadipa Sengupta[1]*, Pankaj K. Singh[2], Sayak Ganguli[3]

[1]Bidhannagar College, Post Graduate Department of Botany, EB-2, Sector-1, Salt Lake, Kolkata 700064, India
[2]Hawkesbury Institute for the Environment, Western Sydney University, Penrith NSW-2753, Australia
[3]St. Xavier's College (Autonomous), Department of Biotechnology, Kolkata 700016, India

*Corresponding author, E-mail: bansubha@gmail.com

Abstract

Root endophytes are considered to be one of the potent environment-friendly substitutes for chemical fertilizers, as they possess an ability to induce crosstalk inside the hosts for growth promotion, nitrogen fixation, phosphate solubilization and iron sequestration. This study aimed to explore and evaluate the key root endophytic bacterial consortia of two widely cultivated varieties of rice (*Oryza sativa* L.), cv. 'Saraswati' (OS01) and cv. 'Kunti' (OS04). Detailed comparative metagenome data were generated for endophytes of OS01 and OS04 and the species richness was calculated. OS01 showed higher endophyte species richness than OS04, with alpha diversity values of 3.10 and 2.40, respectively. *Bacillus, Magnetospirillum, Methanocystis, Desulfomicrobium* and *Pantoea* were identified as common endophyte members for both cultivars. *Solibacillus, Paenibacillus, Candidatus*, and *Melospira* were unique members of OS01, and *Herbaspirillum, Pandoraea, Anabaenopsis* for OS04. Considerable occurrence of nitrogen fixing bacteria and methanogenic bacteria in the cultivars confirmed biological nitrogen fixation, which can contribute to plant development. Core homeotic pathways of amino acid biosynthesis and carbon metabolism were also reflected in endophytes from both cultivars, indicating a supportive environment for microorganisms. Sulfur metabolism pathways were likewise predicted to be active in the niche under study, which may be attributed as a response to arsenic stress. Furthermore, the most abundant genera identified may potentially serve as crucial consortium candidates for host plant development and contribute to better yield in a sustainable manner.

Key words: bacterial diversity, metagenome, microbial consortia, plant growth, rice root endophytes.
Abbreviations: OTUs, operational taxonomic units.

Introduction

Rice (*Oryza sativa* L.) is the most widely grown staple food for over half the world's population. Fifty percent of the Asian population depends on rice for their dietary calorie supply. Therefore, rice production is critical for global food security (Muthayya et al. 2014). West Bengal is the largest source of rice production across India. However, the production faces significant threat from a large number of pathogenic organisms, nematodes, fungi, insects and virus. It was projected that up to 40% of the annual rice production is lost due to the rice-specific infectious diseases (Oerke, Dehne 2004). Further, high temperature, drought, salinity, submergence, nutrient deficiencies etc. also adversely affect rice productivity (Wani, Sah 2014). The availability of chemical fertilizers also limits rice production on a global scale. To achieve the targeted requirement of 321 million tons of rice, it is predicted that 28.8 million tons of chemical fertilizers will be required globally per year (Mahdi et al. 2010). However, the global chemical fertilizer production capacity is 21.6 million tons per year. Thus, a shortage of 7.2 million tons of chemical fertilizers have to be met either by altering cropping cycles or by adopting alternative agricultural strategies. In addition, inappropriate fertilization patterns and excessive use of nitrogen fertilizers have resulted in substantial nitrogen loss through ammonia leaching (Cameron et al. 2013; Ma et al. 2019). Hence, in this scenario, use of biofertilizers seems to be more appropriate approach to not only to counter the need for the scarcity of chemical fertilizer but also to maintain soil health and fertility in a sustainable way.

Endophytes have been found in almost all plant tissues studied (Gaiero et al. 2013; Hardoim et al. 2015; Xia et al 2022). They perform a key function by accelerating mineral uptake by crops and enhancing the bioavailability of various minerals (Sturz et al. 2000; Sessitsch et al. 2004). Endophytes also produce phytohormones, siderophores, antifungal or antibacterial agents that directly or indirectly improve plant development, seed germination, drought tolerance etc. and

protect plants against fungal and bacterial pathogens (Nair, Padmavathy 2014). They play a role in lowering biotic stress, and improving resistance to nematodes, insects, and diseases in host plants (Azevedo et al. 2000; Sturzet al. 2000). Endophytes can be a valuable source of secondary metabolites (Strobel 2002). This may be due to their niche similar to that of phytopathogens. They act as a counter defense system against disease development through de novo synthesis of novel compounds which inhibit the pathogenesis cascade at the very onset. Apart from their role in biocontrol, seedling emergence and sustainable plant growth has been reported to be accelerated by rice seed endophytes (Jana et al. 2022). It has been reported that the presence of *Pseudomonas* and *Bacillus* results in significant improvement of morphological characters and induced systemic resistance in plants (Elekhtyar 2015; Hu Zhou et al. 2021). Endophytes have also been acknowledged to play important roles in increasing plant cell wall strength leading to positive alterations of host physiology (Torres et al. 2012). In addition, overall metabolic responses were found to be increased, which help in enhancing the synthesis of various metabolites (phenolic compounds, oxidative stress enzymes chitinases, peroxidases, polyphenol oxidase, phenylalanine ammonia lyase, etc.) resulting in improved defense response (Xia et al 2022). The synergistic role of 17 strains identified from rice seedling roots was shown, of which five strains (*Rhizobium larrymoorei* E2, *Bacillus aryabhattai* E7, *Bacillus aryabhattai* MN1, *Pseudomonas granadensis* T6, and *Bacillus fortis* T9) showed both nitrogen fixation potential and indole-3-acetic acid production (Shen et al. 2019). Other strains have shown ability to tolerate high doses of fungicides and pesticides, indicating the role of endophytes in chemical sequestration (Sahoo et al. 2017; Fadiji, Babalola 2020). Moreover, endophytes support other crops that are grown through crop rotation with their host plants (Sturz et al. 2000). Therefore, the contribution of endophytes as biological fertilizers is very important (Khare et al. 2018).

There are reports on endophytes showing effect of biofertilizer applications. For example, *Azospirillum* inoculation on wheat, maize, sorghum, and millet significantly increased yield by up to 30% (Okon, Labandera-Gonzalez 1994). Similarly, *Trichoderma* used as a coating agent for rice seeds increased yield by 15 to 20% compared to rice plants receiving full inorganic fertilizer only (Cuevas 1991). Also, a combination of cyanobacteria, microalgae and *Azotobacter* applied on rice significantly increased the rate of germination and growth (Zayadan et al. 2014).

All these studies further emphasize the real need for proper characterization of rice root endophytes, which promises to uncover several beneficial functions each adding to our knowledge required for effective design of biofertilizer consortia. This presents an alternative approach in the direction of an environment-friendly potential natural source for biological control in disease management (Turner et al. 2013). Microbial communities are highly affected by soil type, cultivation practices and geographical location (Trivedi et al. 2016; Correa-Galeote et al. 2018). Climate, pH and the composition of soil and fertilizers are particularly important factors on the local scale (Sun et al. 2013). For example, a study of rice from an area of high salinity (Sundarban, West Bengal, India) reported some new genera with low diversity and the assemblage was dependent on plant genotype as well as on the environment (Kunda et al. 2018).

The advancement of cultivation-independent, high-throughput sequencing based metagenomic analysis allows to gain insight into the bacterial community diversity in plant tissues. In an earlier report, an overview of metagenome data of two rice root samples that were collected from a high productive zone of West Bengal was provided (Sengupta et al. 2017). The present study extends the computational analysis for identification and comparison of enriched biochemical pathways. These results could serve as a source towards the development of suitable bacterial consortia in the form of future bio-fertilizers for rice cultivation, which could be recommended to the rice farmers of West Bengal.

Materials and methods

Sample collection, isolation and sequencing

Rice root samples were collected from 60-day-old rice plants along with bulk rhizospheric soil from cultivated land of South 24 Pargarnas in West Bengal. At these collection sites, rice monoculture cultivation had been practiced for more than 15 years. Two high yield rice cultivars, 'Saraswati' (IET-11271, rainfed semideep long grain) and 'Kunti' (IET-6141, medium land, dwarf grain) were used for the study. These cultivars were designated as OS01 and OS04, respectively. Root samples were collected in triplicate and stored in sterile plastic bags. Initially the excised roots were washed in tap water, and 70% ethanol was used for surface sterilization for 1 min. After that, 1.2% (w/v) NaOCl solution was used to sterilize the tissue for 15 min. Samples were then washed three times with sterile distilled water with shaking (10 min). Root samples were finally dried and stored at –20 °C (Schulz et al. 1993).

