



Short communication

Whole genome sequence of *Pantoea agglomerans* ANP8, a salinity and drought stress-resistant bacterium isolated from alfalfa (*Medicago sativa* L.) root nodules



Fatemeh Noori^a, Hassan Etesami^{b,*}, Somayeh Noori^c, Esmail Forouzan^d, Gholamreza Salehi Jouzani^e, Mohammad Ali Malboobi^f

^a Department of Biotechnology and Plant Breeding, Sari Agricultural Sciences and Natural Resources University, Sari, Iran

^b Agriculture & Natural Resources Campus, Faculty of Agricultural Engineering & Technology, Department of Soil Science, University of Tehran, Tehran, 31587-77871, Iran

^c Qom University of Medical Sciences, Qom, Iran

^d GeneMan Genomics Ltd, Iran

^e Microbial Biotechnology Department, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education and Extension Organization (AREEO), Fahmideh Blvd, P.O. Box: 31535-1897, Karaj, Iran

^f Department of Plant Biotechnology, National Institute of Genetic Engineering and Biotechnology, Tehran, 14965-161, Iran

ARTICLE INFO

Article history:

Received 30 July 2020

Received in revised form 7 November 2020

Accepted 4 February 2021

Keywords:

Nodule endophytes

Plant-microbe interaction

Plant growth-promoting bacteria

Abiotic stress

Legume

ABSTRACT

Environmental abiotic stress conditions, especially drought and salinity, are currently the major factors that reduce crop yields worldwide. It has been reported that plant-associated beneficial bacteria, especially strains resistant to abiotic stresses that could maintain their efficiency under environmental challenging conditions, can contribute to alleviate abiotic stresses of host plants. In this study, we presented the assembly of the whole genome of *Pantoea agglomerans* ANP8, a plant growth-promoting bacterium resistant to salinity and drought stresses. The draft genome assembly contained 4,713,172 bp with 4586 predicted genes. A primary draft genome with a total of 5,115,548 bp and 1916 contigs was obtained (longest contig length being 485,272 bp and smallest contig being 112 bp). Following assembly upgrades, 68 scaffolds and 70 contigs with lengths ≥ 500 bp and an N50 = 209,657 bp were obtained. Number of 5554 and 5472 open reading frames longer than 50 codons were observed in the direct strand and in the reverse strand, respectively. Due to the multiple plant growth-promoting characteristics of this bacterium, genes involved in various indole-3-acetic acid production pathways, e.g., indole-3-pyruvic acid and indole-3-acetamide pathways, were found in the bacterium's genome. In addition, multiple copies of the *gcd* gene, most important enzymes involvement in the solubilization of phosphates, glucose dehydrogenase, were also observed in this genome. The study provides new genomic information to help understanding the way of action of a stress-tolerant plant growth-promoting bacterium.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Salinity stress is one of the most important barriers to agricultural production. More than 20% of the arable land around the world is affected by salinity stress and its amount is increasing day by day [1]. Alfalfa (*Medicago sativa* L.) is one of the most important forage plants in the world due to its palatability and together with its protein and mineral composition [2]. In terms of salinity resistance, alfalfa is in the group of semi-resistant plants; as it is known that salinity (electrical conductivity, EC) more than two dS m^{-1} reduces the growth and yield of this legume plant [3].

Soil salinity also reduces nitrogen fixation and nitrogenase activity in legumes [4].

Alfalfa roots are a good environment for the rapid growth of microorganisms around the roots. A high volume of microbiological activity is due to the natural chemical exudates of the root of this plant [5]. Soil salinity not only inhibits plant growth and development, but also has negative effects on the composition and activity of the rhizospheric bacteria [6]. However, salinity-tolerant plant growth-promoting rhizobacteria (PGPR), or ST-PGPR, may significantly increase the growth and yield of crops including alfalfa under salinity stress conditions [1,2,7]. In addition to rhizobia, it is well known that non-rhizobial bacteria are also inhabitants of legume-root nodules [8]. These bacteria are in close contact with the plant and the metabolites produced by them are

* Corresponding author.

E-mail address: hassanetesami@ut.ac.ir (H. Etesami).

better delivered to the host plant [9]. In recent years, more attention has been paid to the combined inoculation of rhizobia with non-rhizobial PGPR to increase nodulation and ultimately plant growth [4]. However, little information is available on drought and salinity-resistant PGPR isolated from the nodules of legume roots [2]. In addition, the true role of the nodule non-rhizobial bacteria (helper bacteria) in the nodules has not yet been determined. *Pantoea agglomerans* is a Gram-negative bacterium that grows symbiotically with various plants and promotes the plant growth [10].

