Research Article

Genomic Analysis of *Bacillus megaterium* NCT-2 Reveals Its Genetic Basis for the Bioremediation of Secondary Salinization Soil

Bin Wang [iD],[1] **Dan Zhang,**[1] **Shaohua Chu** [iD],[1] **Yuee Zhi,**[1] **Xiaorui Liu** [iD],[2] **and Pei Zhou** [iD][1]

[1]*School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China*
[2]*The International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China*

Correspondence should be addressed to Xiaorui Liu; xiaorui1211@126.com and Pei Zhou; peizhousjtu@163.com

Bacillus megaterium NCT-2 is a nitrate-uptake bacterial, which shows high bioremediation capacity in secondary salinization soil, including nitrate-reducing capacity, phosphate solubilization, and salinity adaptation. To gain insights into the bioremediation capacity at the genetic level, the complete genome sequence was obtained by using a multiplatform strategy involving HiSeq and PacBio sequencing. The NCT-2 genome consists of a circular chromosome of 5.19 Mbp and ten indigenous plasmids, totaling 5.88 Mbp with an average GC content of 37.87%. The chromosome encodes 5,606 genes, 142 tRNAs, and 53 rRNAs. Genes involved in the features of the bioremediation in secondary salinization soil and plant growth promotion were identified in the genome, such as nitrogen metabolism, phosphate uptake, the synthesis of organic acids and phosphatase for phosphate-solubilizing ability, and Trp-dependent IAA synthetic system. Furthermore, strain NCT-2 has great ability of adaption to environments due to the genes involved in cation transporters, osmotic stress, and oxidative stress. This study sheds light on understanding the molecular basis of using *B. megaterium* NCT-2 in bioremediation of the secondary salinization soils.

1. Introduction

Soil application of organic and inorganic fertilizers for crop and vegetable cultivation is the major source for soil nitrate-nitrogen (nitrate-N), which increases agricultural productivity. However, the vegetable yields do not increase continuously with soil nitrate-N [1]. A large accumulation of nitrate in soil results in soil secondary salinization, having various adverse effects on soil productivity, and nitrate accumulation in vegetables [2]. What is more, the reduction of nitrate to nitrite can cause various human diseases [1]. Soil secondary salinization is a severe problem in intensively managed agricultural ecosystems [3]. It is required to develop a low-cost bioremediation method to remove nitrate from soil.

In our previous study, *Bacillus megaterium* NCT-2 was isolated from the secondary nitrate-salinized soil in a greenhouse, which shows high nitrate-reducing capacity and salinity adaptation in secondary salinization soil [4]. It can

remove nitrate at initial nitrate-N concentrations ranging from 100 mg/L to 1,000 mg/L and grow well in inorganic salt medium with 4.0% sodium chloride [4]. In our field trails, the concentrations of NO_3^- in both soil and plant were reduced significantly when we used the NCT-2 strain mixed with straw powder to treat secondary salinization soil (unpublished). Moreover, this strain showed significant phosphate-solubilizing ability of insoluble inorganic phosphates in the culture medium [5]. Strain NCT-2 has the potential to be utilized as a biofertilizer for bioremediation of the secondary nitrate-salinized soil and plant growth promotion [6].

The Gram-positive bacterium *Bacillus megaterium* is found in diverse habitats from soil to sediment, sea, and dried food. It was named after its big size with a volume approximately 100 times than that of *Escherichia coli* [7]. Its big size made it ideal to be used in studies of cell structure, protein localization, sporulation, and membranes [8, 9]. Due to no production of endotoxins associated with the outer membrane and no external alkaline proteases, they are used widely

as desirable cloning hosts in food and pharmaceutical production processes for α- and β-amylases in the baking industry [10, 11], penicillin acylase [12–14], and vitamin B12 [15], such as *Bacillus megaterium* DSM 319, *Bacillus megaterium* QM B1551, and *Bacillus megaterium* WSH 002 [16, 17]. The genomes of them have been sequenced to gain insights into the metabolic versatility that facilitate biotechnological applications, not the bioremediation of secondary salinization soil [18, 19].

Despite the previously published work sequenced the 5.68 Mb draft genome of *B. megaterium* NCT-2 by using the Solexa platform, consisting of the 204 contigs, it focused only on the multiple alignments of nitrate assimilation-related gene sequences [20]. The functional nitrate assimilation-related genes (the nitrate reductase electron transfer subunit, the nitrate reductase catalytic subunit, the nitrite reductase [NAD(P)H] large subunit and small subunit, and the glutamine synthetase) were identified [20]. The genes that could be involved in the full potential of strain NCT-2 in the bioremediation of secondary salinization soil remain unknown. For this, we obtained its complete genome sequence by using a multiplatform strategy involving HiSeq and PacBio sequencing. Furthermore, we performed a comprehensive analysis of nitrogen metabolism and plant growth-promoting features. The comparative analysis might be helpful for use in soil bioremediation.

2. Methods

2.1. DNA Preparation and Genome Sequencing. *B. megaterium* NCT-2, isolated from the secondary salinized greenhouse soil in China, was cultured in a defined inorganic salt medium as previously described [4]. It was registered in China General Microbiological Culture Collection Center under CGMCC No. 4698. Genomic DNA was isolated using QIAGEN DNeasy Blood & Tissue Kit (Hilden, Germany). The concentration and quality of DNA were determined by a Qubit Fluorometer (Thermo Scientific, USA), NanoDrop Spectrophotometer (Thermo Scientific, USA), and agarose electrophoresis. The whole genome of the *B. megaterium* strain NCT-2 was sequenced by the BGI Tech Solutions Co., Ltd. (Shenzhen, China) by using Illumina Hiseq 4000 short-read sequencing platform (Illumina Inc., San Diego, CA, USA) (insert size, 500 bp; 2×125 bp read length) and PacBio RSII long-read sequencing platform (Pacific Biosciences of California, Inc., Menlo Park, CA, USA) (Figure S1).

2.2. Genome Assembly and Annotation. After quality control, the *de novo* assembly of the whole NCT-2 genome was performed using the RS_HGAP Assembly3 in the SMRT Analysis pipeline version 2.2.0 [21]. The HiSeq clean reads were preliminarily assembled into contigs and then were used for hybrid error correction of the subreads from PacBio. There were two rounds of error correction. One was analyzed by using SOAPsnp and SOAPIndel [22] and another was by using the Genome Analysis Toolkit (GATK) [23]. Finally, SSPACE-LongRead [24] and Celera assemble [25] were used to generate a high-quality genome. The finished NCT-2

genome was submitted to GenBank, replacing the previous version of the draft genome [20].

The protein-coding genes were predicted by using Glimmer 3.02 [26], and the tandem repeats were detected with Tandem Repeat Finder 4.04 [27]. The gene function annotation was accomplished by blasting the protein sequences against the database of Kyoto Encyclopedia of Genes and Genomes (KEGG) [28]. In addition, the RAST web server (https://rast.nmpdr.org) with the default parameters was used to catalog all the predicted genes into subsystems according to functional categories [29, 30]. CGView was used to produce the maps of the circular genomes with gene feature information [31]. Genome alignments with locally collinear blocks were performed with MAUVE [32].

