Complete genome sequence of the nitrogen-fixing bacterium *Azospirillum humicireducens* type strain SgZ-5T

Zhen Yu¹, Guiqin Yang¹, Xiaoming Liu¹, Yueqiang Wang¹, Li Zhuang²* and Shungui Zhou³

Abstract

The *Azospirillum humicireducens* strain SgZ-5T, belonging to the Order *Rhodospirillales* and the Family *Rhodospirillaceae*, was isolated from a microbial fuel cell inoculated with paddy soil. A previous work has shown that strain SgZ-5T was able to fix atmospheric nitrogen involved in plant growth promotion. Here we present the complete genome of *A. humicireducens* SgZ-5T, which consists of a circular chromosome and six plasmids with the total genome size of 6,834,379 bp and the average GC content of 67.55%. Genome annotations predicted 5969 protein coding and 85 RNA genes including 14 rRNA and 67 tRNA genes. By genomic analysis, we identified a complete set of genes that is potentially involved in nitrogen fixation and its regulation. This genome also harbors numerous genes that are likely responsible for phytohormones production. We anticipate that the *A. humicireducens* SgZ-5T genome will contribute insights into plant growth promoting properties of *Azospirillum* strains.

Keywords: *Azospirillum humicireducens*, Complete genome, Nitrogen fixation, PGPP

Introduction

*Bacteria* that live in the plant rhizosphere and possess a large array of potential mechanisms to enhance plant growth are considered as PGPR [1–3]. *Azospirillum* represents a well characterized genus of PGPR due to its capacity of fixing atmospheric nitrogen [4, 5]. Although the exact contribution of *Azospirillum* to biological nitrogen fixation in plant growth promotion is debated [2], agricultural applications of the genus *Azospirillum* have been still developed [6, 7]. Another main characteristic of *Azospirillum* proposed to explain plant growth promotion has been related to its ability to produce phytohormones [8, 9].

At present, there are 17 species within the genus *Azospirillum* [10], of which the nitrogen-fixing bacterium *A. humicireducens* SgZ-5T, the focus species of this study, was initially isolated from the anode biofilm of a MFC. A soil sample collected from paddy field in Guangzhou City, Guangdong Province, China (23.18° N 113.36° E) was used as inoculating source of the MFC. In a previous report [11], the nitrogen-fixing capability of strain SgZ-5T was confirmed by acetylene-reduction assay and identification of a *nifH* gene. Furthermore, this strain has the ability to grow under anaerobic conditions via the oxidation of various organic compounds coupled to the reduction of humus [11], showing its potential use in plant rhizosphere. Here, we describe the physiological features together with the whole genome sequence of *A. humicireducens* SgZ-5T.

Organism information

Classification and features

*A. humicireducens* SgZ-5T is a Gram-negative, facultative anaerobic, motile, spiral, straight to slightly curved rod-shaped bacterium (Fig. 1), belonging to the Order *Rhodospirillales* and the Family *Rhodospirillaceae*. The strain grew optimally in the conditions of 30 °C, pH 7.2, and 1% NaCl [11]. On NA, strain SgZ-5T formed cream-colored, round, smooth, convex and non-translucent colonies (Fig. 1). With AQDS as the sole terminal electron acceptor, strain SgZ-5T could utilize pyruvate, glucose and acetate as electron donors under anaerobic conditions [11]. Strain SgZ-5T was able to use a range of carbon substrates including N-Acetyl-glucosamine, citrate, D-ribose,
Fig. 1 Images of the *A. humicireducens* SgZ-5T. **a** Colonies of the strain on NA agar plate, **b** light microscopy and **c** transmission electron microscopy of the strain.

