Complete Genome Sequence of *Pantoea ananatis* Strain NN08200, an Endophytic Bacterium Isolated from Sugarcane

Quan Zeng[1] · GuoYing Shi[1] · ZeMei Nong[1] · XueLian Ye[1] · ChunJin Hu[1]

Abstract

Stain NN08200 was isolated from the surface-sterilized stem of sugarcane grown in Guangxi province of China. The stain was Gram-negative, facultative anaerobic, non-spore-forming bacteria. The complete genome SNP-based phylogenetic analysis indicate that NN08200 is a member of the genus *Pantoea ananatis*. Here, we summarize the features of strain NN08200 and describe its complete genome. The genome contains a chromosome and two plasmids, in total 5,176,640 nucleotides with 54.76% GC content. The chromosome genome contains 4598 protein-coding genes, and 135 ncRNA genes, including 22 rRNA genes, 78 tRNA genes and 35 sRNA genes, the plasmid 1 contains 149 protein-coding genes and the plasmid 2 contains 308 protein-coding genes. We identified 130 tandem repeats, 101 transposon genes, and 16 predicted genomic islands on the chromosome. We found an indole pyruvate decarboxylase encoding gene which involved in the biosynthesis of the plant hormone indole-3-acetic acid, it may explain the reason why NN08200 stain have growth-promoting effects on sugarcane. Considering the pathogenic potential and its versatility of the species of the genus *Pantoea*, the genome information of the strain NN08200 give us a chance to determine the genetic background of interactions between endophytic enterobacteria and plants.

Introduction

The genus *Pantoea* comprises several species that are associated with plants have been found, either as pathogenic or beneficial bacteria to plants [1, 2]. Some of the first identified members of *Pantoea* were plant pathogens, but many studies subsequently indicated that *Pantoea* exist in a multitude of environments and most of them do beneficial to bioremediation and plant growth [3–5]. There are many *Pantoea* strains isolated from plants, soil and environment and are currently being explored for agricultural applications [6, 7]. Approximately, 20 *Pantoea* species have been identified, having diverse characteristics [8]. The ubiquity, versatility and genetic tractability of *Pantoea* make it ideal for exploring niche specific adaptation and opportunism, and for the development of agricultural and environmental products [9, 10].

To obtain endophytes that have growth-promoting effects on host sugarcane plants and have potential for agricultural application, we attempted to isolate and identify endophytic bacteria associated with sugarcane plants grown in Guangxi Province, the major sugarcane and sugar-producing area of China. Bacterial strain NN08200 was isolated from surface-sterilized stems of a ROC22 sugarcane plant grown in Nanning, Guangxi, China. We had determined the plant growth-promoting potential of strain NN08200 to sugarcane under a greenhouse condition [11]. Moreover, we observed the strain NN08200 colonization at the roots and aerial parts of micropropagated sugarcane plantlets with fluorescence microscopy and confocal microscopy. Sequence determinations and phylogenetic analysis of the 16S rRNA gene indicated that strain NN08200 is affiliated with the genus *Pantoea*, and the strain was preserved in the China General Microbiological Culture Collection Center, with the preservation number CGMCC No. 5438. Here, we present a summary of the features of strain NN08200 and its complete genome sequence, which provides a reference for resolving the phylogeny and taxonomy of closely related strains and

[1] Microbiology Research Institute, Guangxi Academy of Agricultural Sciences, Nanning 530007, People's Republic of China

genetic information to study the plant growth-promoting potential and plant-associated lifestyle of strain NN08200.

Organism Information

Classification and General Features

Strain NN08200 is a Gram-negative, non-spore-forming, motile rod with peritrichous flagella (Fig. 1). This bacterium was able to grow in anaerobic using cooked meat medium with thermal melting vaseline and aerobic using beef extract medium, and grew optimally between 28 and 32 °C (Table 1). It forms circular, convex, smooth colonies on nutrient agar; in addition, it grows well on Ashby nitrogen-free culture medium, showing round, transparent colonies. Strain NN08200 is a species of *Pantoea*, showing several differences from the *Pantoea* species described so far. The strain is an endophyte from sugarcane. It is positive for indole production, nitrate reduction and arginine decarboxylase and lysine decarboxylase activity.