2.3. Phylogenetic Analysis. The whole genome-based phylogenetic analysis was performed by using the CVTree 3.0 online server [33, 34]. Fourteen genome sequences were obtained from GenBank. A phylogenetic tree was constructed by the neighbor-joining method using MEGA analysis [35–37]. In addition, FusionDB was used to analyze the functional repertories of *B. megaterium* NCT-2 and identify the nearest "neighbors" based on the functional similarities [38, 39].

3. Results and Discussion

3.1. General Genomic Characteristics. A total of ~1,189 Mb raw data and ~1,147 Mb clean data were obtained after filtering the low-quality reads generated by the HiSeq platform. The PacBio platform yielded 48,392 polymerase reads (with the average size of 12.9 kb) and 622 Mb subreads after quality control. The complete genome was assembled by taking advantage of the higher accuracy short reads from the HiSeq platform and the long subreads from the PacBio platform. The genome consists of a circular chromosome of 5.19 Mb with an average GC content of 38.2% (accession number: CP032527.2) and ten circular plasmids designated as the plasmid pNCT2-1 to pNCT2-10 (accession numbers: CP032528.1-CP032537.1). Sequence information was visualized in CG view Server (Figure 1 and S2). The total genome size is 5.88 Mb with an average GC content of 37.87%. The whole genome contains 6,039 genes, including 5,606 coding sequences, 203 RNA genes, and 230 pseudo genes. There are 127 identified tandem repeat sequences (TRF), 83 minisatellite DNA, and 7 microsatellite DNA.

The general features of *B. megaterium* NCT-2 were compared with five genomes of *Bacillus* strains (*Bacillus megaterium* DSM 319, *Bacillus megaterium* QM B1551, *Bacillus subtilis* subsp. subtilis str. 168, *Bacillus cereus* Q1, and *Bacillus licheniformis* DSM 13) (Table 1). The genome GC contents for three *B. megaterium* strains are around 38%. Strain NCT-2 has the largest genome size and most coding sequences and RNA genes, such as 53 rRNAs and 142 tRNAs. There were 14 rRNA operons on the negative chain and one rRNA operon on the positive strand with a 16S-23S-5S organization. In addition, the positive chain had one unusual rRNA operon with a 16S-23S-5S-5S organization and a single 5S rRNA. The microbial genome size is positively correlated
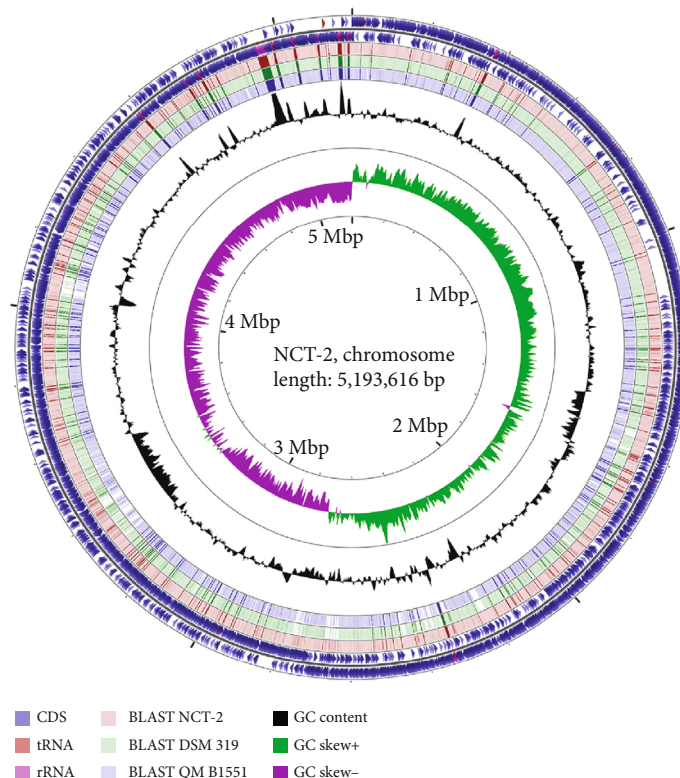
FIGURE 1: Genetic and physical map of the genome of *B. megaterium* NCT-2 prepared using CGView. Circles from the outside to the inside show the position of protein-coding sequences (blue), tRNA gene (red), and rRNA genes (pink) on the positive (circle 1) and negative (circle 2) strands. Circles 3-5 show the positions of BLAST hits detected through BLASTx comparisons of *B. megaterium* NCT-2 against itself (circle 3), *B. megaterium* DSM 319 (circle 4), and *B. megaterium* QM B1551 (circle 5). Circles 6 and 7 show plots of GC content and GC skew plotted as the deviation from the average for the entire sequence.

TABLE 1: General genome features of *B. megaterium* NCT-2 compared with other five *Bacillus* strains.

Strain	*B. megaterium* NCT-2	*B. megaterium* QM B1551	*B. megaterium* DSM 319	*B. subtilis* 168	*B. cereus* Q1	*B. licheniformis* DSM 13
Genome size (Mb)	5.88	5.52	5.10	4.22	5.51	4.22
Chromosome size (Mb)	5.19	5.10	5.10	4.22	5.21	4.22
G+C content (%)	37.8	37.97	38.1	43.5	35.5	46.2
Chromosomal G+C content (%)	38.2	38.3	38.1	43.5	35.6	46.2
Gene number	6039	5674	5245	4536	5856	4382
Coding sequence number	5606	5379	4941	4237	5513	4219
RNA gene number	203	182	153	116	137	98
rRNA genes (5S, 16S, and 23S)	53 (19, 17, 17)	37 (13, 12, 12)	33 (11, 11, 11)	30 (10, 10, 10)	39 (13, 13, 13)	21 (7, 7, 7)
tRNA gene number	142	137	114	86	93	72
Plasmid number	10	7	0	0	2	0

with their environment adaptability [40]. One typical characteristic of soil microorganisms is the high number of rRNAs, which is helpful for fast growth, successful sporulation, germination, and rapid response to changing the availability of nutrients [41–44]. These features indicate that strain NCT-2 has great ability of adaptation to various environments.

Most strains of *Bacillus megaterium* carry multiple plasmids, such as strain QM B1551 has seven resident plasmids

[18], *Bacillus megaterium* strain 216 has ten plasmids [45], and *Bacillus megaterium* NBRC 15308 has six plasmids. As for the ten plasmids in strain NCT-2, the sizes range from 9,625 bp to over 132 kb making up 11.7% of the whole genome (Table S1). The plasmids have significantly lower GC contents than the chromosome (33.7-37.0% versus 38.2%). There are 761 coding sequences and 23 RNA genes. Both plasmids pNCT2-2 and pNCT2-6 had one tRNA. In

addition, pNCT2-7 had 18 tRNAs, one 5S RNA, one large subunit ribosomal RNA (LSU rRNA), and one small subunit ribosomal RNA (SSU rRNA). Additional rRNA operons carried on plasmids slowed the growth rates of *E. coli* on poor carbon sources [46]. Further investigations are needed to clarify the role of plasmids in bacterial growth and adaptations to high-nitrate environments in bioremediation of the secondary salinization soils.