The DNA of each sample was isolated according to the protocol reported by Garcias-Bonet et al. (2012). A Qubits DNA HS assay kit (Life Tech) was used for quantification and 1 μL of each sample was used for determining concentration using a Qubit®2.0 Fluorometer. The amplicon libraries were prepared using Nextera XT index Kit (Illumina) according to the 16S metagenomic sequencing library preparation protocol (Part # 15044223 Rev. B). Primers for the amplification of the V3-V4 hyper-variable region (V3 Forward Oligo: CCTACGGGNBGCASCAG and V4 Reverse Oligo: GACTACNVGGGTATCTAATCC) of 16S rDNA gene of bacteria and Archaea were used.

The amplicons with the Illumina adaptors were amplified using i5 and i7 primers, which added multiplexing index sequences as well as common adapters required for cluster generation (P5 and P7), according to the standard Illumina protocol. The library sizes of samples OS01 and OS04 were 634 bp and 622 bp, respectively. The libraries were sequenced using the Illumina sequencing chemistry to generate ~150 Mb of data per sample. After obtaining the qubit concentration for the library and the mean peak size from the Bioanalyzer profile, the library was loaded onto the Illumina platform at appropriate concentration (10 to 20 pM) for cluster generation and sequencing. The copied reverse strand was then used to achieve sequencing from the fragments at the opposite end (Mbengue et al. 2016).

Raw data quality check

The quality check of sequenced raw reads is one of the most important steps in the pre-processing of sequenced data, where the objective is to understand some relevant properties. The quality of raw reads of Illumina sequencing was checked for ambiguous bases, Phred score > Q30, read length, nucleotide base content and other parameters by using FASTQC.

QIIME analysis

The processed reads were clustered into operational taxonomic units (OTUs) by using QIIME software to identify the microbial community. These OTUs were further used for taxonomic assignment, Phylogenetic Diversity analysis and abundance estimation. QIIME (Quantitative Insights Into Microbial Ecology) is a bioinformatics pipeline designed for analyzing microbial communities. The QIIME software clusters the marker gene nucleotide sequences into OTUs and taxonomically annotates the OTUs by looking for sequences similar to them in a reference taxonomic database.

MEGAN analysis

Megan analysis was started by comparing the reads with the NCBI nr database, using RAPSearch (E value ≤ 0.001, percent identity ≥ 30). Then, MEGAN assigned a taxon ID to processed read results based on NCBI taxonomy, which created a MEGAN file that contained the required information for statistical and graphical analysis. Lastly, the lowest common ancestor algorithm was run to inspect assignments, to analyze data and to create summaries of data based on different NCBI taxonomy levels.

Kraken was additionally used for assigning taxonomic labels to metagenomic DNA sequences. Kraken's classification algorithm performs mapping of k-mers to taxa by querying a pre-computed database. Kraken creates this database through a multistep process, beginning with the selection of a library of genomic sequences (Wood et al. 2014). The Kraken library is based on completed microbial genomes in the National Center for Biotechnology Information's (NCBI) RefSeq database. The classification results were visualized using Krona.

PICRUSt workflow was used to predict the gene content of the metagenomes. PICRUSt enables the estimation of the gene families contributed to a metagenome by bacteria or archaea identified using 16S rRNA sequencing. These gene content predictions are pre-calculated for protein-coding genes present in KEGG and for 16S rRNA gene copy number. The QIIME OTU table was used to predict functions and pathways using PICRUSt. The first step was to correct the OTU table based on the predicted 16S rRNA copy number for each organism in the OTU table. The functional predictions of KEGG Ortholog (KOs) were carried out using the corrected OTU table as input. Finally, the KO's were collapsed to KEGG pathways because one KO can map to many KEGG Pathways. Finally, the pathways were mapped for their enrichment using KEGG Mapper [https://www.genome.jp/kegg/mapper.html].

Soil analysis

The soil samples were analyzed for physical and chemical properties (pH, texture, total organic carbon and total nitrogen). Soil pH was measured in 1:2.5 soil:water suspension using a glass electrode pH meter (Jackson 1962) and soil texture was determined by hydrometer method. Total organic carbon of soil and nitrogen were detected by the Walkley-Black method (Jackson 1967) and Kjeldahl method (Jackson 1973), respectively.

Data visualization

Comparative analysis was performed using Venny 2.0 to identify the unique and the common microbes from the two datasets and to generate representative Venn diagrams. Also, a heat map was generated using an in-house R program that enabled the visualization of the abundance of the common microbes in graphical form.

Results

Analysis of species richness and operational taxonomic unit generation

QIIME analysis indicated species richness to be much higher in sample OS01 than in OS04, shown by rarefaction curves (Fig. 1) and Shannon alpha diversity – 3.10 and 2.40, respectively. The filtered reads were grouped into OTU clusters. Total number of identified OTU's was 420 in OS01 and 297 in OS04.

Comparative endophytic profile

The four major classes of endophytes were identified, with the highest percentage of sequences assigned to Fermicutes (almost 90% in OS04 and 50% in OS01) followed by Cyanobacteria, Proteobacteria and Actinobacteria (Fig. 2a). The comparative analysis of the community profile revealed absence of Crenarchaeota, Epsilonbacteraota,

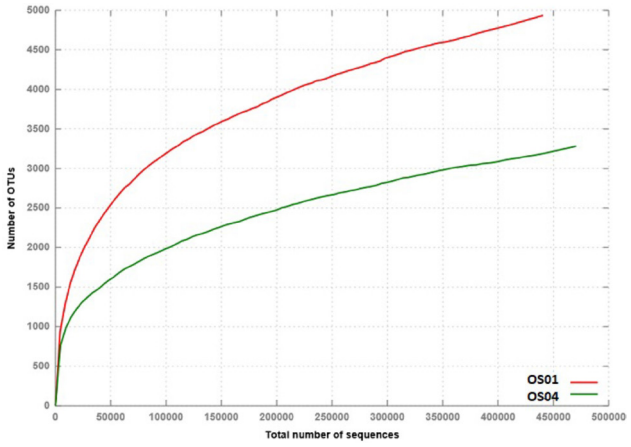

Fig. 1. Rarefaction curve indicating endophyte species richness of OS01 and OS04. Rarefaction curve for sample OS01 and OS04 was generated using FASTQC and validated by SILVAngs pipeline. The alpha diversity of OS01 was found to be higher than OS04 (3.10 and 2.40, respectively). The X axis represents total number of sequences whereas Y axis represents the number of OTUs obtained.

Latescibacteria and Planctomycetes in OS04; while OS01 lacked members of Calditrichaeota, Hydrogenedentes, and Nitrospinae. However, members of Actinobacteria were found in both samples. A Venn diagram was generated (Figure 2b) to show the most abundant, shared and unique OTUs. A total of 115 common phyla were identified among the two samples. *Magnetospirillum, Methanocystis, Desulfomicrobium*, and *Pantoea* were the most common genera. Unique members of OS01 were *Candidatus, Clavaria, Melospira*, and *Cellulomonas*, while *Phaeospirillum, Anaerococcus, Pandoraea* and *Anabaenopsis* were found only in OS04. A detailed list of genera for OS01 and OS04 is given in Appendix 1. Archeal sequences were detected in small amounts (less than 100 sequences), which belonged to two classes – Crenarchaeota and Euryarchaeota. Crenarchaeota was absent in OS04 along with Epsilonbacteraota, Latescibacteria and Planctomycetes. Krona interactive charts were generated to validate the data obtained in a heat map (Fig. 3).