A PHYML method phylogenetic tree based on SNP of complete genomes for strain belonging to the genus Pantoea constructed by TreeBeST (Fig. 2) showed that strain NN08200 is most closely related to strains belonging to

the *Pantoea ananatis* [20]. Genomes gene sequences from the following strains were used to construct the phylogenetic tree: *P. sesame* Si-M154, taxonomy ID: 1881110; *P. ananatis* LMG 20103, taxonomy ID: 706191; *P. ananatis* AJ13355, taxonomy ID: 932677; *P. ananatis* R100, taxonomy ID:; *P. ananatis* PA13, taxonomy ID: 1095774; *P. allii* LMG_24248, taxonomy ID: 574096; *P. stewartii subsp_indologenes* LMG 2632, taxonomy ID: 66270; *P. agglomerans* Eh318, taxonomy ID: 1408177; *P. septica* LMG 5345, taxonomy ID: 472695; *P. rwandensis* ND04, taxonomy ID: 1076550; *P. eucrina* LMG 5346, taxonomy ID: 472693; *P. wallisii* LMG 26277, taxonomy ID: 1076551; *P. cypripedii* LMG 2657, taxonomy ID: 55209; *P. alhagi* LTYR-11Z, taxonomy ID: 1891675; *1* 342, taxonomy ID: 1465635.

Genome Sequencing Information Genome Project History

Pantoea ananatis strain NN08200 was selected for sequencing based on its taxonomic significance and because it could be used in promoting plant growth. The genome sequence is deposited in GenBank with the accession number CP035034. Information about the genome sequencing and its association with MIGS version 2.0 compliance is shown in Table 2.

Growth Conditions and DNA Isolation

P. ananatis strain NN08200 was grown in liquid Luria–Bertani medium at 28 °C until stationary phase. Genomic DNA was extracted using a TIANamp bacterial DNA kit (Tiangen Biotech, Beijing, China). The quantity and quality of DNA were assessed using a NanoDrop spectrophotometer (Thermo Scientific, USA).

Genome Sequencing and Assembly

The genomic DNA of *P. ananatis* strain NN08200 was first constructed into a 10-kb SMRT Bell library and sequenced using the PacBio RS II sequencing system. Low-quality reads were filtered by the SMRT portal (version 2.3.0) and the filtered reads were assembled to generate five contigs containing 5,176,640 bases [21, 22]. The final assembly of the genome provided an average of 166-fold coverage. The five contigs were scaffolding to three circular sequences. The fully assembled *P. ananatis* strain NN08200 genome is composed of a 4.7-M base pair chromosome, and two plasmids, whose sizes were 125k and 307k base pairs, respectively.
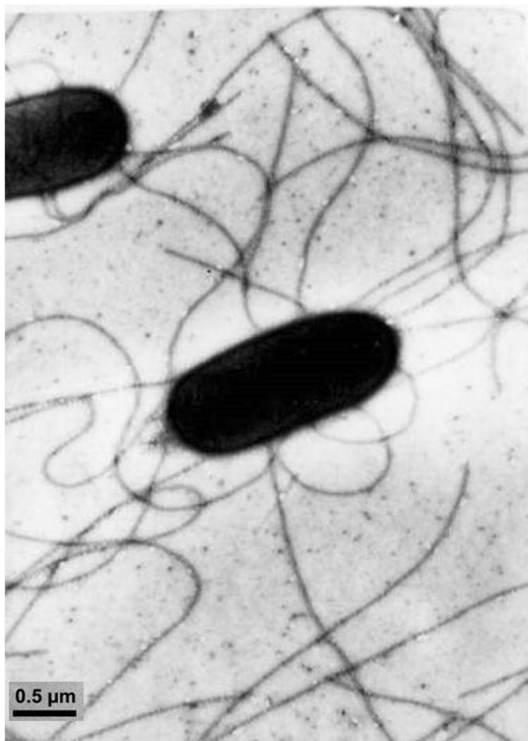

Fig. 1 Transmission electron microphotograph of the *Pantoea ananatis* strain NN08200