Table 1  Classification and general features of *A. humicireducens* SgZ-5\textsuperscript{T} according to the MIGS recommendations \[16\]

| MIGS ID | Property                        | Term                                                                 | Evidence code* |
|---------|---------------------------------|----------------------------------------------------------------------|----------------|
|         | Current Classification          | Domain **Bacteria**                                                  | TSA \[22\]    |
|         | Phylum                           | **Proteobacteria**                                                   | TSA \[33\]    |
|         | Class                             | **Alphaproteobacteria**                                             | TSA \[34\]    |
|         | Order                             | **Rhodospirillales**                                                | TSA \[35, 36\]|
|         | Family                            | **Rhodospirillaceae**                                               | TSA \[35, 36\]|
|         | Genus                             | **Azospirillum**                                                     | TSA \[37, 38\]|
|         | Species                           | **Azospirillum humicireducens**                                     | TSA \[11\]    |
|         | Type strain                       | SgZ-5=CCUG AB 2012021=KACC 16605                                    | TSA \[11\]    |
|         | Gram stain                        | Negative                                                            | TSA \[11\]    |
|         | Cell shape                        | Spiral, straight to slightly curved rods                           | TSA \[11\]    |
|         | Motility                          | Motile                                                              | TSA \[11\]    |
|         | Sporulation                       | Nonsporulating                                                      | NSA            |
|         | Temperature range                 | 25–37 °C                                                            | TSA \[11\]    |
|         | Optimum temperature               | 30 °C                                                               | TSA \[11\]    |
|         | pH range; Optimum                 | 5.5–8.5; 7.2                                                        | TSA \[11\]    |
|         | Carbon source                     | Acetate, L-lactate, citrate, D-ribose, L-rhamnose, D-glucose, N-Acetyl-glucomamine, meso-inositol, D-saccharose, D-maltose, suberic acid, malonate, L-serine, salicin, L-alanine, gluconate, glycogen, 2-keto-gluconate, D-mannitol, D-melibiose, L-fucose, D-sorbitol, L-arabinose, L-histidine, 3-hydroxy-butryic acid, 4-hydroxy-benzoic acid, L-proline, capric acid, adipic acid and malic acid | TSA \[11\]    |
| MIGS-6  | Habitat                           | Paddy soil                                                          | TSA \[11\]    |
| MIGS-6.3| Salinity                          | NaCl 0–1% (w/v)                                                     | TSA \[11\]    |
| MIGS-22 | Oxygen requirement                | Facultative anaerobic                                               | TSA \[11\]    |
| MIGS-15 | Biotic relationship               | Free living                                                         | NAS            |
| MIGS-14 | Pathogenicity                     | Not reported                                                        | NAS            |
| MIGS-4  | Geographic location               | Guangzhou City, Guangdong Province, China                          | NAS            |
| MIGS-5  | Sample collection time            | Dec 2011                                                            | NAS            |
| MIGS-4.1| Latitude                          | 23.18° N                                                            | NAS            |
| MIGS-4.2| Longitude                         | 113.36° E                                                           | NAS            |
| MIGS-4.3| Depth                             | 0.1 m beneath the surface                                           | NAS            |
| MIGS-4.4| Altitude                          | 40 m                                                                | NAS            |

*Evidence code: 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). These evidence codes are from the Gene Ontology Project \[39\]*
meso-inositol, D-saccharose, D-maltose, L-rhamnose, suberic acid, malonate, acetate, L-serine, salicin, L-lactate, L-alanine, gluconate, 2-keto-gluconate, glycogen, D-mannitol, D-glucose, D-melibiose, L-fucose, D-sorbierte, L-arabinose, L-histidine, 3-hydroxy-butyric acid, 4-hydroxy-benzoic acid, L-proline, capric acid, adipic acid and malic acid [11] (Table 1).

A phylogenetic tree was constructed from aligning the 16S rRNA gene sequences of strain SgZ-5T and type strains of the genus *Azospirillum* by MEGA 5 using the neighbour-joining method [12]. The phylogenetic position of strain SgZ-5T is shown in Fig. 2, where *A. humicireducens* can be grouped as a *Azospirillum* species, forms a distinct subclade together with *A. lipoferum* that are known as a biofertilizer widely used for agricultural production [13, 14]. The 16S rRNA gene of strain SgZ-5T is 98% similar to that of *A. lipoferum* NCIMB 11861T. Since *nifH* gene is highly conserved among nitrogen-fixing Proteobacteria [15], a *nifH*-based phylogenetic tree was constructed to identify the relationship of *A. humicireducens* to other species within the genus *Azospirillum* and related genus (Additional file 1). The phylogenetic reconstruction indicated the close relationship of the *A. humicireducens* SgZ-5T *nifH* gene with that from *Azospirillum* sp. B510.