Pathway analysis

Amino acid biosynthetic pathways and carbon metabolism pathways were found to be the most abundant primary metabolic pathways enriched in both samples (Fig. 4). Secondly, sulphur and nitrogen metabolism pathways were also detected in both samples. Pathways found to be enriched in OS01 samples belonged mostly to primary metabolism pathways, whereas for OS04, apart from

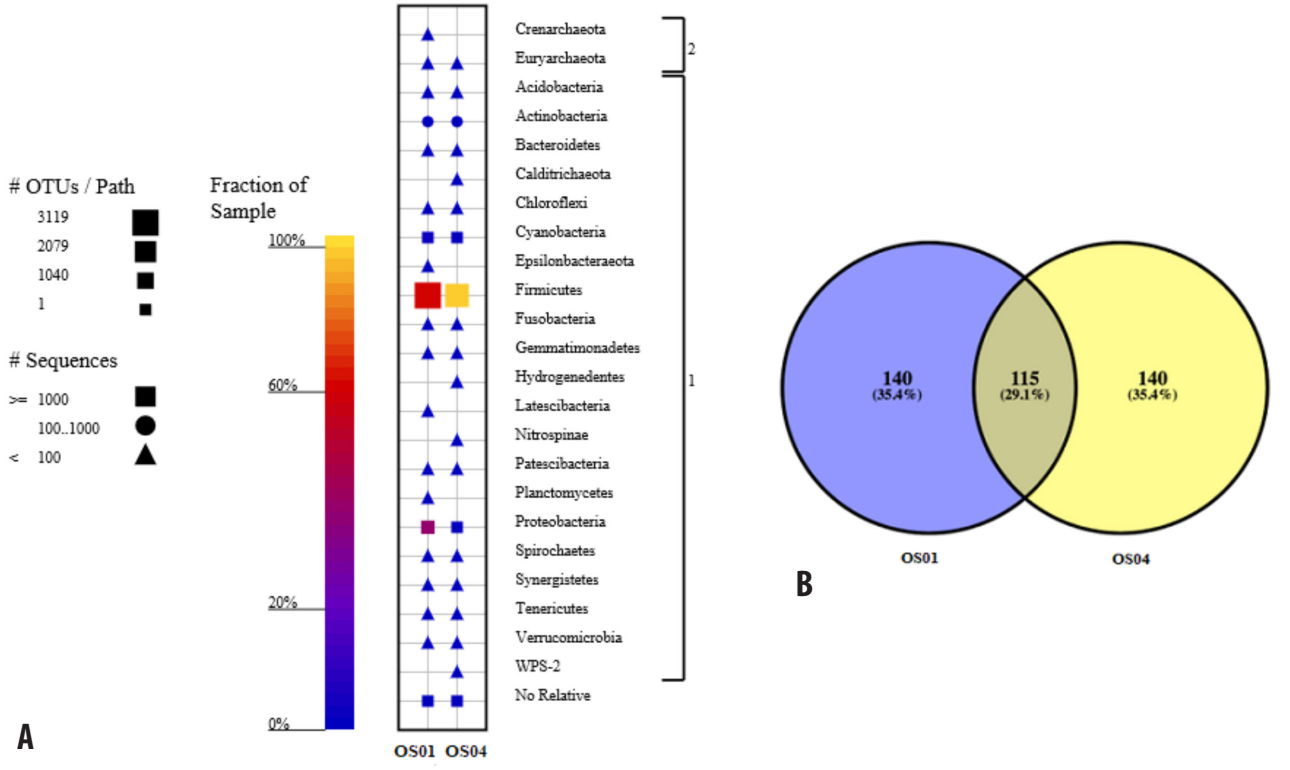

Fig. 2. Comparative analysis between the sample sets: heat map (A) and Venn diagram (B) depicting the distribution of microbial abundance in OS01 and OS04. Heat map depicted the phylum level abundances of root endophytes. The sample size is represented in form of squares where larger squares indicated higher abundances. In both the samples Firmicutes are represented as the highest abundant phylum followed closely by proteobacteria in sample OS01 and OS04. Cyanobacteria members are also found to be equally abundant in OS04. Venn diagram is showing a total of 115 common genera identified between the two samples. *Bacillus, Magnetospirillum, Methanocystis, Desulfomicrobium, Pantoea* are amongst the common genera. Unique members of OS01 are *Solibacillus, Paenibacillus, Clavaria, Melospira,* and *Cellulomonas*, while for OS04 are *Herbaspirillum, Anaerococcus, Pandoraea* and *Anabaenopsis*.

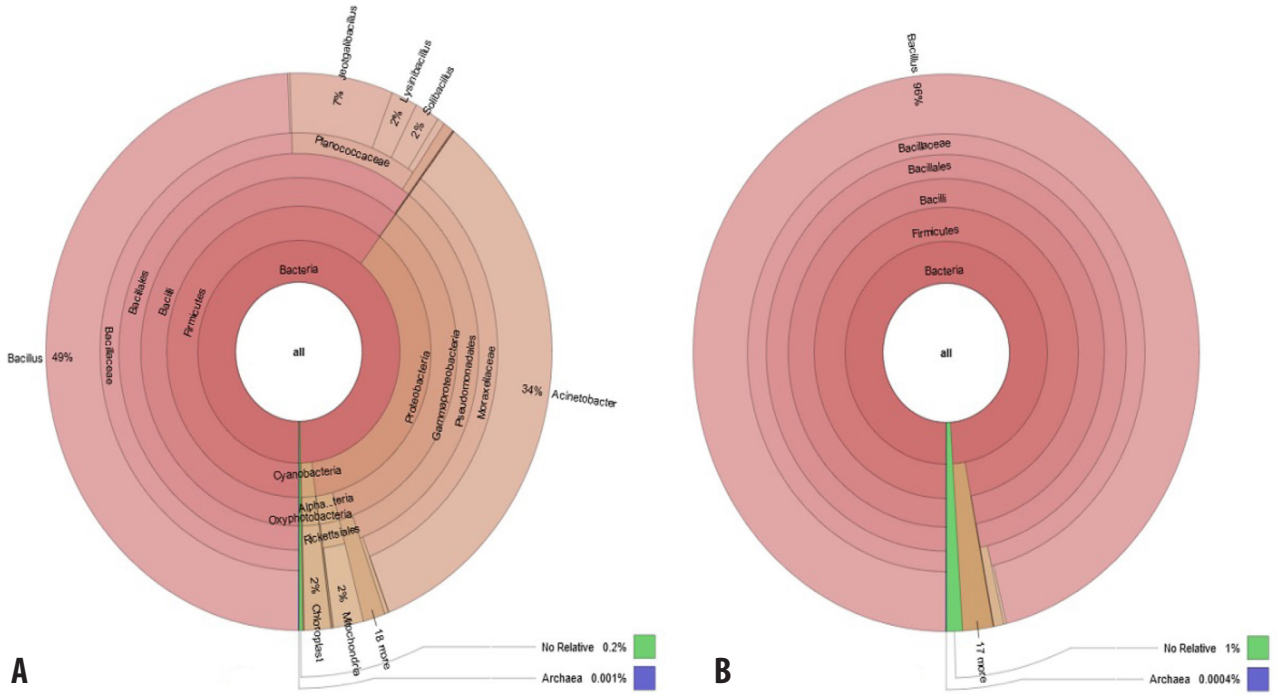

Fig. 3. Phylogenetic analysis of the obtained OTUs representing abundance and hierarchy simultaneously using KRONA graph. A, OS01 sample exhibits the presence of different bacterial members, where *Bacillus* is of maximum abundance. B, OS04 exhibits the overwhelming presence of 96% *Bacillus*.

primary metabolic pathways, streptomycin biosynthetic pathways were also detected. Detailed pathway data containing the KEGG Orthology accessions of OS01 and OS04 are available in Appendix 2 and 3, respectively.

Soil analysis

Values of soil parameters at the physicochemical level were similar between sampling sites, but sampling site OS04 had a higher organic carbon level and total nitrogen level, compared to sampling site OS01. OS04 also had a slightly bit higher sand content compared to OS01 (Table 1). However, the pH was same for both samples.

Discussion

In present study on rice endophytes, we were able to identify four major phyla among the domain bacteria: Firmicutes, Proteobacteria, Cyanobacteria and Actinobacteria. Firmicutes were present in the highest abundance in both samples. The relative frequency of Firmicutes was above 90% in sample OS04 and nearly 50% in sample OS01. Bacillus is the major genus and a ubiquitous member responsible for plant growth promotion (Vacheron et al. 2013). These species are able to dissolve insoluble forms of phosphate, thus playing one of the key functions in plant nutrition (Alori et al. 2017). In addition, *Bacillus* produces siderophores, indole-3-acetic acid, and 1-aminocyclopropane-1-carboxylic acid deaminase, which help plants to regulate abiotic stress. They are also involved in mitigating biotic stress by activating the plant defense system (Radhakrishnan, 2017).

The second most abundant phylum was Proteobacteria and the most represented class was Gammaproteobacteria, a large heterogenous group of organisms and a major colonizer in rice roots (Mano, Morisaki 2008). Presence of Gammaproteobacteria indicates some protection from pathogen attack, as reported for other crop plants (Köberl et al. 2017). Cyanobacteria, which was the third most abundant group among the endophytes, are important in nitrogen fixation.