3.2. Phylogenetic Lineage Analysis.
We used CVTree 3.0 to construct a phylogenetic tree based on the complete proteomes with *Macrococcus caseolyticus* JCSC5402 as an outgroup. The obtained tree (Figure 2(a)) indicated that *B. megaterium* NCT-2 was most homologous to *B. megaterium* DSM 319 and then *B. megaterium* QM B1551. Similarly, genome comparison using the RAST Prokaryotic Genome Annotation Server also showed that the genomic sequence of NCT-2 had a higher comparison score with *B. megaterium* QM B1551 and *B. megaterium* DSM 319 (Figure S3). Furthermore, 16S rDNA sequences from 15 *Bacillus* strains were used to construct a phylogenetic tree by MEGA7 with the neighbor-joining method. The neighbor-joining phylogenetic tree shows that strain NCT-2 is closest to *B. megaterium* QM B1551, *B. megaterium* DSM 319, and *B. megaterium* WSH 002 (Figure 2(b)). Whole-genome alignment of *B. megaterium* NCT-2 to closely related QM B1551 and DSM 319 by using MAUVE revealed that the chromosomes of the three strains showed overall collinearity (Figure 2(c)).

3.3. Functional Annotations of B. megaterium NCT-2.
To investigate the function of the 5,606 coding sequences, the GO database, the KEGG database, the COG database, and RAST web server were used. The 3,159 genes annotated by GO were classified into biological processes, cellular components, and molecular functions (Figure S4). The top five categories were catalytic activity (1,822), metabolic process (1,786), cellular process (1,567), single-organism process (1,400), and binding (1,214).

2,338 chromosomal genes (44%) were assigned into 477 subsystems by RAST (Figure S5a). Subsystem category comparisons among six related *Bacillus* strains showed that the number of genes involved in "Amino Acids and Derivatives" and "Carbohydrates" was highest in the genome of the six strains (Figure 3(a)). In addition, *Bacillus megaterium* has more genes involved in "Cofactors, Vitamins, Prosthetic Groups, Pigments." The top five categories in strain NCT-2 were the "Amino Acids and Derivatives" (538), "Carbohydrates" (500), "Cofactors, Vitamins, Prosthetic Groups, Pigments" (340), "Protein Metabolism" (283), and "Fatty Acids, Lipids, and Isoprenoids" (180).

Likewise, 2,962 genes annotated by the KEGG database were assigned to 38 pathways (Figure 3(b)). The top five enriched pathways were "Biosynthesis of other secondary metabolism" (710), "Signaling molecules and interaction in Environmental information processing" (542), "Substance dependence" (540), "Nucleotide metabolism" (475), and "Immune disease" (472).

Like most strains of *B. megaterium*, which carry more than four plasmids, strain NCT-2 harbors ten indigenous plasmids. Only 75 genes (10%) were assigned into 37 subsystems by RAST (Figure S5b), including genes for riboflavin metabolism, butanol biosynthesis, and xylose utilization, and parts of genes in benzoate degradation and metabolism of central aromatic intermediates. There are also genes for cobalt-zinc-cadmium resistance, oxidative stress, and nitrosative stress.

3.4. Microbial Functional Similarities.
The translated protein sequence of *B. megaterium* NCT-2 was downloaded from RAST and submitted to the FusionDB web server (https://services.bromberglab.org/fusiondb/mapping) [38]. The submitted proteome (containing 5,364 proteins) matched to 3,662 FusionDB functions, while 228 proteins could not be mapped to any function in their database. The functional similarities of *B. megaterium* NCT-2 with 1,374 taxonomically distinct bacteria (with similarity > 40%) were shown in Table S2, most of them were soil bacterium. Strain NCT-2 is most functionally similar to *B. megaterium* DSM 319 (90%) and *B. megaterium* QM B1551 (89%). The functional relationships among nine *Bacillus* strains were demonstrated by the fusion+ networks (Figure 4(a)). There were 1,290 functions shared by all of them. The common functional annotations related to nitrogen metabolism were nitrite transporter NirC, nitrogen-fixing NifU domain protein, nitroreductase, nitrate transport protein, and 2-nitropropane dioxygenase. Notably, there are 3,047 functions shared among three strains of *B. megaterium* (strain NCT-2, strain QM B1551, and strain DSM 319) (Figure 4(b)). Strain NCT-2 has most of the core genes and pathways, including vitamin biosynthesis and nitrogen metabolism. The nitrogen metabolism-related genes, such as those encoding nitrate transport protein, nitrate/nitrite sensor protein, nitric oxide reductase activation protein, nitrite reductase [NAD(P)H] large subunit, nitrite reductase [NAD(P)H] small subunit, nitrite transporter, nitrite-sensitive transcriptional repressor, nitrogen regulatory protein P-II, nitrogen-fixing NifU domain protein, nitroreductase, and nitroreductase family protein, were located on the chromosome of the three strains. Furthermore, only strain NCT-2 carries the gene encoding for periplasmic nitrate reductase.