Table 1 Classification and general features of *Pantoea ananatis* strain NN08200

MIGS ID	Property	Term	Evidence code
	Current	Domain *Bacteria*	TAS [12]
	Classification	Phylum *Proteobacteria*	TAS [13]
		Class *Gammaproteobacteria*	TAS [14–17]
		Order *Enterobacterales*	TAS [18]
		Family *Erwiniaceae*	TAS [19]
		Genus *Pantoea*	TAS [11]
		Species *Pantoea ananatis*	TAS [11]
		Strain: NN08200	TAS [11]
	Gram strain	Negative	TAS [11]
	Cell shape	Rod	TAS [11]
	Motility	Motile	TAS [11]
	Sporulation	Non-sporulating	TAS [11]
	Temperature range	Mesophile	TAS [11]
	Optimum temperature	28–30℃	IDA
	Carbon source	Sucrose, flucose, fructose, galactose, maltose	IDA
	Energy source	Chemoorganotroph	IDA
MIGS-6	Habitat	Soil, plants	TAS [11]
MIGS-6.3	Salinity	0–4% NaCl	IDA
MIGS-22	Oxygen	Anaerobic and aerobic	TAS [11]
MIGS-23	Isolation	Stem of sugarcane cultivar GT22	TAS [11]
MIGS-15	Biotic relationship	Free-living, endophytic	TAS [11]
MIGS-14	Pathogenicity	No reported	
MIGS-4	Geographic location	Nanning, Guangxi, China	TAS [11]
MIGS-5	Sample collection time	2008	TAS [11]
MIGS-4.1	Longitude	107.37	NAS
MIGS-4.2	Latitude	22.40	NAS
MIGS-4.3	Depth	0.3–0.5 m above the surface	IDA
MIGS-4.4	Altitude	123 m	NAS

Evidence codes: IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence)
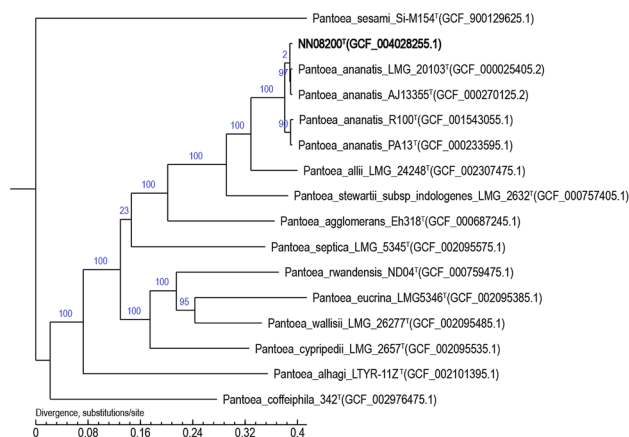
Fig. 2 Phylogenetic tree based on the genome sequences showing the phylogenetic position of strain NN08200 and other strains belonging to the genus *Pantoea*. A PHYML method was beedn used to build the phylogenetic tree based on SNP of complete genomes for strain belonging to the genus Pantoea constructed by TreeBeST

Table 2 Genome sequencing project information for *Pantoea ananatis* NN08200

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	10 kb SMRT Bell library
MIGS-29	Sequencing platforms	PacBio RS II
MIGS-31.2	Fold coverage	166×
MIGS-30	Assemblers	SMRT portal
MIGS-32	Gene calling method	GeneMarkS
	Genome database release	Genbank
	Genbank ID	CP035034
	Genbank date of release	Jan 17,2019
MIGS-13	Source material identifier	NN08200
	Project relevance	Taxonomy, biotechnology

Table 3 Nucleotide content and gene count levels of the *P. ananatis* NN08200 genome

Attribute	Value	% of total
Size (bp)	5,176,640	100.00
G+C content (bp)	2,834,728	54.76
Coding region (bp)	4,480,173	86.55
Total genes	4733	100.00
RNA genes	135	2.85
Protein-coding genes	4598	97.15
Genes assigned to COGs	4369	86.43
Genes with signal peptides	412	8.96
Genes with transmembrane helices	341	7.42
Chromosome size (bp)	4,743,568	91.6%
Chromosome G+C content (bp)	2,552,158	53.8%
Plasmid 1 size (bp)	125,402	2.4%
Plasmid 1 G+C content (bp)	70,096	56.47%
Plasmid 2 size (bp)	307,670	7.1%
Plasmid 2 G+C (bp)	160,525	52.17%

Genome Annotation

The complete sequence of *P. ananatis* strain NN08200 was analyzed using GeneMarkS (version 4.17) to retrieve protein coding genes [23]. Transfer RNA (tRNA) genes were predicted by tRNAscan-SE [24]. Ribosomal RNA (rRNA) genes were analyzed by rRNAmmer [25]. Transposon PSI was used to predict transposons based on the homologous blast method. RepeatMasker (version open-4.0.5) and TRF (tandem repeats finder, version 4.07b) were used for identification of interspersed nuclear elements and tandem repeats, respectively [26, 27]. SlandPath-DIOMB (version 0.2) was used for identification of genomic islands [28].