Two Archeabacterial classes, Crenarchaeota and Euryarchaeota, were detected in small amounts (less than 100 sequences per sample) in the samples. Crenarchaeota, along with Epsilonbacteraota, Latescibacteria and Planctomycetes, were absent in OS04. The absence of these members in OS04 may be associated with the rhizospheric conditions and interactions with other endophytic microbes. Species of Crenarchaeota, which were uniquely found in OS01, but were surprisingly absent in OS04, and considered to be the most abundant ammonia-oxidizing organisms in soil (Daebeler et al. 2012). Members of Epsilonbacteraota in OS01 are mostly associated with the assimilation of nitrogen from ammonia taken up from the environment or generated from environmental nitrate or nitrite, by employing a variety of functional redox modules (Koch et al. 2019).

The presence of Latescibacteria in the OS01 sample may be associated with the presence of Cyanobacteria and

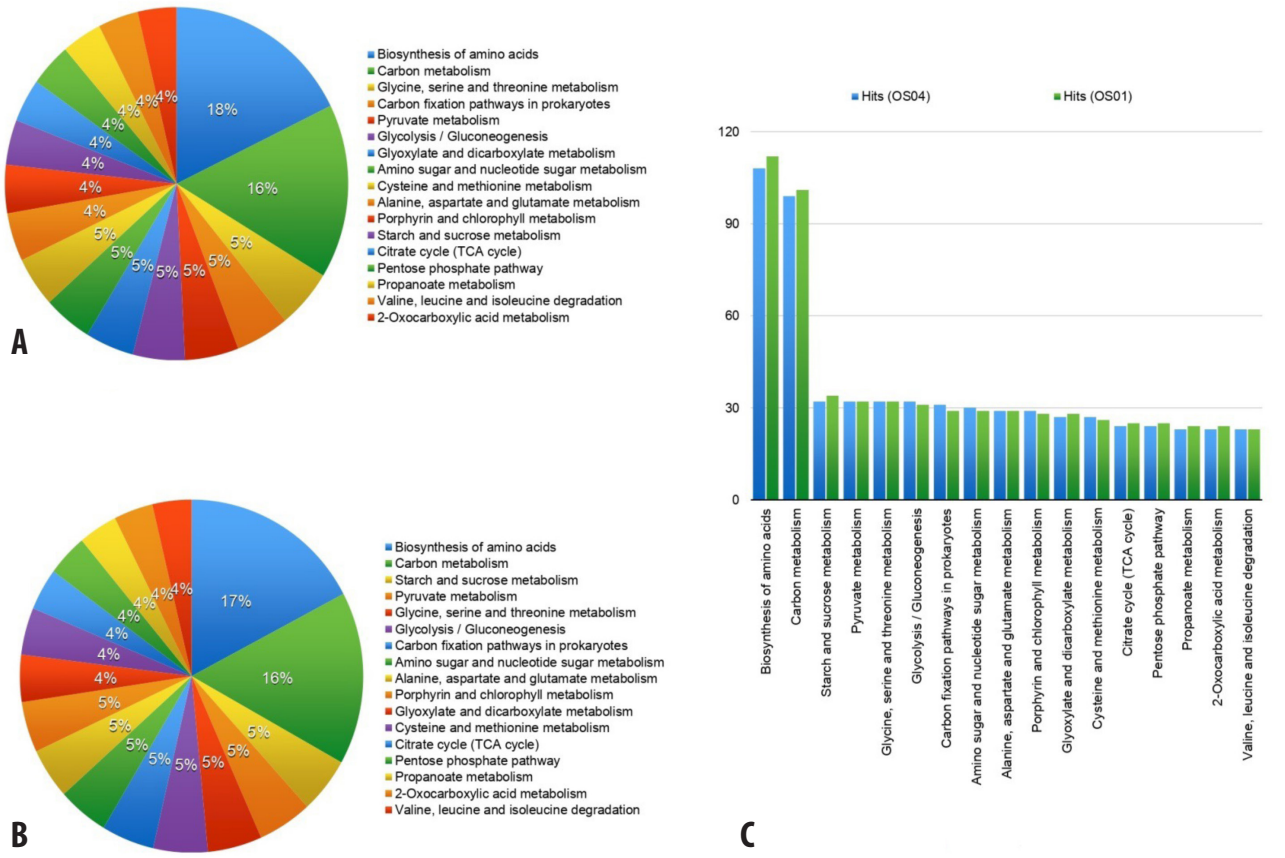

Fig. 4. PICRUSt2 predicted functional composition of endophytes in OS01 (A) and OS04 (B) in the form of pie charts. The overall metabolic pathways in the form of pie charts showed mostly the enrichment of primary metabolism pathways in the both samples. C, comparative bar graph of the samples based on their mapping to KEGG metabolic pathways. The details of the network enrichment pathways with *P* value file are given in the Appendix 2 and Appendix 3.

other algal members, which may function in the regulation and processing of byproducts of algal metabolism. Latescibacteria (formerly WS3) is a member of the Fibrobacteres-Chlorobi-Bacteroidetes super phylum that has anaerobic metabolism of the fermentative type. They also have the capability to degrade multiple polysaccharides and glycoproteins such as pectin, fucan, alginate, hydroxyproline-rich glycoproteins that represent cell walls components of Chlorophycean members. They also produce extensive machinery for the catabolism of all transported sugars, including a bacterial microcompartment to sequester propionaldehyde (Youssef et al. 2015). Sample OS01 lacked members of the Calditrichaeota, Hydrogenedentes, and Nitrospinae. The functions of these groups of organisms have been previously reported by several workers (Marshall et al 2017; Sun et al. 2019; Chen et al. 2021). Most of the Calditrichaeota genomes have been found to contain genes that render the organism O_2-tolerant (Marshall et al. 2008). The candidate phylum 'Hydrogenedentes' was named for the abundance of hydrogenases and putative H_2-utilizing pathways in the four SAGs that were the first partial genomes representing the phylum. The first partial genomes of the candidate phylum Hydrogenedentes revealed that there was an abundance of hydrogenases (Rinke et al. 2013), which was later confirmed using advanced methodologies revealing that they were effective lipolytic glycerol

Table 1. Physicochemical parameters of the soil from the study sites

Character	Parameter	Method adopted	OS01	OS04	Unit
Physical	pH at 25 °C	Using glass electrode pH meter (Jackson 1973)	6.18	6.14	pH units
	Sand	Hydrometer method	32.0	38.67	%
	Silt	Hydrometer method	26.7	27.85	%
	Clay	Hydrometer method	29.43	26.54	%
Chemical	Total organic carbon	Walkey Black method (Jackson 1962)	1.98	2.21	mg kg^{-1} soil
	Total nitrogen	Kjeldahl method (Jackson 1967)	34.56	46.56	mg kg^{-1} soil

degraders (Nobu et al. 2015). Absence of Nitrospinae in OS01 may be attributed to the fact that the soil composition of the area had lower nitrogen availability in the soil. Thus, the presence of these members in the OS04 sample and their subsequent absence in the OS01 sample is a direct indicator of the soil quality of the area, since the absence of the bacterial members is due to the lack of elements that these microbes utilize as metabolic intermediates in their life cycles. Excess ammonia or a rich algal assemblage can be detrimental to the overall growth of the plant as a whole and as a result, the absence of these microbes indicates that the soil and the subsequent endophytic compartment are quite suitable for growth. We also found many unique genera in both samples, which was assumed to be the result of differences in soil physiochemical properties.

The pathway analysis results revealed a large number of primary metabolic pathways that were abundantly enriched in both endophytic samples. Amino acid biosynthetic pathways and carbon metabolism pathways were found to be the most abundant, indicating a healthy environment for the growth of microorganisms and synergistic interactions with plant roots. A few interesting observations were also made, especially regarding the abundant sulphur metabolism pathways detected in the samples. Sulphur metabolism in the rice rhizospheric region leads to iron plaque formation, which prevents the uptake of arsenic in rice roots (Hu et al. 2007). Apart from sulphur metabolism, a few other enriched pathways such as tetracycline and streptomycin biosynthesis specific to the microbes were also detected. This appears as an interesting outcome, since the potential for these antibiotic genes to be transmitted via translocation to the edible parts of the plant represents a potential pathway for the spread of antimicrobial resistance by increasing the pool of antibiotic genes in the human gut (Ganguli et al. 2019). Furthermore, these genes may taken up by animals or leach into the soil, as dried plants are often used as animal fodder, and through the field residues in the form of stubble and stalks (Adijaya et al. 2021).