3.5. Genome Inventory for Nitrogen Metabolism.
In our field experiment, strain NCT-2 shows high nitrate-reducing capacity in secondary salinization soil (unpublished). The functional nitrate assimilation-related genes that are involved in the process of converting nitrate to glutamine have been identified [20]. The genes encoding nitrate and nitrite reductase were cloned and overexpressed in *Escherichia coli* [47]. Here, the whole genomic analysis also revealed the genes encoding sensor, transporter, and enzymes are involved in nitrogen metabolism. The genes were scattered in the chromosome. Genes encoding nitrite-sensitive transcriptional repressor (NsrR), which is directly sensitive to nitrosative stress, were found in both the chromosome and the plasmid (Table S3 and Figure S6). *B. megaterium* NCT-2 possessed
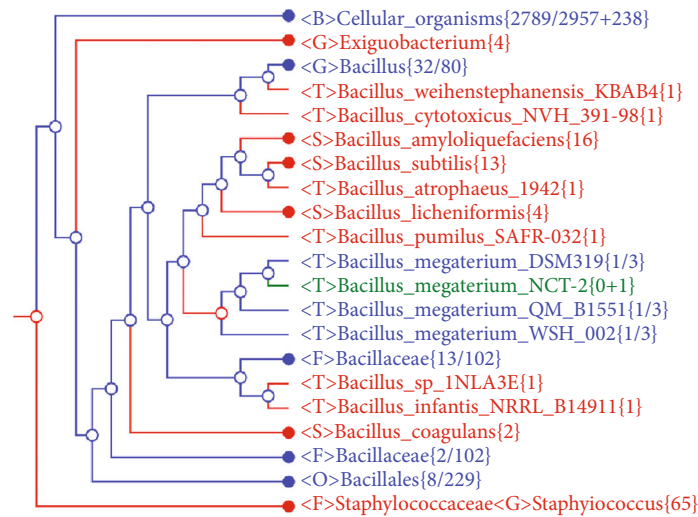
(a)
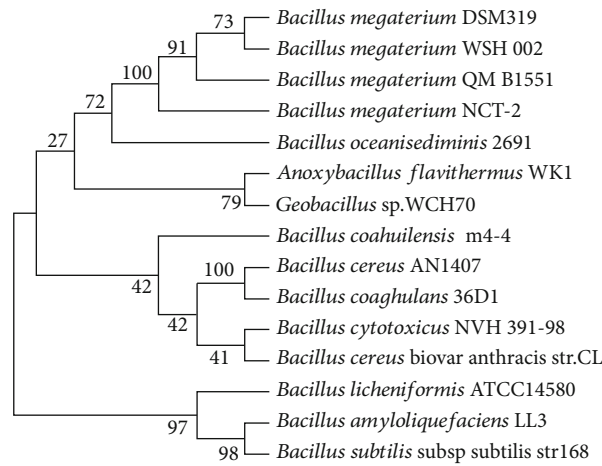
(b)
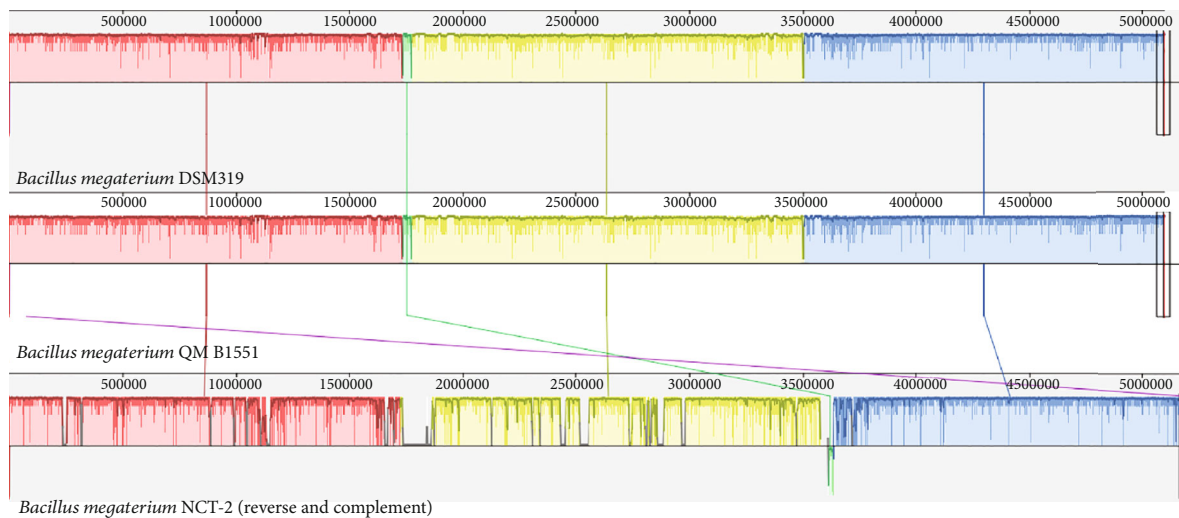
(c)

FIGURE 2: Phylogenetic tree showing the position of *Bacillus megaterium* NCT-2. (a) The tree was constructed based on the frequency of >6-string predicted peptides from the whole genome sequences by using CVTree 3.0. (b) The tree is based on 16S rDNA phylogenetic analysis by using MEGA7 with the neighbor-joining method. The bootstrap consensus tree inferred from 1,000 replicates is taken to represent the evolutionary history of the taxa analyzed. (c) Chromosomal similarity among strain NCT-2, DSM 319, and QM B1551 by using Mauve alignments. Three local collinear blocks (LCBs) on the chromosomes were identified and joined by connecting lines in the three genomes.
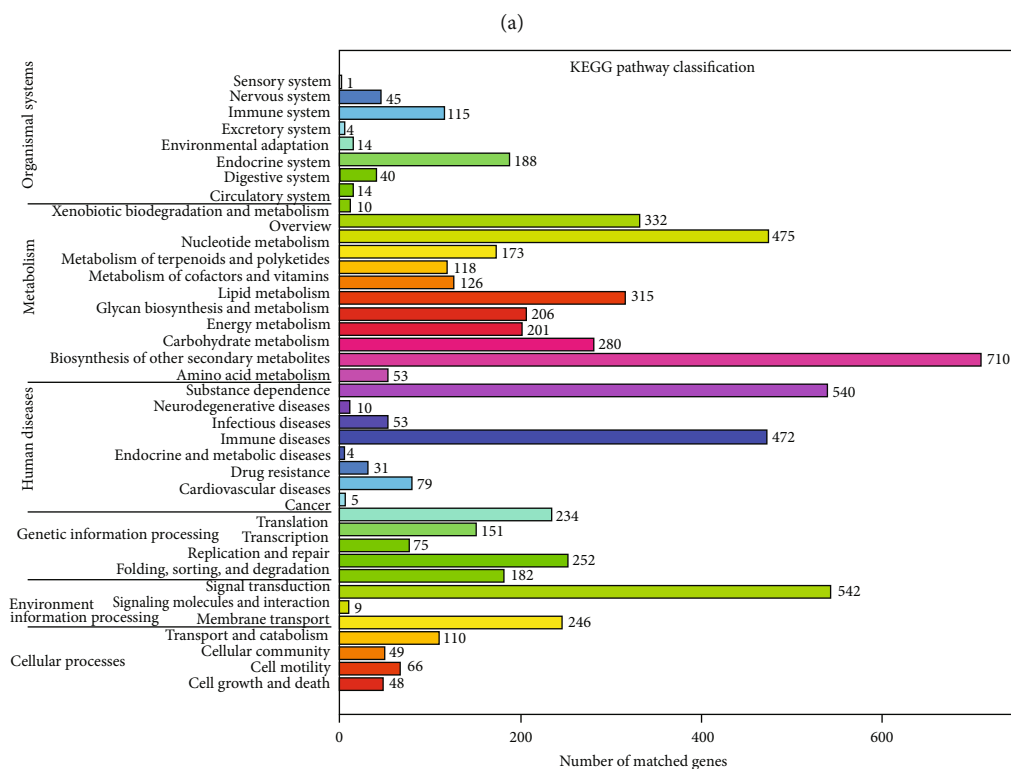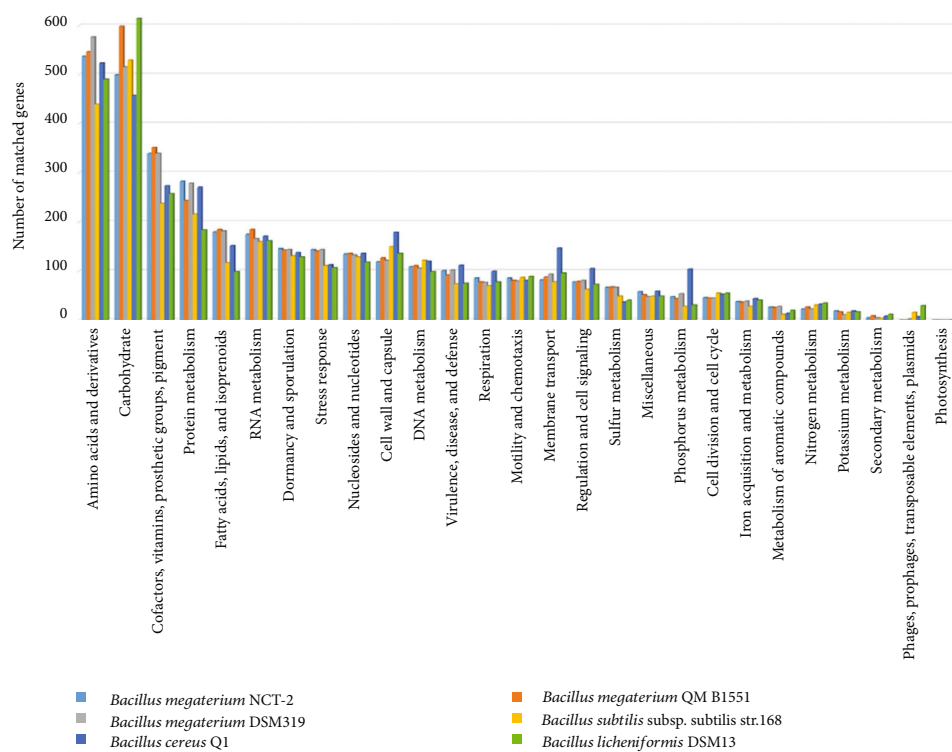
(a)