**Genome sequencing information**

**Genome project history**

*A. humicireducens* SgZ-5T was selected for genome sequencing on the basis of its biotechnological potential in agricultural applications as a PGPR likely harboring multiple PGPP [11]. The complete genome sequences have been deposited at Gen-Bank/EMBL/DDBJ under the accession numbers CP015285.1, CP028902-CP028907. Project information is available from Genome Online database number Gp0150267 at Joint Genome Institute.

**Table 2** Genome sequencing project information

| MIGS ID | Property            | Term                          |
|---------|---------------------|-------------------------------|
| MIGS-31 | Finishing quality   | Complete                      |
| MIGS-28 | Libraries used      | Three libraries (a paired-end library and two mate-pair libraries) |
| MIGS-29 | Sequencing platforms| Illumina Hiseq 2500           |
| MIGS-31.2| Fold coverage      | 259x                          |
| MIGS-30 | Assemblers          | SOAPdenovo 2.04 [17]          |
| MIGS-32 | Gene calling method | GeneMarkS+ [18]               |
|         | Locus Tag           | A6A40                         |
|         | Genbank ID          | CP015285.1, CP028902-CP028907 |
|         | Genbank Date of Release | April 18, 2018             |
|         | GOLD ID             | Gp0150267                     |
|         | Bioproject          | PRJNA318554                   |
| MIGS-13 | Source Material Identifier | SgZ-5T                      |
|         | Project relevance   | Type strain, nitrogen fixation, plant growth promotion |

**Fig. 2** Phylogenetic tree highlighting the position of *A. humicireducens* SgZ-5T relative to other type strains within the genus *Azospirillum*. The strains and their corresponding GenBank accession numbers of 16S rRNA genes were indicated in parentheses. The sequences were aligned using Clustal W and the neighbor-joining tree was constructed based on kumura 2-paramenter distance model by using MEGA 5. Bootstrap values above 50% were obtained from 1000 bootstrap replications. Bar, 0.01 substitutions per nucleotide position. *Rhodovulum adriaticum* DSM 2781T was used as an outgroup.
In Table 2, we summarize the project information and its association with Minimum Information about a Genome Sequence (MIGS) [16].

**Growth conditions and genomic DNA preparation**

*A. humicireducens* SgZ-5T was routinely cultured in NB medium containing (L⁻¹) 5 g peptone, 3 g beef extract and 5 g NaCl at 30 °C. For genome sequencing, total genomic DNA was extracted from 10 mL overnight cultures using a DNA extraction kit following the manufacturer’s instructions (Aidlab). Quantification and quality control of the genomic DNA were completed by using a Qubit fluorometer (Invitrogen, CA, USA) with Qubit dsDNA BR Assay kit and 0.7% agarose gel electrophoresis with λ-Hind III digest DNA marker.

**Genome sequencing and assembly**

Complete genome sequencing was performed on an Illumina HiSeq 2500 system by constructing three DNA libraries (a paired-end library with insert size of 491 bp, and two mate pair libraries with insert sizes of 2.5 and 6.9 kb). After filtering low quality and Illumina PCR adapter reads, a total of 1967 Mb clean data were obtained from 2052 Mb raw data. Subsequently, all reads data were *denovo* assembled into a circular contig with 259 folds of genomic coverage, using SOAPdenovo v.2.04 [17]. Detailed genome sequencing project information is shown in Table 2.

**Genome annotation**

Gene prediction was carried out by GeneMarkS v.4.6 [18]. Function annotation of predicted ORFs was performed based on a BLASTP search against NCBI nonredundant protein database and COG database. Transfer RNAs, rRNAs and sRNA were predicted using tRNAscan-SE v.1.31 with the bacterial model, RNAmmer v.1.2 and Rfam database v.9.1, respectively [19–21]. The CRISPRs were identified by using the CRISPR database [22]. The prediction of genes with signal peptides and transmembrane helices were performed by SignalP server v.4.1 [23] and TMHMM server v.2.0 [24], respectively. The secondary metabolism gene cluster was predicted according to the antiSMASH v.3.0 procedure [25].