Nitrogen metabolism pathways were also found to be enriched, which is associated with the high abundance of *Bacillus* sp. in the microbiomes. Several studies have reported the presence of endophytic nitrogen fixing microorganisms and their beneficial effects on tropical crop plants (Cocking 2003; Muangthong et al. 2015; Suman et al. 2016). Thus, we can safely conclude that the endophytic consortia provided the plants with the benefits of nitrogen fixation. Biotin and geraniol metabolism pathways were also detected in the consortia. Both of these compounds have economic importance – the former is used as an essential supplement for hair follicle development (Okon, Labandera-Gonzalez 1994) and the latter is a very important commercial terpene alcohol used in the fragrance industry (Stepanyuk, Kirschning 2019). The ability of some members of the microbial consortia to effectively process these compounds indicates the possibility of molecular farming of these essential byproducts at the synthetic level. In addition, multi organism metabolic crosstalk may induce a beneficial synergistic effect (Santoyo 2022)

Finally, the data provided here gave interesting insights into the unique bacterial members present in the endophytic assemblages of the two varieties of rice studied. *Solibacillus* (endospore formers) and *Paenibacillus* (nitrogen fixation) were unique to OS01, while *Herbaspirillum* (nitrogen fixer) and *Pantoea* (lactose fermenter) were unique to OS04. *Solibacillus*, which have been reported from different soil assemblages, are a marker of good soil health (Hartmann et al. 2014). They may also offer protection from pathogen attack. In the future, we aim to explore whether these members can be used to serve as important markers of a healthy microbial consortium specific to a particular rice variety or may serve as a promising candidate of biofertilizer.

It is eminent that the efforts to evaluate the total microbial populations in a particular plant species may produce varied results, depending on the growth conditions of the host plant, and the mode in which the plant tissues were used (Santoyo et al. 2015). Our study gave important insights into the possible prevalent root endophytic microbiome of rice collected from cultivated land. As the threat of climate change looms large in the future, we have to seek alternative cropping cycles which could save us from losses incurred due to the natural calamities and flooding (Kruger et al. 2005; McInernerny et al. 2008; Breidenbach, Conrad 2015). Cropping cycles can become more effective if the soil and endophytic micro-environments synergistically interact with each other. Moreover, these endophytic microbes can be utilized and tested as an effective environment-friendly alternative of chemical fertilizers to suitably modify and reclaim degraded soils. It is recommended that longitudinal studies from diverse geographical locations of West Bengal be carried out in order to predict a possible universal core microbial consortium with immense agricultural and commercial potential.

Acknowledgements

Equal contribution of Subhadipa Sengupta and Sayak Ganguli is acknowledged. The work was supported by the Minor Research Project grants given to Subhadipa Sengupta (Grant No. PSW-039/14-15 ERO) by the University Grants Commission (UGC), New Delhi, India. The authors hereby acknowledge the Post Graduate Department of Botany, Bidhannagar College, Kolkata for providing the required space for conducting the work and Xcelris Labs Limited, Ahmedabad, Gujarat for conducting the sequencing part. Metagenome data is available at NCBI under the biosample accession no. SAMN06209694 and SAMN06209718.

References

Adijaya I., Budiari N., Sari R, Elizabeth P. 2021. The agronomy performance and financial feasibility of hybrid maize varieties for consumption and cattle feed in difference planting system.

IOP Conf. Ser. Earth Environ. Sci. 759: 012044.
Alori E.T., Glick B.R., Babalola O. 2017. Microbial phosphorus solubilization and its potential for use in sustainable agriculture. *Front. Microbiol.* 8: 971.
Azevedo J., Maccheroni W., Pereira J., Araújo W. 2000. Endophytic microorganisms: a review on insect control and recent advances on tropical plants. *Electron. J. Biotechnol.* 3: 15–16.
Breidenbach B., Conrad R. 2015. Seasonal dynamics of bacterial and archaeal methanogenic communities in flooded rice fields and effect of drainage. *Front. Microbiol.* 5: 752.
Cameron K.C., Di H.J., Moir J.L. 2013. Nitrogen losses from the soil/plant system: a review. *Ann. Appl. Biol.* 162: 145–173.
Chen P., Zhou H., Huang Y., Xie Z., Zhang M., Wei Y., Li J., Ma Y., Luo M., Ding W., Cao J., Jiang T., Nan P., Fang J., Li X. 2021. Revealing the full biosphere structure and versatile metabolic functions in the deepest ocean sediment of the Challenger Deep. *Genome Biol.* 22: 207.
Cocking E.C. 2003. Endophytic colonization of plant roots by nitrogen-fixing bacteria. *Plant Soil* 252: 169–175.
Correa-Galeote D., Bedmar E.J., Arone G.J. 2018. Maize endophytic bacterial diversity as affected by soil cultivation history. *Front. Microbiol.* 9: 484.
Cuevas V.C. 1991. Rapid composting for intensive rice land use. In: *Innovation for Rural Development.* SEAMEO-SEARCA, Los Baños, Philippines, pp. 5–10.
Daebeler A., Abell G.C., Bodelier P.L., Bodrossy L., Frampton D.M., Hefting M.M., Laanbroek H.J. 2012. Archaeal dominated ammonia-oxidizing communities in Icelandic grassland soils are moderately affected by long-term N fertilization and geothermal heating. *Front. Microbiol.* 3: 352.
Elekhtyar N. 2015. Efficiency of *Pseudomonas fluorescens* as plant growth-promoting rhizobacteria (PGPR) for the enhancement of seedling vigor, nitrogen uptake, yield and its attributes of rice (*Oryza sativa* L.). *Int. J. Sci. Res. Agric. Sci.* 2: 57–67.
Fadiji A.E., Babalola O.O. 2020. Elucidating mechanisms of endophytes used in plant protection and other bioactivities with multifunctional prospects. *Front. Bioeng. Biotechnol.* 8: 467.
Gaiero J.R., McCall C.A., Thompson K.A., Day N.J., Best A.S., Dunfield K.E. 2013. Inside the root microbiome: bacterial root endophytes and plant growth promotion. *Amer. J. Bot.* 100: 1738–1750.
Ganguli S., Pal S., Das K., Banerjee R., Bagchi S.S. 2019. Gut microbial dataset of a foraging tribe from rural West Bengal – insights into unadulterated and transitional microbial abundance. *Data Brief* 25: 103963.
Garcias-Bonet N., Arrieta J.M., De Santana C.N., Duarte C.M., Marbà N. 2012. Endophytic bacterial community of a Mediterranean marine angiosperm (*Posidonia oceanica*). *Front. Microbiol.* 3: 342.
Hardoim P.R, van Overbeek L.S., Berg G., Pirttilä A.M., Compant S., Campisano A., Döring M., Sessitsch A. 2015. The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol. Mol. Biol. Rev.* 79: 293–320.
Hartmann M., Frey B., Mayer J., Mäder P., Widmer F. 2014. Distinct soil microbial diversity under long-term organic and conventional farming. *ISME J.* 9: 1177–1194.
Hu Z.Y., Zhu Y.G., Li M., Zhang L.G., Cao Z.H., Smith F.A. 2007. Sulfur (S)-induced enhancement of iron plaque formation in the rhizosphere reduces arsenic accumulation in rice (*Oryza sativa* L.) seedlings. *Environ. Pollut.* 147: 387–393.
Jackson M.L 1962. *Soil Chemical Analysis*. Constable and Co., London.
Jackson M.L 1967. *Soil Chemical Analysis*. Prentice Hall of India, New Delhi.
Jackson M.L 1973. *Soil Chemical Analysis*. Prentice Hall of India, New Delhi.
Jana S.K., Islam M.M., Hore S., Mandal S. 2022. Rice seed endophytes transmit into the plant seedling, promote plant growth and inhibit fungal phytopathogens. *Plant Growth Regul.* https://doi.org/10.1007/s10725-022-00914-w.
Khare E., Mishra J., Arora N.K. 2018. Multifaceted interactions between endophytes and plant: developments and prospects. *Front. Microbiol.* 9: 2732.
Köberl M., Dita M., Martinuz A., Staver C., Berg G. 2017. Members of Gammaproteobacteria as indicator species of healthy banana plants on *Fusarium* wilt-infested fields in Central America. *Sci. Rep.* 7: 45318.
Koch H., van Kessel M.A.H.J., Lücker S. 2019. Complete nitrification: insights into the ecophysiology of *Comammox nitrospira. Appl. Microbiol. Biotechnol.* 103: 177–189.
Krüger M., Frenzel P., Kemnitz D., Conrad R. 2005. Activity, structure and dynamics of the methanogenic archaeal community in a flooded Italian rice field. *FEMS Microbiol. Ecol.* 51: 323–331.
Kunda P., Dhal P., Mukherjee A. 2018. Endophytic bacterial community of rice (*Oryza sativa* L.) from coastal saline zone of West Bengal: 16S rRNA gene based metagenomics approach. *Meta Gene* 18: 79–86.
Ma Z., Yue Y., Feng M., Li Y., Ma X., Zhao X., Wang S. 2019. Mitigation of ammonia volatilization and nitrate leaching via loss control urea triggered H-bond forces. *Sci. Rep.* 9: 15140.
Mahdi S.S., Hassan G.I., Samoon S.A., Rather H.A., Dar S.A., Zehra B. 2010. Bio-fertilizers in organic agriculture. *J. Phytol* 2: 42–54.
Mano H., Morisaki H. 2008. Endophytic bacteria in the rice plant. *Microbes Environ.* 23: 109–117.
Marshall I.P.G., Starnawski P., Cupit C., Fernández Cáceres E., Ettema T.J.G., Schramm A., Kjeldsen K.U. 2017. The novel bacterial phylum Calditrichaeota is diverse, widespread and abundant in marine sediments and has the capacity to degrade detrital proteins. *Environ. Microbiol. Rep.* 9: 397–403.
Mbengue A., Namdev P., Kumar T., Halder K., Bhattacharjee S. 2016. Next generation whole genome sequencing of *Plasmodium falciparum* using Next Seq 500 technology in India. *bioRxiv* doi: https//doi.org/10.1101/068676.
McInerney M.J., Struchtemeyer C.G., Sieber J., Mouttaki H., Stams A.J., Schink B., Rohlin L., Gunsalus R.P. 2008. Physiology, ecology, phylogeny, and genomics of microorganisms capable of syntrophic metabolism. *Ann. NY Acad. Sci.* 1125: 58–72.
Muangthong A., Youpensuk S., Rerkasem B. 2015. Isolation and characterisation of endophytic nitrogen fixing bacteria in sugarcane. *Trop. Life Sci. Res.* 26: 41–51.
Muthayya S., Sugimoto J.D., Montgomery S., Maberly G.F. 2014. An overview of global rice production, supply, trade, and consumption. *Ann. NY Acad. Sci.* 1324: 7–14.
Nair D.N., Padmavathy S. 2014. Impact of endophytic microorganisms on plants, environment and humans. *Sci. World J.* 2014: 250693.
Nobu M.K., Narihiro T., Rinke C., Kamagata Y., Tringe S.G., Woyke T., Liu W.T. 2015. Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic

bioreactor. *ISME J.* 9: 1710–1722.
Oerke E.C., Dehne H.W. 2004. Safeguarding production-losses in major crops and the role of crop protection. *Crop Prot.* 23: 275–285.
Okon Y., Labandera-Gonzalez C.A. 1994. Agronomic applications of *Azospirillum*: an evaluation of 20 years worldwide field inoculation. *Soil Biol. Biochem.* 26: 1591–1601.
Patel J.K., Archana G. 2017. Diverse culturable diazotrophic endophytic bacteria from Poaceae plants show cross-colonization and plant growth promotion in wheat. *Plant Soil* 417: 99–116
Radhakrishnan R., Hashem A., Abd Allah E.F. 2017. Bacillus: a biological tool for crop improvement through bio-molecular changes in adverse environments. *Front. Physiol.* 8: 667.
Rinke C., Schwientek P., Sczyrba A., Ivan ova N.N., Anderson I.J., Cheng J.F., Darling A., Malfatti S., Swan B.K., Gies E.A., Dodsworth J.A., Hedlund B.P., Tsiamis G., Sievert S.M., Liu W.T., Eisen J.A., Hallam S.J., Kyrpides N.C., Stepanauskas R., Rubin E.M., Hugenholtz P., Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499: 431–437.
Sahoo S., Sarangi S., Kerry R.G. 2017. Bioprospecting of endophytes for agricultural and environmental sustainability. In: Patra J., Vishnuprasad C., Das G. (Eds.) *Microbial Biotechnology.* Springer, Singapore, pp. 429–458.
Santoyo G. 2022. How plants recruit their microbiome? New insights into beneficial interactions. *J. Adv Res.* 40: 45–58.
Santoyo G., Moreno-Hagelsieb G., Orozco-Mosqueda M.C., Glick B.R. 2015. Plant growth-promoting bacterial endophytes. *Microbiol. Res.* 183: 92–99.
Schulz B., Wanke U., Draeger S., Aust H.J. 1993. Endophytes from herbaceous plants and shrubs: effectiveness of surface sterilization methods. *Mycol. Res.* 97: 1447–1450.
Sengupta S., Ganguli S., Singh P.K. 2017. Metagenome analysis of the root endophytic microbial community of Indian rice (*O. sativa* L.). *Genom. Data* 12: 41–43.
Sessitsch A., Reiter B., Berg G. 2004. Endophytic bacterial communities of field-grown potato plants and their plant-growth-promoting and antagonistic abilities. *Can. J. Microbiol.* 50: 239–249.
Shen F.T., Yen J.H., Liao C.S., Chen W.C., Chao Y.T. 2019. Screening of rice endophytic biofertilizers with fungicide tolerance and plant growth-promoting characteristics. *Sustainability* 11: 1133.
Stepanyuk A. Kirschning A. 2019 A Synthetic terpenoids in the world of fragrances: Iso E Super® is the showcase. *Beilstein J. Org. Chem.* 15: 2590–2602.
Strobel G.A. 2002. Rainforest endophytes and bioactive products. *Crit. Rev. Biotechnol.* 22: 315–333.
Sturz A.V., Christie B.R., Nowak J. 2000. Bacterial endophytes: potential role in developing sustainable systems of crop production. *Crit. Rev. Plant Sci.* 19: 1–30.
Sturz A.V, Kimpinski J. 2004. Endo root bacteria derived from marigolds (*Tagetes* spp.) can decrease soil population densities of root-lesion nematodes in the potato root zone. *Plant Soil* 262 :241–249.
Suman A., Yadav A.N., Verma P. 2016. Endophytic microbes in crops: diversity and beneficial impact for sustainable agriculture. In: Singh D., Singh H., Prabha R. (Eds.) *Microbial Inoculants in Sustainable Agricultural Productivity*. Springer, New Delhi, pp. 117–143.
Sun B., Wang X., Wang F., Jiang Y., Zhang X.X. 2013. Assessing the relative effects of geographic location and soil type on microbial communities associated with straw decomposition. *Appl. Environ. Microbiol.* 79: 3327–3335.
Sun X., Kop L.F.M., Lau M.C.Y., Frank J., Jayakumar A., Lücker S., Ward B.B. 2019. Uncultured *Nitrospina*-like species are major nitrite oxidizing bacteria in oxygen minimum zones. *ISME J.* 13: 2391–2402.
Torres M., White J., Zhang X., Hinton D. 2012. Endophyte-mediated adjustments in host morphology and physiology and effects on host fitness traits in grasses. *Fungal Ecol.* 5: 322–330.
Trivedi P., Delgado-Baquerizo M., Anderson I.C., Singh B.K. 2016. Response of soil properties and microbial communities to agriculture: implications for primary productivity and soil health indicators. *Front. Plant Sci.* 12: 990.
Turner T.R., James E.K., Poole P.S. 2013. The plant microbiome. *Genome Biol.* 14: 209.
Vacheron J., Desbrosses G., Bouffaud M.L., Touraine B., Moënne-Loccoz Y., Muller D., Legendre L., Wisniewski-Dyé F., Prigent-Combaret C. 2013. Plant growth-promoting rhizobacteria and root system functioning. *Front. Plant Sci.* 4: 356.
Wani S.H., Sah S.K. 2014. Biotechnology and abiotic stress tolerance in rice. *J. Rice Res.* 2: e105.
Wood D.E., Salzberg S.L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15: R46.
Xia Y., Liu J., Chen C., Mo X., Tan Q., He Y., Wang Z., Yin J., Zhou G. 2022. The multifunctions and future prospects of endophytes and their metabolites in plant disease management. *Microorganisms* 10: 1072.
Youssef N.H., Farag I.F., Rinke C., Hallam S.J., Woyke T., Elshahed M.S. 2015. *In silico* analysis of the metabolic potential and niche specialization of candidate phylum "Latescibacteria" (WS3). *PLoS ONE* 10: e0127499.
Zayadan B.K., Matorin D.N., Baimakhanova G.B., Bolathan K., Oraz G.D., Sadanov A.K. 2014. Promising microbial consortia for producing biofertilizers for rice fields. *Microbiology* 83: 391–397.