(b)

FIGURE 3: Distribution of genes based on Cluster of Orthologous Groups (COG) classification and KEGG pathways in *B. megaterium* NCT-2. Only genes assigned to the COG and KEGG categories were used for analysis. (a) Functional categorization of *B. megaterium* NCT-2 compared with 5 related *Bacillus* strains based on the COG database. (b) KEGG classification of *B. megaterium* NCT-2.

nitrate/nitrite sensor protein (NaNiS) and nitrate/nitrite transporter (NaNiT) for sensing and transporting the NO_3^- and NO_2^-. In the process of nitrate and nitrite ammonification,

assimilatory nitrate reductase (NaRas) and nitrite reductase (NiRas) catalyzed the reduction of nitrate to ammonia through nitrite [48]. Then, ammonia was assimilated into
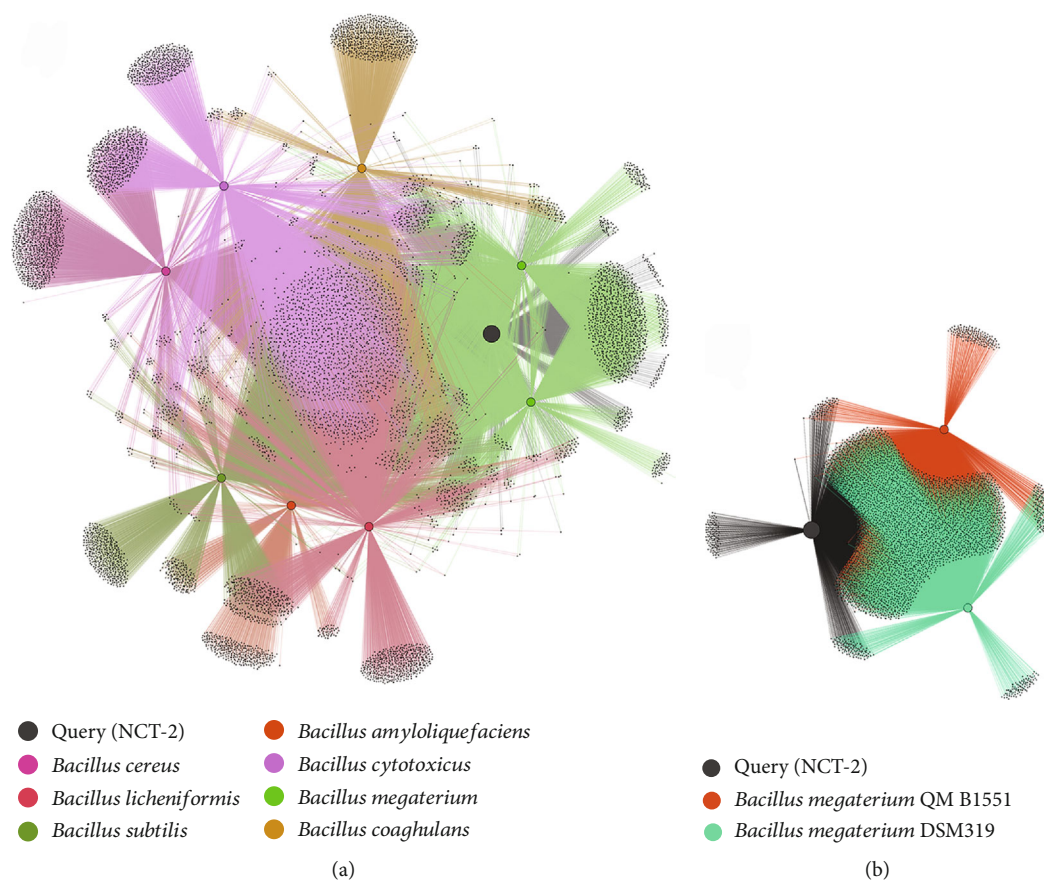
FIGURE 4: Functional similarity networks analysis by fusion+. (a) Functional similarity network among 9 strains of *Bacillus*. (b) Functional similarity network among 3 strains of *Bacillus megaterium*. The networks contain nodes of organisms and functions. The colored organism nodes are connected to black function nodes.

amino acids through L-Glutamine and L-Glutamate by glutamine synthetase type I (GSI), Ferredoxin-dependent glutamate synthase (GOGATF), glutamate synthase [NADPH] large chain (GOGDP1), and glutamate synthase [NADPH] small chain (GOGDP2). Ammonium transporter (Amt) was also encoded in the genome. Ammonium is an important nitrogen source for plant growth. Environmental NH_4^+/NH_3 was imported across membranes by Amt for cell growth in prokaryotes and plants [49]. Bacterial Amt proteins act as passive channels for the uncharged gas ammonia (NH_3) [50]. It means that *B. megaterium* NCT-2 might scavenge NH_4^+/NH_3 in soil instead of providing. In the face of nitrosative stress, genes encoding nitrite-sensitive transcriptional repressor (NsrR) were found in both the chromosome and the plasmid. NsrR played a pivotal role in the regulation of NirK (nitrite reductase), which was expressed aerobically in response to the increasing concentration of NO_2^- and decreasing pH [51]. However, no functional NirK could be found. Instead, two nitric oxide reductase activation proteins (NorD and NorQ) for denitrifying reductase gene clusters were found but without nitric oxide reductase, making the function of denitrification highly unlikely. Thus, the genome analysis proposed that *B. megaterium* NCT-2 could convert nitrate from secondary salinization soil into biomass

through glutamate rather than reduce nitrate to nitrous oxide or dinitrogen, which are lost from the soil (Figure 5). It is an effective bioremediation approach to remove nitrate from soils.

3.6. Genes Associated with Plant Growth-Promoting Features. Our previous studies on the plant growth promotion of *B. megaterium* NCT-2 revealed that it could produce organic acids (lactic acid, acetic acid, propionic acid, and gluconic acid) and phosphatase in culture medium, showing significant phosphate-solubilizing ability [5]. Inoculation with *B. megaterium* NCT-2 significantly increased the root fresh weight of maize [6]. The genome of NCT-2 contains genes encoding for glucose 1-dehydrogenase (EC 1.1.1.47) and alkaline phosphatase (EC 3.1.3.1). Glucose dehydrogenase can oxidize glucose to gluconic acid, which is the most frequent organic acid produced by phosphate-solubilizing bacteria [52]. Additionally, the phosphate starvation system for phosphate uptake encoded by *pst*S, *pst*C, *pst*A, and *pst*B was also found in the genome. The phosphate solubilization capacity of strain NCT-2 plays a positive role in promoting plant growth by dissolving unavailable P (PO_4^{3-}) in soil to plant available forms.