**Genome properties**

The genome of *A. humicireducens* SgZ-5T comprises a circular chromosome of 3,181,617 bp and six circular plasmids, designated as pYZ1 (715,112 bp), pYZ2 (1,098,603 bp), pYZ3 (252,411 bp), pYZ4 (338,445 bp), pYZ5 (626,509 bp) and pYZ6 (711,682 bp) (Table 3).

The total size of the genome is 6,834,379 bp, and the average GC content is 67.55%. The genome contains 6054 genes with the total length of 5,902,731 bp, of which 5969 (98.6%) are protein coding genes. There are 85 RNA genes (1.4%), including 14 rRNA and 67 tRNA genes. A total of 4844 genes (80.0%) have been assigned a predicted function while the rest have been designated as hypothetical proteins. Genome statistics are summarized in Table 4 and a graphical map is represented in Fig. 3. Furthermore, 4550 (75.2%) genes were assigned to 21 COG functional categories. The distribution of genes into different COG functional categories is provided in Table 5. Six *Azospirillum* species genomes (including *A. humicireducens*) of characterized strains are compared in Table 6. Almost all of these *Azospirillum* genomes consisting of 6–8 replicons have the total size of 6.5–7.6 Mb and the average GC content of 67.5–70.7%, and contain the total genes in the range of 5951 to 6982 [3, 6, 26, 27]. Furthermore, the main features of *A. humicireducens* SgZ-5T genome are close to those of *A. lipoferum* 4B genome.

### Table 3 Summary of genome: one chromosome and six plasmids

| Label | Size (bp) | Topology | INSDC identifier | RefSeq ID |
|-------|-----------|----------|------------------|-----------|
| Chromosome | 3,181,617 | Circular | CP015285.1 | NZ_CP015285.1 |
| pYZ1 | 715,112 | Circular | CP028902.1 | NA |
| pYZ2 | 1,098,603 | Circular | CP028903.1 | NA |
| pYZ3 | 252,411 | Circular | CP028904.1 | NA |
| pYZ4 | 338,445 | Circular | CP028905.1 | NA |
| pYZ5 | 626,509 | Circular | CP028906.1 | NA |
| pYZ6 | 711,682 | Circular | CP028907.1 | NA |

### Table 4 Genome statistics of *A. humicireducens* SgZ-5T

| Attribute | Genome (total) | % of total |
|-----------|----------------|------------|
| Genome size (bp) | 6,834,379 | 100.00 |
| DNA coding (bp) | 5,902,731 | 86.37 |
| DNA G + C (bp) | 4,616,422 | 67.55 |
| DNA scaffolds | 7 | |
| Total genes | 6054 | 100.00 |
| Protein coding genes | 5969 | 98.60 |
| RNA genes | 85 | 1.40 |
| rRNA genes | 14 | 0.23 |
| tRNA genes | 67 | 1.11 |
| Pseudo genes | 194 | 3.20 |
| Genes in internal clusters | NA | |
| Genes with function prediction | 4844 | 80.01 |
| Genes assigned to COGs | 4550 | 75.16 |
| Genes with signal peptides | 425 | 7.02 |
| Genes with transmembrane helices | 1022 | 16.88 |
| CRISPR repeats | 3 | |
Insights into the genome sequence

Nitrogen fixation is the major proposed mechanism, by which *Azospirillum* affects plant growth [2, 4]. A complete set of genes encoding enzymes involved in nitrogen fixation was found in the genomic analysis of *A. humicireducens* SgZ-5T (Table 7). The main genes involved in this process are *nif* genes, of which *nifDK* genes (A6A40_02900 and A6A40_02895) annotated as nitrogenase molybdenum-iron proteins and *nifH* gene (A6A40_02905) encoding dinitrogenase reductase protein have been identified. In the upstream region of the *nifHDK* operon, we have found that *nifEN* genes (A6A40_02875 and A6A40_02870) involved in synthesis of the molybdenum-iron cofactor of nitrogenase are clustered into a single operon together with *nifX* (A6A40_02865). Furthermore, the genome of *A. humicireducens* SgZ-5T has *nifUSVW* genes (A6A40_02235, A6A40_02230, A6A40_02225 and A6A40_02215), which are separated from the structural *nifENX* operon by about 160 kb.