Appendix 1. Endophyte genera found in data sets of OS01 and OS04

Common genera found in both data sets OS01 and OS04	Unique genera in data OS01	Unique genera in data OS04
Magnetospirillum	*Lamprocystis*	*Phaeospirillum*
unclassified (derived from Rhizobiales)	*Listeria*	*Desulfobotulus*
Intrasporangium	*Thermanaerovibrio*	*Truepera*
Clostridium	*Ktedonobacter*	*Fusobacterium*
unclassified (derived from Firmicutes)	*Candidatus Thiobios*	*Melittangium*
Methylocystis	unclassified (derived from Desulfobacteraceae)	*Bacteroides*
Porphyromonas	*Nonomuraea*	*Anaerococcus*
Planctomyces	*Salinibacter*	*Desulfococcus*
Phormidium	*Desulfuromonas*	*Dehalococcoides*
Desulfomicrobium	*Acidaminococcus*	*Brevundimonas*
Acholeplasma	*Achnanthes*	*Hymenobacter*
Spirobacillus	*Agromyces*	*Delftia*
Pantoea	*Jeotgalibacillus*	*Desulforhabdus*
Alcaligenes	*Clavaria*	*Sphaerobacter*
unclassified (derived from Alcaligenaceae)	*Nitrosovibrio*	*Aeromicrobium*
Coraliomargarita	*Navicula*	*Desulfobacterium*
Megasphaera	*Prosthecobacter*	*Nostoc*
Hydrogenobacter	*Extubocellulus*	*Pandoraea*
Azovibrio	*Erythrobacter*	*Erythromicrobium*
Spirochaeta	*Rhodoplanes*	*Lysobacter*
Marinobacter	*Candidatus Magnetobacterium*	*Leptonema*
Herbaspirillum	unclassified (derived from Peptococcaceae)	*Anaerobranca*
Variovorax	*Melosira*	*Anaerostipes*
Chlorella	*Thiocapsa*	*Massilia*
unclassified (derived from Chroococcales)	*Kinetoplastibacterium*	*Cytophaga*
Thermosinus	*Trachelomonas*	*Desulfonauticus*
Thalassiosira	*Methylobacter*	*Anabaenopsis*
Candidatus Glomeribacter	*Cellulomonas*	*Rhodovibrio*
unclassified (derived from Lachnospiraceae)	*Methylonatrum*	*Faecalibacterium*
Seinonella	*Oxobacter*	*Ruminobacter*
Chlamydomonas	*Nephroselmis*	*Collimonas*
Couchioplanes	*Alkaliphilus*	*Paenibacillus*
unclassified (derived from Desulfobulbaceae)	*Moritella*	*Heliobacterium*
Lachnospira	*Kocuria*	*Acetivibrio*
Rubritalea	*Methanobacterium*	*Dermacoccus*
Syntrophobacter	*Phyllobacterium*	*Dethiosulfovibrio*
unclassified (derived from Alphaproteobacteria)	*Azospirillum*	*Agrostis*
unclassified (derived from unclassified sequences)	*Pirellula*	*Candidatus Tremblaya*
Candidatus Odyssella	*Actinomadura*	*Stenotrophomonas*
Calypogeia	*Rhodococcus*	*Sporosarcina*
Treponema	*Thiobaca*	*Legionella*
Denitromonas	*Scytonema*	*Thiodictyon*
Brevibacillus	*Chlamydophila*	*Thalassionema*
Rhodobacter	*Rhizobium*	*Penicillium*
Cellulosilyticum	unclassified (derived from Verrucomicrobia)	*Spiroplasma*
Nitrosospira	*Actinomyces*	*Candidatus Hamiltonella*
Aquimonas	*Oikopleura*	*Reinekea*
Methylocaldum	*Dialister*	*Saccharum*
Flavobacterium	*Thioalkalivibrio*	*Brochothrix*
Candidatus Desulforudis	*Bulleidia*	*Ectothiorhodospira*
Alistipes	*Laceyella*	*Ixodes*
Methylohalomonas	*Derxia*	*Brachymonas*
Marinilabilia	*Nannochloropsis*	*Mechercharimyces*

Continued

Appendix 1. *Continued*

Common genera found in both data sets OS01 and OS04	Unique genera in data OS01	Unique genera in data OS04
Candidatus Chloracidobacterium	unclassified (derived from Methylococcaceae)	*Lolium*
Taylorella	unclassified (derived from Nitrosomonadales)	unclassified (derived from Bacillaceae)
Succinimonas	*Terrabacter*	*Candidatus Protochlamydia*
Myxococcus	*Vitis*	*Flexithrix*
Syntrophothermus	*Acidimicrobium*	*Sphingobacterium*
Lyngbya	*Homo*	*Atrichum*
Streptococcus	*Arabidopsis*	*Methylomicrobium*
Corynebacterium	unclassified (derived from Comamonadaceae)	unclassified (derived from Gammaproteobacteria)
Cycloclasticus	*Amphiprora*	*Desulfatibacillum*
Prochlorococcus	*Carboxydothermus*	*Cylindrospermopsis*
Riemerella	unclassified (derived from Thermomonosporaceae)	*Anaerobaculum*
Rhodopseudomonas	*Shewanella*	*Bacteriovorax*
Hydrogenophaga	unclassified (derived from Salinisphaeraceae)	*Mycoplasma*
Achromobacter	*Sulfobacillus*	unclassified (derived from Alteromonadaceae)
Shuttleworthia	*Desulforegula*	*Haslea*
Aeromonas	*Capreolia*	*Pimelobacter*
Phenylobacterium	*Leyanella*	*Agrobacterium*
Desulfobacula	*Oscillatoria*	unclassified (derived from Ruminococcaceae)
Spirulina	*Plesiocystis*	unclassified (derived from Betaproteobacteria)
Ornithobacterium	*Magnetococcus*	*Cupriavidus*
Pinus	*Sphingomonas*	*Psychromonas*
Microbacterium	*Caldanaerobacter*	*Chromatium*
Peptostreptococcus	*Chondromyces*	*Marichromatium*
Rubrivivax	*Thiohalomonas*	*Alkalispirillum*
Deinococcus	*Bdellovibrio*	*Oscillochloris*
Bifidobacterium	*Bosea*	*Methanosarcina*
Sphingosinicella	unclassified (derived from Opitutaceae)	*Symbiobacterium*
Polymorphospora	*Micromonospora*	*Desulfovibrio*
Blastomonas	*Azoarcus*	unclassified (derived from Cyanobacteria)
unclassified (derived from Bacteroidetes)	*Rhodovulum*	*Streptosporangium*
Cyclotella	*Halochromatium*	*Maricaulis*
unclassified (derived from Nitrosomonadaceae)	*Curvibacter*	*Pseudonocardia*
Aminobacterium	*Geobacter*	*Porphyrobacter*
Desulfonema	*Olisthodiscus*	*Cryobacterium*
Desulfocapsa	*Actinocorallia*	*Dyella*
Candidatus Koribacter	*Leptothrix*	*Nesterenkonia*
Gemmatimonas	*Nocardia*	*Alteromonas*
Leptospira	*Blastopirellula*	*Methylothermus*
Microbispora	*Coptotermes*	*Tetrasphaera*
Eucampia	*Desulfoglaeba*	*Nitrospira*
Anaplasma	*Crenothrix*	*Xanthomonas*
Afipia	*Flexibacter*	*Cyanidium*
Candidatus Nitrosocaldus	*Heliophilum*	*Acetobacterium*
Thermoleophilum	*Verminephrobacter*	*Halobacillus*
Tropheryma	*Candidatus Nitrososphaera*	*Syntrophus*
Synechococcus	*Brachybacterium*	unclassified (derived from Burkholderiales)
Pseudomonas	*Oligella*	*Burkholderia*
Beutenbergia	*Candidatus Portiera*	*Volvox*
Thermoanaerobacter	*Dorea*	*Ammonifex*
Ethanoligenens	*Geoalkalibacter*	*Azospira*
Parabacteroides	*Leptolyngbya*	*Megamonas*