In this report, the genome assembly of the *P. agglomerans* ANP8, isolated from the root nodules of alfalfa plants grown on a saline soil (EC: 10.94 dS m⁻¹) in a field in Qom, Iran (N 34° 43', E 51° 06', and 854 m above sea level) is presented. Genome sequencing of ANP8 strain was performed using De novo strategy and Illumina MiSeq platform. The aim of the research was access to the genome information of a salinity and drought-resistant non-rhizobial PGP rhizobacterium isolated from root-nodules. *P. agglomerans* strain ANP8 was selected among a set of nodule non-rhizobial bacteria due to having some unique features (see below).

Based on a previous standard method [11] and using yeast extract-mannitol agar (YMA) medium supplemented with congo red and nutrient agar (NA) medium, the nodule endophytic bacteria (rhizobial and non-rhizobial bacteria) were isolated and purified. To differentiate the symbiotic isolates with alfalfa plant from non-symbiotic isolates (non-rhizobial bacteria), the nodule formation test for each bacterial isolate was checked by inoculation of alfalfa seedlings [11]. Based on a series of assays (e.g., screening for PGP traits, resistance to salinity and drought, and plant growth stimulation potential in the presence of salinity), the non-rhizobial strain ANP8 was selected as an effective strain among other non-rhizobial strains. This strain was able to grow in NA medium supplemented with 11.6% NaCl and in NB (nutrient broth) medium having osmotic potential of -30 bar.

The growth optical density (OD₆₀₀) of this strain reached 1.0 in NB medium after 6 h of incubation at 28 ± 2 °C and 200 r.p.m. The growth rate of this strain was considerable compared to the growth rate of *E. coli* bacterium as the fastest growing bacterium (OD₆₀₀ = 1.0 after 4 h of incubation). *P. agglomerans* ANP8 could also solubilize mineral and organic phosphates, was able to fix N₂, and produced indole-3-acetic acid (IAA) (Data not shown). The remarkably number of positive traits in this strain motivated us to approach its full genome sequencing. DNA was extracted by QIAGEN genomic DNA extraction kit following the manufacturer's instructions and sent to Bji company, China, for whole genome sequencing using De novo strategy and the Illumina MiSeq platform.

The whole genome sequencing of the strain ANP8 resulted in 1GB raw data with 7.8 million paired-end reads with an average

length of 300 bp from the paired-end library. To ensure the quality of reads, raw data were first controlled using FastQC v.0.11.9. Then, based on the results of the previous step, low-quality reads/bases were filtered using Trimmomatic version 0.36, if necessary. The criterion for low quality reads is a Phred quality score lower than 20, which is equivalent to a 1% error probability. After merging the pairs of reads using Flash 1.2.11, the primary assembly was performed on high-quality filtered reads with a total of 6,817,112 reads using SPAdes 3.13.0. Afterward, a primary draft genome with a total of 5,115,548 bp and 1916 contigs was obtained (longest contig length being 485,272 bp and smallest contig being 112 bp). Then, using FastANI 0.1.2, the average nucleotide similarity between this genome and all genomes of the same genus in the gene database were calculated. Accordingly, the genomes with a genomic similarity more than 98% were used to increase assembly quality. Then, using the SPAdes 3.13.0, the assembly was upgraded as much as possible and the best possible assembly was obtained. Following assembly upgrades, 4,713,172 bp in 68 scaffolds and 70 contigs with lengths ≥ 500 bp and an N50 = 209,657 bp were obtained. Average coverage depth was 338X, with more than 95% of the read pairs (a total of 3.5 million read pairs) aligned on it correctly. The GC content of the genome was 55.05%. This shows that the assembly is also of good quality in terms of accuracy. BUSCO (v1.1) and QUAST (v2.2) assessment tools were used to control assembly quality. About 98% of the genes predicted by the BUSCO which most likely should be present in the genome were also present in the assembly. This means that the assembly is of acceptable quality in terms of completeness. A summary of the results of assembly in numbers is presented in Table 1.

Genome annotation was performed using Prokka 1.14.1 and 4586 genes were predicted, which include 4401 CDS (coding sequences) and 103 misc_RNA, 81 tRNA, and 1 tmRNA. Out of a total of 4401 coding DNA sequences (CDS), 1574 genes had predicted functions and 1308 were hypothetical proteins. A summary of the annotation results is shown in Table 2.