Genome Properties

The genome of strain NN08200 contains a single chromosome of 4,743,568 nucleotides with 53.8% G+C content and two plasmids, one of 125,402 nucleotides with 56.47% G+C content and another of 307,670 nucleotides with 52.17% G+C content. The chromosome contains 4733

predicted genes: 4598 protein-coding genes and 135 RNA genes including 78 tRNA genes, 35 sRNA genes, and 22 rRNA genes (Table 3; Fig. 3). The plasmid 1 contains 149 protein-coding genes and the plasmid 2 contains 308 protein-coding genes. Ciros was used to show the genome and the result of gene function annotation [29]. In total, 4369 genes were assigned in Clusters of Orthologous Groups of proteins (COG) functional categories and they are listed in Table 4.

Insights from the Genome

Here we present the complete genome sequence of *Pantoea ananatis* strain NN08200. Protein-coding sequences accounted for 4598(97.15%) of the total of 4733 genes identified. 54 complete genomes of P. ananatis have been download from NCBI to performed an Average Nucleotide Identity (ANI) analysis with strain NN08200 [30]. The results justified the conclusion of phylogenetic analysis, strain NN08200 with other strains resulted in a high ANI (> 95%). The results suggested that the strain NNo8200 belongs to the P. ananatis. NN08200 and P. ananatis S8 resulted in the highest ANI (99.2%) and show that they are similar than other strains.

Conclusion

In this study, we present the complete genome sequence of *Pantoea ananatis* strain NN08200, an endophyte from sugarcane. The genome of *P. ananatis* NN08200 consists of a 4,743,568-bp long chromosome, containing 4598 protein coding genes. *P. ananatis* NN08200 also contains two plasmids. To analyze the complete genome sequence of *Pantoea ananatis* strain NN08200, we found an indole pyruvate decarboxylase encoding gene which involved in the biosynthesis of the plant hormone indole-3-acetic acid [31], it may promote plant growth by improving the synthesis of indoleacetic acid. The new genomic data will facilitate future applications of this strain in agricultural production.
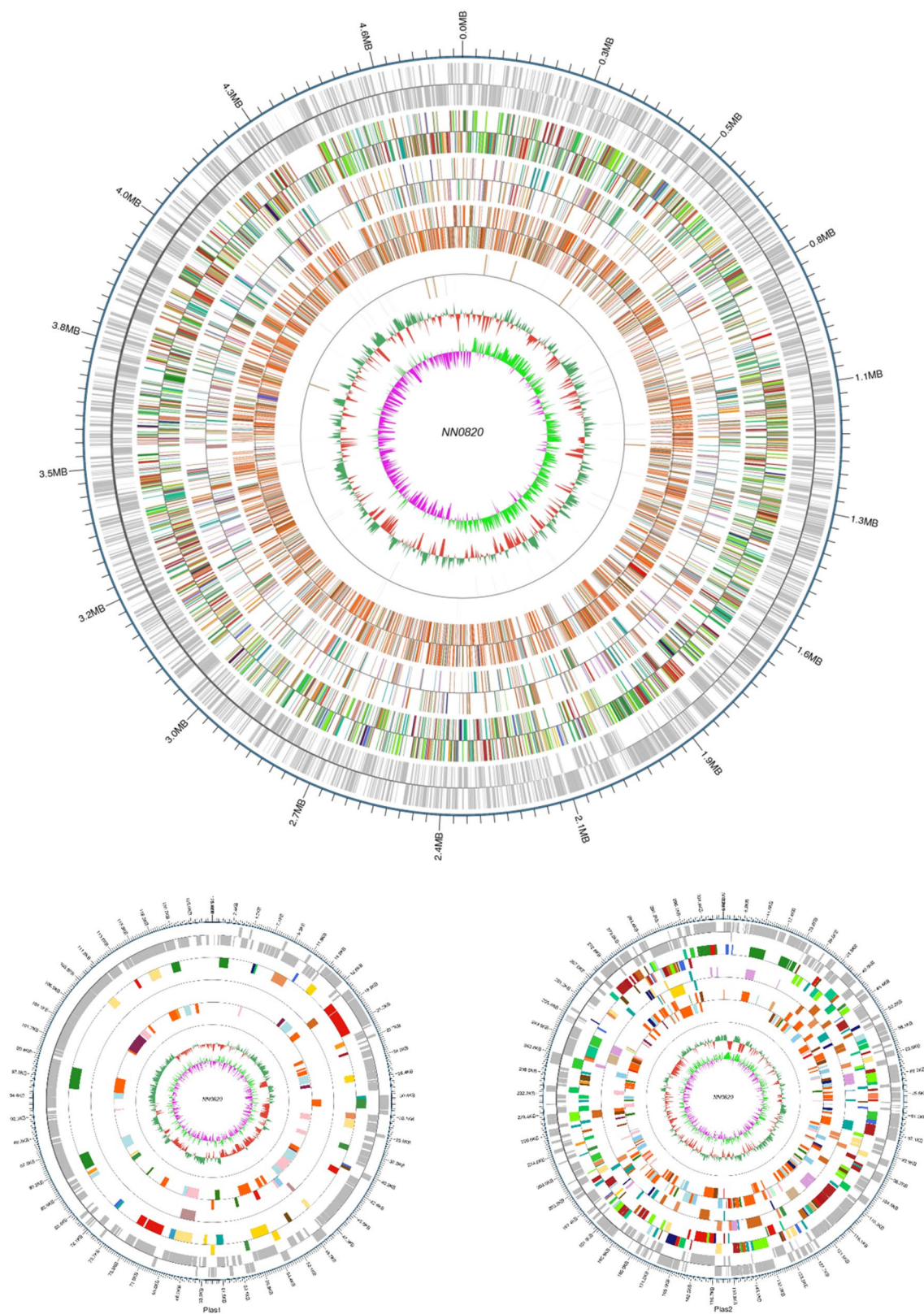
Fig. 3 Graphical circular map of the chromosome and plasmids of Pantoea ananatisNN08200 by Circos. From outside to the center: Coding genes on forward andreverse strands, the results of gene function annotation (including genes, COG,KEGG, GO), ncRNAs