Many plant growth-promoting bacteria have the ability to synthesize plant auxins (indole-3-acetic acid, IAA)
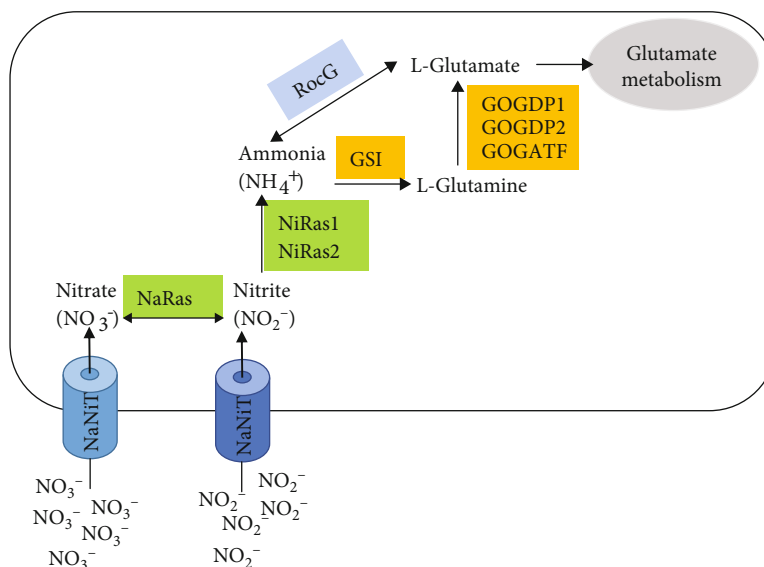
FIGURE 5: The proposed nitrogen metabolism in *B. megaterium* NCT-2. NaNiT: nitrate/nitrite transporter; NaRas: assimilatory nitrate reductase large subunit (EC 1.7.99.4); NiRas1: nitrite reductase [NAD(P)H] large subunit (EC 1.7.1.4); NiRas2: nitrite reductase [NAD(P)H] small subunit (EC 1.7.1.4); GSI: glutamine synthetase type I (EC 6.3.1.2); GOGDP1: glutamate synthase [NADPH] large chain (EC 1.4.1.13); GOGDP2: glutamate synthase [NADPH] small chain (EC 1.4.1.13); GOGATF: Ferredoxin-dependent glutamate synthase (EC 1.4.7.1); RocG: NAD-specific glutamate dehydrogenase (EC 1.4.1.2).

[53, 54], which is a key regulator for plant growth and development, such as cell division and elongation, lateral root production, and flowering [55]. Large-scale genomic analysis of IAA synthesis pathways suggested that plenty of bacteria could synthesize IAA via multiple incomplete pathways, and Firmicutes genomes had the simplest Trp-dependent IAA synthetic system [56]. According to the KEGG analysis, strain NCT-2 could assimilate tryptophan (Trp) (Figure S7) but had incomplete Trp-dependent IAA synthesis pathways, such as the indole-3-acetamide (IAM) pathway and indole-3-pyruvate (IPA) pathway (Figure S8). It had aldehyde dehydrogenase (NAD$^+$) (EC 1.2.1.3) and amidase (EC 3.5.1.4) catalyzing the final step of IAA synthesis. However, we could not find the enzymes which convert Trp into IAM and IPA. These results suggested that strain NCT-2 might synthesize IAA from intermediates.

Both the phosphate solubilization and IAA synthesis play important roles in plant growth promotion of strain NCT-2 during biocontrol and bioremediation of the secondary salinization soils.

3.7. Genes Involved in Stress Response. B. megaterium NCT-2 showed high salinity adaptation in secondary salinization soil in our previous study [4]. From the genome perspective, we can see genes involved in cation transporters (magnesium transport and copper transport system) and stress response, such as osmotic stress, oxidative stress, and detoxification. Glycine betaine, a very efficient osmoprotectant, can be synthesized or acquired from exogenous sources [57]. There are glycine betaine ABC transport systems (*opu*A, *opu*C, and *opu*D) for choline uptake and genes for the glycine betaine biosynthetic enzymes (choline dehydrogenase, *gbs*B, and betaine-aldehyde dehydrogenase, *gbs*A) in strain

NCT-2 genome. Moreover, the genome contains genes encoding for superoxide dismutase (EC 1.15.1.1), catalase (EC 1.11.1.6), and ferroxidase (EC 1.16.3.1), protecting bacteria from oxidative stress. It implied that NCT-2 has great ability of adaption to environments.

4. Conclusion

A hybrid approach with multiple assembler was used to assemble the complete genome of *B. megaterium* NCT-2. The deeper investigation identified clues associated with the features of the bioremediation of secondary salinization soil and plant growth promotion at the gene level, such as nitrogen metabolism, phosphate uptake, synthesis of organic acids and phosphatase for phosphate-solubilizing ability, and Trp-dependent IAA synthetic system. Furthermore, the genes involved in cation transporters, osmotic stress, and oxidative stress implied that NCT-2 has great ability of adaption to environments. In summary, these results provide valuable genomic resources for further studies and applications of using *B. megaterium* NCT-2 in bioremediation processes of secondary salinization soil.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files. The genome sequence of *B. megaterium* NCT-2 has been deposited in GenBank. The accession number for the *B. megaterium* NCT-2 chromosome is CP032527.2, and those for ten plasmids are CP032528.1, CP032529.1, CP032530.1, CP032531.1, CP032532.1, CP032533.1, CP032534.1, CP032535.1, CP032536.1, CP032537.1.

Disclosure

The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

Xiaorui Liu and Pei Zhou conceptualized the study. Bin Wang and Xiaorui Liu performed formal analysis. Dan Zhang and Shaohua Chu took care of funding acquisition. Bin Wang, Dan Zhang, and Shaohua Chu performed the investigation. Yuee Zhi and Xiaorui Liu performed methodology. Pei Zhou acquired resources. Yuee Zhi and Pei Zhou performed supervision of the study. Bin Wang wrote the original draft. Dan Zhang and Xiaorui Liu reviewed and edited the manuscript.