As mentioned above, strain ANP8 was positive for several PGP traits and was also resistant to salinity and drought stress. Genome annotation was able to identify multiple gene clusters associated with PGP traits and the resistance to salinity and drought stress. Some of these genes include the *ipdC* gene, which produces the enzyme indole-3-pyruvate decarboxylase, a key enzyme in the indole-3-pyruvic acid (IPyA) pathway for the production of indole-3-acetic acid (IAA). The *laaH* gene is also active in the indole-3-acetamide (IAM) pathway, the most common pathway known to produce IAA in bacteria. This bacterium has the *aam* gene, which is a homologous gene to *iaaM* gene and is involved in the IAM pathway. Strain ANP8 also contains the Trp operon genes (*trpA*, *B*, *C*, *E*, *S*, *R*, and *GD* genes), which are involved in producing IAA.

Table 1
The results of assembly quality control.

Assembly	Assembly output (scaffolds)	Assembly_output_broken (contigs)
# contigs	68	70
Largest contig	485,272	385,906
Total length	4,713,172	4,712,978
GC (%)	55.05	55.05
N50	209,657	195,423
# total reads	6,817,112	6,817,167
Mapped (%)	99.61	99.61
Properly paired (%)	96.42	96.42
Avg. coverage depth	338	338
# N's per 100 kbp	4.12	0
Complete BUSCO (%)	97.97	97.97
Partial BUSCO (%)	0	0

*Statistics are for contigs and scaffolds with length of 500 bp and more.

Table 2
Genome annotation results.

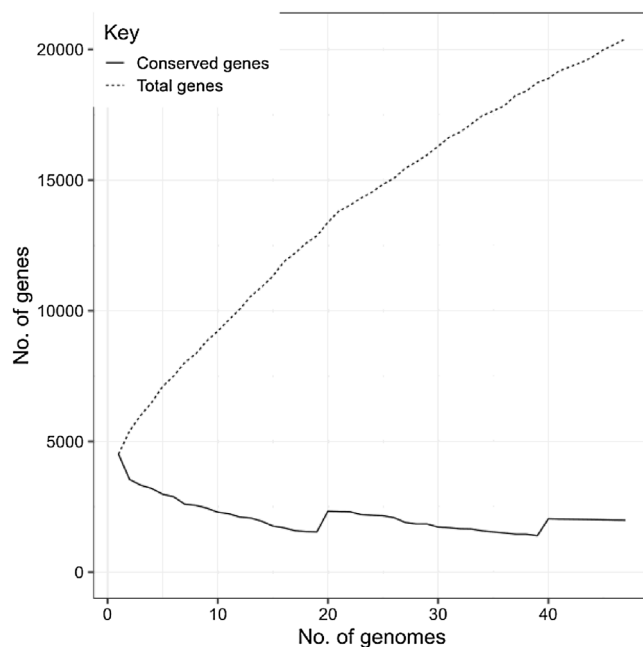
Value	Features
1916	Contigs
5,115,548	Bases
4586	Genes
4401	CDS
1547	Translated products with predicted functions
1308	Hypothetical protein
103	misc_RNA
81	tRNA
1	tmRNA
25	rDNA

Genome analysis showed that strain ANP8 has three copies of the *gcd* gene, which is responsible for producing the most important enzymes involved in the solubilization of phosphates, e.g., glucose dehydrogenase. This bacterial strain had a very high ability to solubilize mineral phosphates (data not shown) and the multiplicity of *gcd* genes in the genome of this bacterium confirmed this phenotype. Also, based on the necessity of pyrroloquinolone quinone cofactor for glucose dehydrogenase activity, it was found that strain ANP8 contains *pqqB*, *C*, *D*, and *E* genes from *pqq* gene cluster, which are responsible for synthesizing the cofactor that is necessary for glucose dehydrogenase activity.

Some of the genes involved in regulating potassium concentrations, which are associated with salinity tolerance, include: *nhaA*. This gene is needed to tolerate sodium and potassium and to express resistance to alkaline pH in the presence of Na^+ . NhaK belongs to the CPA1 antiporter family, which is involved in the flow of Na^+ , K^+ , Li^+ , and Rb^+ . The antiporter activity K^+/H^+ is expressed under alkaline pH conditions. Strain ANP8 is equipped with the NhaK, NhaP and NhaA antiporter systems. The presence of such several antiporter systems is in agreement with the ability of this bacterium to correct different ionic unbalances under stressful conditions. Strain ANP8 genome analysis showed that this strain has a *ChaA* gene in its genome, which encodes Na^+ (Ca^{2+})/ H^+ antiporter and is also effective in potassium balance. *ybaL* encodes a K^+/H^+ flow pump, part of the large family of cation/proton (CPA-2) antiporters in Gram-positive bacteria, Gram-negative bacteria, and other organisms. The KdpFABC protein complex, which is present in strain ANP8, imports potassium into the cell, and is one of the few operons used to facilitate K^+ uptake. The Trk system is made up of a transmembrane protein called TrkH, or TrkG, which is a K^+ transfer subunit. Strain ANP8 has the essential *TrkA* and *TrkH* genes and has the ability to accumulate potassium. This bacterium has also the *kup* gene encoding potassium transport system.