Table 4 Number of genes associated with the 25 general COG functional categories

Code	Value	% of total	Description
A	1	0.02	RNA processing and modification
C	191	3.78	Energy production and conversion
D	52	1.03	Cell cycle control, cell division, chromosome partitioning
E	418	8.27	Amino acid transport and metabolism
F	105	2.08	Nucleotide transport and metabolism
G	451	8.92	Carbohydrate transport and metabolism
H	200	3.96	Coenzyme transport and metabolism
I	150	2.97	Lipid transport and metabolism
J	269	5.32	Translation, ribosomal structure and biogenesis
K	351	6.94	Transcription
L	162	3.2	Replication, recombination and repair
M	287	5.68	Cell wall/membrane/envelope biogenesis
N	133	2.63	Cell motility
O	146	2.89	Posttranslational modification, protein turnover, chaperones
P	256	5.06	Inorganic ion transport and metabolism
Q	89	1.76	Secondary metabolites biosynthesis, transport and catabolism
R	352	6.96	General function prediction only
S	226	4.47	Function unknown
T	256	5.06	Signal transduction mechanisms
U	98	1.94	Intracellular trafficking, secretion, and vesicular transport
V	90	1.78	Defense mechanisms
W	30	0.59	Extracellular structures
X	56	1.1	Mobilome: prophages, transposons
–	686	13.57	Not in COGs

References

1. Feng YJ, Shen DL, Dong XZ et al (2003) In vitro symplasmata formation in the rice diazotrophic endophyte Pantoea agglomerans YS19. Plant Soil 255(2):435–444
2. Brady C, Cleenwerck I, Venter S et al (2008) Phylogeny and identiication of Pantoea species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). Syst Appl Microbiol 31:447–460
3. Choi Q, Kim H, Lee Y et al (2012) First report of sheathrot of rice caused by Pantoea ananatis in Korea. Plant Pathol J 28(3):331
4. Zhang J, Zhang CW, Yang J et al (2019) Insights into endophytic bacterial community structures of seeds among various *Oryza sativa* L. rice genotypes. J Plant Growth Regul 38(1):93–102
5. Volksch B, Thon S, Jacobsen I et al (2009) Polyphasic study of plant and clinic-associated *Pantoea agglomerans* strains reveals indistinguishable virulence potential. Infect Genet Evol 9:1381–1391
6. Nadarasah G, Stavrinides J (2014) Quantitative evaluation of the host colonizing capabilities of the enteric bacterium Pantoea using plant and insect hosts. Microbiology 160:602–615
7. Selvakumar G, Kundu S, Joshi P et al (2008) Characterization of a cold-tolerant plant growth-promoting bacterium Pantoea dispersa 1A isolated from a sub-alpine soil in the North Western Indian Himalayas. World J Microbiol Biotechnol 24(7):955–960
8. Guo HB, He SW, Wang X et al (2017) Phylogenetic diversity and plant growth-promoting characteristics of endophytic Pantoea spp. in rice seeds. Acta Microbiol Sin. https://doi.org/10.13343/j.cnki.wsxb.20180581