Continued

Appendix 1. *Continued*

Common genera found in both data sets OS01 and OS04	Unique genera in data OS01	Unique genera in data OS04
Sinorhizobium	*Staphylococcus*	*Salinispora*
Erwinia	*Raphidonema*	*Cycas*
unclassified (derived from Thermoanaerobacterales)	*Thiohalorhabdus*	*Capnocytophaga*
Atopobium	*Sisymbrium*	*Ruminococcus*
Nocardioides	unclassified (derived from Enterobacteriaceae)	*Subtercola*
Finegoldia	unclassified (derived from Epsilonproteobacteria)	*Piscirickettsia*
Acidithiobacillus	*Arthrobacter*	*Rhodospirillum*
unclassified (derived from Desulfovibrionales)	*Lepeophtheirus*	*Viridibacillus*
Methanosphaerula	*Geranium*	*Oceanimonas*
Cyclobacterium	*Dolichospermum*	*Kribbella*
Natranaerobius	*Hyphomicrobium*	*Pseudacidovorax*
	Allochromatium	*Mesoplasma*
	Hydrogenobaculum	*Desulfosarcina*
	Mycobacterium	*Acipenser*
	Tetrathiobacter	*Lysinibacillus*
	Sphaerospermopsis	*Escherichia*
	Pleurocapsa	*Vaccinium*
	Kurthia	unclassified (derived from Methylocystaceae)
	Geodermatophilus	unclassified (derived from Podoviridae)
	Phaeodactylum	*Nocardiopsis*
	Chitinophaga	*Desulfitobacterium*
	Bradyrhizobium	unclassified (derived from Sphingobacteriaceae)
	Janibacter	*Sideroxydans*
	Criblamydia	*Sulfurimonas*
	Geitlerinema	*Vaucheria*
	Streptomyces	unclassified (derived from Prasinophyceae)
	Marinoscillum	*Paucimonas*
	Halothermothrix	*Persephonella*
	Alicyclobacillus	unclassified (derived from Clostridiales Family XI. Incertae Sedis)
	Methylomonas	*Pedobacter*
	Zoogloea	*Alkalilactibacillus*
	unclassified (derived from Chromatiaceae)	*Rothia*
	Trichomonas	*Planomonospora*
	unclassified (derived from Euryarchaeota)	*Aneurinibacillus*
	Acidobacterium	*Thermobaculum*
	unclassified (derived from Rhodocyclaceae)	*Dictyoglomus*

Appendix 2. Detailed pathway data containing the KEGG Orthology accessions of OS01

Pathway	Total	Expected	Hits	Pval	FDR
Biosynthesis of amino acids	222	57.7	112	1.36E–16	2.02E–14
Carbon metabolism	249	64.7	101	6.69E-08	4.95E–06
Lipopolysaccharide biosynthesis	17	4.42	14	1.73E–06	8.3E–05
Glyoxylate and dicarboxylate metabolism	51	13.3	29	2.24E–06	8.3E–05
Carbon fixation pathways in prokaryotes	60	15.6	32	4.42E–06	0.000131
Selenocompound metabolism	15	3.9	12	1.78E–05	0.000439
Valine, leucine and isoleucine biosynthesis	15	3.9	11	0.00016	0.00339
Glycine, serine and threonine metabolism	78	20.3	34	0.000455	0.00812
Streptomycin biosynthesis	12	3.12	9	0.000519	0.00812
Amino sugar and nucleotide sugar metabolism	64	16.6	29	0.000549	0.00812
Citrate cycle (TCA cycle)	53	13.8	25	0.000633	0.00836
Alanine, aspartate and glutamate metabolism	62	16.1	28	0.00073	0.00836
Folate biosynthesis	29	7.54	16	0.000734	0.00836
Pyruvate metabolism	74	19.2	32	0.000791	0.00837
Terpenoid backbone biosynthesis	23	5.98	13	0.00174	0.0172
Propanoate metabolism	55	14.3	24	0.00307	0.0284
Cysteine and methionine metabolism	71	18.5	29	0.00394	0.0343
Polyketide sugar unit biosynthesis	4	1.04	4	0.00454	0.0373
Porphyrin and chlorophyll metabolism	69	17.9	28	0.00511	0.0389
D-Glutamine and D-glutamate metabolism	6	1.56	5	0.00552	0.0389
Geraniol degradation	6	1.56	5	0.00552	0.0389
Valine, leucine and isoleucine degradation	58	15.1	24	0.00697	0.0456
Peptidoglycan biosynthesis	13	3.38	8	0.00715	0.0456
Glycolysis / Gluconeogenesis	80	20.8	31	0.0074	0.0456
Starch and sucrose metabolism	65	16.9	26	0.00857	0.0479
Carbon fixation in photosynthetic organisms	35	9.1	16	0.00874	0.0479
Histidine metabolism	35	9.1	16	0.00874	0.0479
C5-Branched dibasic acid metabolism	11	2.86	7	0.00936	0.0495
Pentose phosphate pathway	63	16.4	25	0.0111	0.0555
Biotin metabolism	19	4.94	10	0.0116	0.0555
2-Oxocarboxylic acid metabolism	57	14.8	23	0.0116	0.0555
Nitrogen metabolism	17	4.42	9	0.0158	0.0728
Chloroalkane and chloroalkene degradation	23	5.98	11	0.0194	0.0869
Cyanoamino acid metabolism	10	2.6	6	0.0237	0.103
One carbon pool by folate	24	6.24	11	0.0276	0.117
Nicotinate and nicotinamide metabolism	36	9.36	15	0.0284	0.117
Fatty acid metabolism	52	13.5	20	0.0313	0.125
Fatty acid biosynthesis	28	7.28	12	0.0382	0.149
Tetracycline biosynthesis	6	1.56	4	0.0429	0.16
Fructose and mannose metabolism	44	11.4	17	0.0433	0.16
Sulfur metabolism	38	9.88	15	0.0465	0.168

Appendix 3. Detailed pathway data containing the KEGG Orthology accessions of OS04

Pathway	Total	Expected	Hits	Pval	FDR
Biosynthesis of amino acids	222	57.3	108	8.73E–15	1.29E-12
Carbon metabolism	249	64.3	99	2.09E–07	1.54E–05
Lipopolysaccharide biosynthesis	17	4.39	14	1.57E–06	7.77E–05
Carbon fixation pathways in prokaryotes	60	15.5	31	1.27E–05	0.000469
Glyoxylate and dicarboxylate metabolism	51	13.2	27	2.64E–05	0.000782
Starch and sucrose metabolism	65	16.8	32	3.3E–05	0.000815
Selenocompound metabolism	15	3.87	11	0.00015	0.00317
Amino sugar and nucleotide sugar metabolism	64	16.5	30	0.000184	0.0034
Alanine, aspartate and glutamate metabolism	62	16	29	0.000247	0.00406
Streptomycin biosynthesis	12	3.1	9	0.000491	0.00726
Folate biosynthesis	29	7.49	16	0.000675	0.00856
Pyruvate metabolism	74	19.1	32	0.000694	0.00856
Histidine metabolism	35	9.04	18	0.000957	0.0106
Valine, leucine and isoleucine biosynthesis	15	3.87	10	0.00101	0.0106
Citrate cycle (TCA cycle)	53	13.7	24	0.0015	0.0148
Glycine, serine and threonine metabolism	78	20.1	32	0.00203	0.0183
Porphyrin and chlorophyll metabolism	69	17.8	29	0.00211	0.0183
One carbon pool by folate	24	6.2	13	0.00272	0.0224
Glycolysis / Gluconeogenesis	80	20.7	32	0.00329	0.0256
Polyketide sugar unit biosynthesis	4	1.03	4	0.00441	0.0326
D-Glutamine and D-glutamate metabolism	6	1.55	5	0.00534	0.0344
Geraniol degradation	6	1.55	5	0.00534	0.0344
Tetracycline biosynthesis	6	1.55	5	0.00534	0.0344
Terpenoid backbone biosynthesis	23	5.94	12	0.00591	0.0365
Propanoate metabolism	55	14.2	23	0.00643	0.0381
Peptidoglycan biosynthesis	13	3.36	8	0.00683	0.0389
Carbon fixation in photosynthetic organisms	35	9.04	16	0.00814	0.0446
C5-Branched dibasic acid metabolism	11	2.84	7	0.00898	0.0475
2-Oxocarboxylic acid metabolism	57	14.7	23	0.0107	0.0536
Biotin metabolism	19	4.91	10	0.011	0.0536
Nicotinate and nicotinamide metabolism	36	9.3	16	0.0112	0.0536
Valine, leucine and isoleucine degradation	58	15	23	0.0135	0.0612
Fatty acid metabolism	52	13.4	21	0.0142	0.0612
Fatty acid biosynthesis	28	7.23	13	0.0143	0.0612
Cysteine and methionine metabolism	71	18.3	27	0.0145	0.0612
Pantothenate and CoA biosynthesis	29	7.49	13	0.0199	0.0804
Pentose phosphate pathway	63	16.3	24	0.0201	0.0804
Cyanoamino acid metabolism	10	2.58	6	0.0229	0.089
Thiamine metabolism	23	5.94	10	0.0491	0.186

Received 13 September 2022; received in revised form 27 November 2022; accepted 1 December 2022