In order to perform a core- and pangenome analysis, the Roary 3.11.2 software was used. In this analysis, 46 genomes were used with a set of 20,402 genes in the strain level. Comparison of translated CDS collections and strain-specific CDSs were determined throughout the pangenome. The pangenome diagram (Fig. 1) shows the variation of the core genes and the pan genes (all genes except for the core genes) based on the number of genomes present in the analysis. It was found that the number of core genes decreased with the addition of more genomes to the analysis and the Pan genes increased. This shows that this species has an open pangenome. It means that the more genomes of this species are examined, the more (new) genes are found in at least one of the existing genomes.

Results also showed that, there were: 532 genes in the core group present in all genomes, 2902 genes in the hard-core group present in 80–90% of genomes, 262 genes in the soft core group present in 50–80% of genomes, 255 genes in the shell group present in 20–50% of genomes, 271 genes in the cloud group

**Fig. 1.** Available genes in core-pan genome of *P. agglomerans* ANP8.

present in less than 20% of genomes but which were not unique, and 179 genes present in only one genome (i.e. singletons) (Table 3). Thus, out of a total of 4401 genes in the genome of this strain, 4% genes are unique genes specific to this strain.

In order to provide a functional overview, an analysis of Gene Ontology (GO) was performed with the results summarized in Fig. 2.

According to Fig. 2, 23.12% of the genes are predicted to be involved in the biosynthetic process, which are a set of chemical reactions and pathways leading to the production of substances that are typically a part of the metabolism required for energy, and in which simpler substances are converted to more complex substances. These pathways include the synthesis of cellular carbohydrates, nitrogen compounds, auxins, cytokines, antibiotics, cofactors, mucilage, cellular regulators, etc. Results also showed that 11.82% of the genes are involved in small-molecules metabolic processes. Given that the genome studied is related to a plant growth promoting bacterium, the presence of a large number of genes involved in metabolic pathways related to small molecules, might be indicative of their role in communication with the plant or other plant associated beneficial bacteria. Results also showed that 11.38% of the genes are involved in the metabolism of nitrogen-containing compounds. Such compounds play important roles in the transfer of nitrogen between bacteria and plants.

Results also showed that 6.9% of genes are involved in transport processes. About 7% of genes involved in transport processes can explain the high tolerance of this bacterium to salinity. Also, 2.46% of the genes are involved in the processes of response to stress.

The results of 16S rRNA gene blast with the sequences in the NCBI database showed that this strain was 98.83959% similar to

Table 3
The results of Core-Pan genome analysis.

Number of genes in each group (%)	Group type
12.1	Core
66	HardCore
6	SoftCore
5.8	Shell
6.1	Cloud
4	Unique