9. Yoshida A, Kohchi C, Inagawa H et al (2009) Improvement of allergic dermatitis via regulation of the Th1/Th2 immune systembalance bymacrophages activated with lipopolysaccharide derived from Pantoea agglomerans (IP-PA1). Anticancer Res 29:4867–4870

10. Walterson AM, Stavrinides J (2015) Pantoea: insights into a highly versatile and diverse genus within the Enterobacteriaceae. FEMS Microbiol Rev 39(6):968–984

11. Shi GY, Zeng Q, Nong ZM et al (2012) Identification of an endophytic nitrogen-fixing bacterium NN08200 from sugarcane and its growth promotion of sugarcane. Microbiol China. https ://doi.org/10.13344/j.microbiol.china.180523

12. Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87:4576–4579

13. Garrity GM, Bell JA, Lilburn T (2005) Phylum XIV. Proteobacteria phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds) Bergey's manual of systematic bacteriology, 2nd edn, vol. 2, Part B. Springer, New York, p 1

14. Garrity GM, Bell JA, Lilburn T et al (2015) Gammaproteobacteria class. nov. Criminal behaviour and mental health

15. Krieg NR, Staley JT, Garrity GM (eds) (2005) Bergey's manual of systematic bacteriology, 2nd edn, vol. 2. Springer, New York, p 1

16. List Editor (2005) Validation of publication of new names and new combinations previously effectively published outside the IJSEM. List no. 106. Int J Syst Evol Microbiol 55:2235–2238.

17. Williams KP, Kelly DP (2013) Proposal for a new class within the phylum Proteobacteria, Acidithiobacillia classis nov., with the type order Acidithiobacillales, and emended description of the class Gammaproteobacteria. Int J Syst Evol Microbiol 63:2901–2906

18. Garrity GM, Holt JG (2001) Taxonomic outline of the archaea and bacteria. In: Garrity GM, Boone DR, Castenholz RW (eds) Bergey's manual of systematic bacteriology, vol 1, 2nd edn. Springer, New York, pp 155–166

19. Palmer M, Steenkamp ET, Coetzee MPA et al (2018) Mixta gen. nov. a new genus in the Erwiniaceae. Int J Syst Evol Microbiol 68(4):1396

20. Tamura K, Peterson DS, Peterson N et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739

21. Konstantin B, Sergey K, Chen-Shan C et al (2015) Assembling large genomes with single-molecule sequencing and locality sensitive hashing. Nat Biotechnol 33:623–630

22. Sergey K, Adam MP (2015) One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. Curr Opin Microbiol 23:110–120

23. Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29(12):2607–2618

24. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25(5):955–964

25. Lagesen K, Hallin P, Rødland EA et al (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35(9):3100–3108

26. Saha S, Bridges S, Magbanua ZV et al (2008) Empirical comparison of ab initio repeat finding programs. Nucleic Acids Res 36(7):2284–2294

27. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27(2):573

28. Hsiao W, Wan I, Jones SJ et al (2003) IslandPath: aiding detection of genomic islands in prokaryotes. Bioinformatics 19(3):418–420

29. Krzywinski M et al (2009) Circos: an Information Aesthetic for Comparative Genomics. Genome Res 19:1639–1645

30. Richter M, Rossellomora R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA 106(45):19126–19131

31. Schutz A, Sandalova T, Ricagno S et al (2003) Crystal structure of thiamindiphosphate-dependent indolepyruvate decarboxylase from Enterobacter cloacae, an enzyme involved in the biosynthesis of the plant hormone indole-3-acetic acid. FEBS J 270(10):2312–2321