Acknowledgments

Supplementary Materials

Supplementary 1. Figure S1: whole genome sequencing and assembly workflow. Supplementary 2. Figure S2: circular representation of the ten plasmids of *B. megaterium* NCT-2. Supplementary 3. Figure S3: genome similarity of strain NCT-2. The genome of strain NCT-2 was submitted to the web service RAST and was compared with genomes of other strains. A higher comparison score means higher similarity. Supplementary 4. Figure S4: histogram of GO classifications. The results are summarized in three categories: biological process (blue), cellular component (brown), and molecular function (orange). Supplementary 5. Figure S5: genes connected to subsystems according to functional categories. (a) The subsystems of genes from chromosome. (b) The subsystems of genes from plasmids. Supplementary 6. Figure S6: enzymes involved in the nitrogen metabolism of *B. megaterium* NCT-2 from KEGG. Genes of *B. megaterium* NCT-2 were shown in green boxes. Supplementary 7. Figure S7: enzymes involved in phenylalanine, tyrosine, and tryptophan biosynthesis of *B. megaterium* NCT-2 from KEGG. Genes of *B. megaterium* NCT-2 were shown in green boxes. Supplementary 8. Figure S8: enzymes involved in the tryptophan metabolism of *B. megaterium* NCT-2 from KEGG. Genes of *B. megaterium* NCT-2 were shown in green boxes. Supplementary 9. Table S1. plasmids features of *B. megaterium* NCT-2. Supplementary 10. Table S2. the functional similarities of *B. megaterium* NCT-2 with 1,374 bacterial genomes. Supplementary 11. Table S3. gene cluster involved in nitrogen metabolism of *B. megaterium* NCT-2. *(Supplementary Materials)*

References

[1] Z. H. Wang and S. X. Li, "Effects of nitrogen and phosphorus fertilization on plant growth and nitrate accumulation in vegetables," *Journal of Plant Nutrition*, vol. 27, no. 3, pp. 539–556, 2004.

[2] W. S. Shen, X. G. Lin, W. M. Shi et al., "Higher rates of nitrogen fertilization decrease soil enzyme activities, microbial functional diversity and nitrification capacity in a Chinese polytunnel greenhouse vegetable land," *Plant and Soil*, vol. 337, no. 1-2, pp. 137–150, 2010.

[3] L. Lassaletta, G. Billen, B. Grizzetti, J. Anglade, and J. Garnier, "50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland," *Environmental Research Letters*, vol. 9, no. 10, article 105011, 2014.

[4] S. WeiWei, H. HongYan, W. Nan, Z. YueE, L. QunLu, and Z. Pei, "Characterization of a nitrate-uptake bacterial strain *Bacillus megaterium* NCT-2," *Fresenius Environmental Bulletin*, vol. 22, no. 2, pp. 412–417, 2013.

[5] X. Wei, X. U. Lurong, D. Zhang, Y. Zhi, and P. Zhou, "Phosphate solubilizing characteristics of a nitrate-tolerating bacterium *Bacillus megaterium*," *Acta Scientiae Circumstantiae*, vol. 35, no. 7, pp. 2052–2058, 2015.

[6] S. Chu, D. Zhang, Y. Zhi et al., "Enhanced removal of nitrate in the maize rhizosphere by plant growth-promoting *Bacillus megaterium* NCT-2, and its colonization pattern in response to nitrate," *Chemosphere*, vol. 208, pp. 316–324, 2018.

[7] P. S. Vary, R. Biedendieck, T. Fuerch et al., "*Bacillus megaterium*—from simple soil bacterium to industrial protein production host," *Applied Microbiology and Biotechnology*, vol. 76, no. 5, pp. 957–967, 2007.

[8] S. Hrafnsdottir, J. W. Nichols, and A. K. Menon, "Transbilayer movement of fluorescent phospholipids in *Bacillus megaterium* membrane vesicles," *Biochemistry*, vol. 36, no. 16, pp. 4969–4978, 1997.

[9] G. J. McCool and M. C. Cannon, "PhaC and PhaR are required for polyhydroxyalkanoic acid synthase activity in *Bacillus megaterium*," *Journal of Bacteriology*, vol. 183, no. 14, pp. 4235–4243, 2001.

[10] R. Kittsteiner-Eberle, I. Ogbomo, and H. L. Schmidt, "Biosensing devices for the semi-automated control of dehydrogenase substrates in fermentations," *Biosensors*, vol. 4, no. 2, pp. 75–85, 1989.

[11] T. Nagao, T. Mitamura, X. H. Wang et al., "Cloning, nucleotide sequences, and enzymatic properties of glucose dehydrogenase isozymes from *Bacillus megaterium* IAM1030," *Journal of Bacteriology*, vol. 174, no. 15, pp. 5013–5020, 1992.

[12] K. Suga, Y. Shiba, T. Sorai, S. Shioya, and F. Ishimura, "Reaction kinetics and mechanism of immobilized penicillin acylase from *Bacillus megaterium*," *Annals of the New York Academy of Sciences*, vol. 613, pp. 808–815, 1990.

[13] L. Martin, M. A. Prieto, E. Cortes, and J. L. Garcia, "Cloning and sequencing of the pac gene encoding the penicillin G acylase of *Bacillus megaterium* ATCC 14945," *FEMS Microbiology Letters*, vol. 125, no. 2-3, pp. 287–292, 1995.

[14] W. Panbangred, K. Weeradechapon, S. Udomvaraphant, K. Fujiyama, and V. Meevootisom, "High expression of the penicillin G acylase gene (pac) from *Bacillus megaterium* UN1 in its own pac minus mutant," *Journal of Applied Microbiology*, vol. 89, no. 1, pp. 152–157, 2000.

[15] E. Raux, A. Lanois, M. J. Warren, A. Rambach, and C. Thermes, "Cobalamin (vitamin B12) biosynthesis: identification and characterization of a Bacillus megaterium cobI operon," *Biochemical Journal*, vol. 335, no. 1, pp. 159–166, 1998.

[16] C. Korneli, F. David, R. Biedendieck, D. Jahn, and C. Wittmann, "Getting the big beast to work—systems biotechnology of *Bacillus megaterium* for novel high-value proteins," *Journal of Biotechnology*, vol. 163, no. 2, pp. 87–96, 2013.

[17] "Manipulation of *B. megaterium* growth for efficient 2-KLG production by *K. vulgare*," *Process Biochemistry*, vol. 45, no. 4, pp. 602–606, 2010.

[18] M. Eppinger, B. Bunk, M. A. Johns et al., "Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319," vol. 193, no. 16, pp. 4199–4213, 2011.

[19] L. Liu, Y. Li, J. Zhang et al., "Complete genome sequence of the industrial strain *Bacillus megaterium* WSH-002," *Journal of Bacteriology*, vol. 193, no. 22, pp. 6389-6390, 2011.

[20] W. Shi, W. Lu, Q. Liu, Y. Zhi, and P. Zhou, "The identification of the nitrate assimilation related genes in the novel *Bacillus megaterium* NCT-2 accounts for its ability to use nitrate as its only source of nitrogen," *Functional & Integrative Genomics*, vol. 14, no. 1, pp. 219–227, 2014.

[21] J. Eid, A. Fehr, J. Gray et al., "Real-time DNA sequencing from single polymerase molecules," *Science*, vol. 323, no. 5910, pp. 133–138, 2009.

[22] R. Li, Y. Li, K. Kristiansen, and J. Wang, "SOAP: short oligonucleotide alignment program," *Bioinformatics*, vol. 24, no. 5, pp. 713-714, 2008.

[23] A. McKenna, M. Hanna, E. Banks et al., "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data," *Genome Research*, vol. 20, no. 9, pp. 1297–1303, 2010.

[24] M. Boetzer and W. Pirovano, "SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information," *BMC Bioinformatics*, vol. 15, no. 1, article 211, 2014.

[25] E. W. Myers, G. G. Sutton, A. L. Delcher et al., "A whole-genome assembly of Drosophila," *Science*, vol. 287, no. 5461, pp. 2196–2204, 2000.