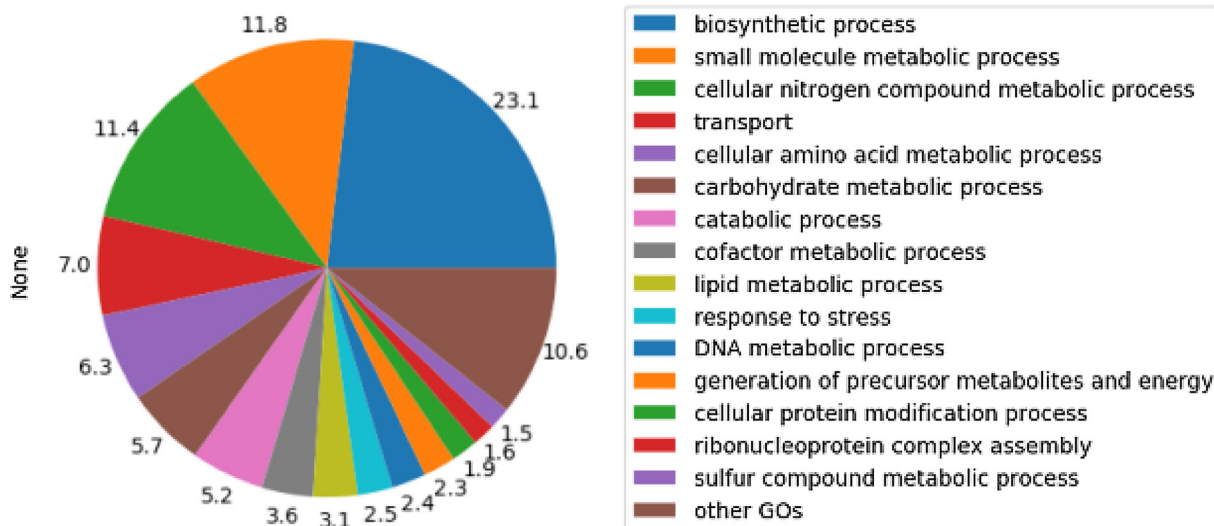


Fig. 2. Annotation results based the Gene Ontology (GO).

Pantoea agglomerans DAPP-PG7. The average nucleotide similarity between the studied genome and all available genomes for the respective genus (the existing primary genus was determined based on 16S), which are available in the GeneBank database, was determined through the use of FastANI 0.1.2. The results showed that the strain ANP8 presented the highest similarity to the bacterium *P. agglomerans* K1, thus suggesting that both strains belong to the same species.

Information on deposited data

This Whole Genome Shotgun project was deposited at DDBJ/ENA/GenBank under the accession JACBKS000000000. The version described in this paper is version JACBKS010000000, BioProject ID PRJNA644144, BioSample ID SAMN15449403.

Author contributions

All authors have contributed equally to this study.

Declaration of Competing Interest

The authors have no conflict of interest.

Acknowledgements

We are grateful for the financial support of the University of Tehran for doing the research.

References

- [1] H. Etesami, B.R. Glick, Halotolerant plant growth-promoting bacteria: prospects for alleviating salinity stress in plants, *Environ. Exp. Bot.* (2020) 104124.
- [2] F. Noori, H. Etesami, H. Najafi Zarini, N.A. Khoshkholgh-Sima, G. Hosseini Salekdeh, F. Alishahi, Mining alfalfa (*Medicago sativa* L.) nodules for salinity tolerant non-rhizobial bacteria to improve growth of alfalfa under salinity stress, *Ecotoxicol. Environ. Saf.* 162 (2018) 129–138.
- [3] E.V. Maas, G.J. Hoffman, Crop salt tolerance-current assessment, *J. Irrigation Drainage Div.* 103 (1977) 115–134.
- [4] H. Etesami, S.M. Adl, Can interaction between silicon and non-rhizobial bacteria benefit in improving nodulation and nitrogen fixation in salinity-stressed legumes? A review, *Rhizosphere* (2020)100229.
- [5] D. Putnam, M. Russelle, S. Orloff, J. Kuhn, L. Fitzhugh, L. Godfrey, A. Kiess, R. Long, Alfalfa, wildlife and the environment, Calif. Alfalfa Forage Assoc. (2001).
- [6] J. Yang, H. Jiang, G. Wu, H. Dong, Salinity shapes microbial diversity and community structure in surface sediments of the Qinghai-Tibetan Lakes, *Sci. Rep.* 6 (2016) 25078.
- [7] M. Kumar, H. Etesami, V. Kumar, *Saline Soil-based Agriculture by Halotolerant Microorganisms*, Springer, 2019.
- [8] P. Martínez-Hidalgo, A.M. Hirsch, The nodule microbiome: N2-fixing rhizobia do not live alone, *Phytobiomes* 1 (2017) 70–82.
- [9] G. Berg, M. Alavi, C.S. Schmidt, C. Zachow, D. Egamberdieva, F. Kamilova, B. Lugtenberg, Biocontrol and osmoprotection for plants under salinated conditions, *Mol. Microb. Ecol. Rhizosphere* 1 (2013) 561–573.
- [10] V. Shariati, M.A. Malboobi, Z. Tabrizi, E. Tavakol, P. Owilia, M. Safari, Comprehensive genomic analysis of a plant growth-promoting rhizobacterium *Pantoea agglomerans* strain P5, *Sci. Rep.* 7 (2017) 1–12.
- [11] J.M. Vincent, *A Manual for the Practical Study of the Root-nodule Bacteria. A Manual for the Practical Study of the Root-nodule Bacteria*, (1970) .