[26] A. L. Delcher, K. A. Bratke, E. C. Powers, and S. L. Salzberg, "Identifying bacterial genes and endosymbiont DNA with Glimmer," *Bioinformatics*, vol. 23, no. 6, pp. 673–679, 2007.

[27] G. Benson, "Tandem repeats finder: a program to analyze DNA sequences," *Nucleic Acids Research*, vol. 27, no. 2, pp. 573–580, 1999.

[28] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori, "The KEGG resource for deciphering the genome," *Nucleic Acids Research*, vol. 32, Supplement 1, pp. D277–D280, 2004.

[29] R. K. Aziz, D. Bartels, A. A. Best et al., "The RAST Server: rapid annotations using subsystems technology," *BMC Genomics*, vol. 9, no. 1, p. 75, 2008.

[30] R. Overbeek, R. Olson, G. D. Pusch et al., "The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST)," *Nucleic Acids Research*, vol. 42, no. D1, pp. D206–D214, 2014.

[31] P. Stothard and D. S. Wishart, "Circular genome visualization and exploration using CGView," *Bioinformatics*, vol. 21, no. 4, pp. 537–539, 2005.

[32] A. C. E. Darling, B. Mau, F. R. Blattner, and N. T. Perna, "Mauve: multiple alignment of conserved genomic sequence with rearrangements," *Genome Research*, vol. 14, no. 7, pp. 1394–1403, 2004.

[33] J. Qi, B. Wang, and B. I. Hao, "Whole proteome prokaryote phylogeny without sequence alignment: a K-string composition approach," *Journal of Molecular Evolution*, vol. 58, no. 1, pp. 1–11, 2004.

[34] G. H. Zuo and B. L. Hao, "CVTree3 web server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy," *Genomics, Proteomics & Bioinformatics*, vol. 13, no. 5, pp. 321–331, 2015.

[35] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.

[36] K. Tamura, M. Nei, and S. Kumar, "Prospects for inferring very large phylogenies by using the neighbor-joining method," *Proceedings of the National Academy of Sciences*, vol. 101, no. 30, pp. 11030–11035, 2004.

[37] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Molecular Biology and Evolution*, vol. 33, no. 7, pp. 1870–1874, 2016.

[38] C. Zhu, Y. Mahlich, M. Miller, and Y. Bromberg, "fusionDB: assessing microbial diversity and environmental preferences via functional similarity networks," *Nucleic Acids Research*, vol. 46, no. D1, pp. D535–D541, 2018.

[39] C. Zhu, T. O. Delmont, T. M. Vogel, and Y. Bromberg, "Functional basis of microorganism classification," *PLoS Computational Biology*, vol. 11, no. 8, article e1004472, 2015.

[40] K. T. Konstantinidis and J. M. Tiedje, "Trends between gene content and genome size in prokaryotic species with larger genomes," *Proceedings of the National Academy of Sciences*, vol. 101, no. 9, pp. 3160–3165, 2004.

[41] J. A. Klappenbach, J. M. Dunbar, and T. M. Schmidt, "rRNA operon copy number reflects ecological strategies of bacteria," *Applied and Environmental Microbiology*, vol. 66, no. 4, pp. 1328–1333, 2000.

[42] P. M. Shrestha, M. Noll, and W. Liesack, "Phylogenetic identity, growth-response time and rRNA operon copy number of soil bacteria indicate different stages of community succession," *Environmental Microbiology*, vol. 9, no. 10, pp. 2464–2474, 2007.

[43] C. Andres-Barrao, F. F. Lafi, I. Alam et al., "Complete genome sequence analysis of *Enterobacter* sp SA187, a plant multistress tolerance promoting endophytic bacterium," *Frontiers in Microbiology*, vol. 8, 2017.

[44] K. Yano, T. Wada, S. Suzuki et al., "Multiple rRNA operons are essential for efficient cell growth and sporulation as well as

outgrowth in *Bacillus subtilis*," *Microbiology*, vol. 159, Part 11, pp. 2225–2236, 2013.

[45] B. C. Carlton and D. R. Helinski, "Heterogeneous circular DNA elements in vegetative cultures of Bacillus megaterium," *Proceedings of the National Academy of Sciences*, vol. 64, no. 2, pp. 592–599, 1969.

[46] B. S. Stevenson and T. M. Schmidt, "Growth rate-dependent accumulation of RNA from plasmid-borne rRNA operons in *Escherichia coli*," *Journal of Bacteriology*, vol. 180, no. 7, pp. 1970–1972, 1998.

[47] S. H. Chu, D. Zhang, D. X. Wang, Y. E. Zhi, and P. Zhou, "Heterologous expression and biochemical characterization of assimilatory nitrate and nitrite reductase reveals adaption and potential of *Bacillus megaterium* NCT-2 in secondary salinization soil," *International Journal of Biological Macromolecules*, vol. 101, pp. 1019–1028, 2017.

[48] D. J. Richardson and N. J. Watmough, "Inorganic nitrogen metabolism in bacteria," *Current Opinion in Chemical Biology*, vol. 3, no. 2, pp. 207–219, 1999.

[49] D. Kleiner, "NH4+ transport systems: CRC Press," in *Alkali cation transport systems in prokaryotes*, pp. 379–396, CRC Press, Inc., Boca Raton, Fla, 1993.

[50] E. Soupene, L. He, D. Yan, and S. Kustu, "Ammonia acquisition in enteric bacteria: physiological role of the ammonium/methylammonium transport B (AmtB) protein," *Proceedings of the National Academy of Sciences*, vol. 95, no. 12, pp. 7030–7034, 1998.

[51] H. J. Beaumont, S. I. Lens, W. N. Reijnders, H. V. Westerhoff, and R. J. van Spanning, "Expression of nitrite reductase in Nitrosomonas europaea involves NsrR, a novel nitrite-sensitive transcription repressor," *Molecular Microbiology*, vol. 54, no. 1, pp. 148–158, 2004.

[52] N. Oteino, R. D. Lally, S. Kiwanuka et al., "Plant growth promotion induced by phosphate solubilizing endophytic *Pseudomonas* isolates," *Frontiers in Microbiology*, vol. 6, article 745, 2015.

[53] D. Duca, J. Lorv, C. L. Patten, D. Rose, and B. R. Glick, "Indole-3-acetic acid in plant-microbe interactions," *Antonie Van Leeuwenhoek International Journal of General and Molecular*, vol. 106, no. 1, pp. 85–125, 2014.

[54] O. S. Olanrewaju, B. R. Glick, and O. O. Babalola, "Mechanisms of action of plant growth promoting bacteria," *World Journal of Microbiology and Biotechnology*, vol. 33, no. 11, article 197, 2017.

[55] A. W. Woodward and B. Bartel, "Auxin: regulation, action, and interaction," *Annals of Botany*, vol. 95, no. 5, pp. 707–735, 2005.

[56] P. Zhang, T. Jin, S. Kumar Sahu et al., "The distribution of tryptophan-dependent indole-3-acetic acid synthesis pathways in bacteria unraveled by large-scale genomic analysis," *Molecules*, vol. 24, no. 7, article 1411, 2019.

[57] B. Kempf and E. Bremer, "Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments," *Archives of Microbiology*, vol. 170, no. 5, pp. 319–330, 1998.