Organization of the nitrogen fixation gene cluster in *A. humicireducens* SgZ-5T is presented in Fig. 4. Except for the separately transcribed *nifA* (A6A40_09040), *nifB* (A6A40_09050) and *nifZ* genes (A6A40_09070 and A6A40_09075), all the *nif* genes have resided in the nitrogen fixation gene cluster of 176.7 kb. Besides, an operon containing *fixABCX* genes (A6A40_02185, A6A40_02190, A6A40_02195 and A6A40_02220) responsible for electron transfer to nitrogenase is located upstream of this gene cluster. Nevertheless, the *fixABCX* operon is generally regulated by a transcriptional activator NifA protein for all nitrogen-fixing bacteria in the genus *Azospirillum* studied so far [5]. Furthermore, *draTG* genes (A6A40_02920 and A6A40_02925) implicated in posttranslational regulatory process of nitrogenase activity were found in the downstream of and divergently oriented with respect to *nifHDK* genes. On the whole, the nitrogen fixation gene cluster of *A. humicireducens* SgZ-5T was in agreement with that in *A. brasilense*, *A. lipoferum* and *Azospirillum* sp.
### Table 5 Number of genes associated with general COG functional categories

| Code | Value | % of total | Description                                      |
|------|-------|------------|--------------------------------------------------|
| J    | 182   | 2.98       | Translation, ribosomal structure and biogenesis   |
| A    | 0     | 0.00       | RNA processing and modification                   |
| K    | 357   | 5.85       | Transcription                                     |
| L    | 175   | 2.87       | Replication, recombination and repair             |
| B    | 1     | 0.02       | Chromatin structure and dynamics                  |
| D    | 38    | 0.62       | Cell cycle control, cell division, chromosome partitioning |
| V    | 80    | 1.31       | Defense mechanisms                                |
| T    | 338   | 5.54       | Signal transduction mechanisms                    |
| M    | 218   | 3.57       | Cell wall/membrane/envelope biogenesis            |
| N    | 73    | 1.20       | Cell motility                                     |
| U    | 58    | 0.95       | Intracellular trafficking, secretion, and vesicular transport |
| O    | 162   | 2.65       | Posttranslational modification, protein turnover, chaperones |
| C    | 342   | 5.60       | Energy production and conversion                  |
| G    | 263   | 4.31       | Carbohydrate transport and metabolism             |
| E    | 448   | 7.34       | Amino acid transport and metabolism               |
| F    | 81    | 1.33       | Nucleotide transport and metabolism               |
| H    | 160   | 2.62       | Coenzyme transport and metabolism                 |
| I    | 139   | 2.28       | Lipid transport and metabolism                    |
| P    | 333   | 5.45       | Inorganic ion transport and metabolism            |
| Q    | 144   | 2.36       | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 227   | 3.72       | General function prediction only                  |
| S    | 731   | 11.97      | Function unknown                                  |
| –    | 1555  | 25.47      | Not in COGs                                       |

*aThe total is based on the total number of protein coding genes in the annotated genome*

### Table 6 Genome statistics comparison among characterized *Azospirillum* species

| Genome name | 1 | 2 | 3 | 4 | 5 | 6 |
|-------------|---|---|---|---|---|---|
| Sp 7        | 6.6 | 7.4 | 7.5 | 6.8 | 6.5 | 7.6 |
| Az39        | 5 | 6 | 6 | NA | 7 | 6 |
| Sp245       | 68.3 | 68.5 | 68.5 | 67.7 | 70.7 | 68.2 |
| DSM 3675    | 6713 | 6982 | 6137 | 5999 | 6684 | 6692 |
| BV-S        | 6054 | 67.5 | 67.5 | 67.5 | 67.5 | 67.5 |
| BS10        | 6054 | 67.5 | 67.5 | 67.5 | 67.5 | 67.5 |
| SgZ-5       | 6054 | 67.5 | 67.5 | 67.5 | 67.5 | 67.5 |

*aThe *Azospirillum* species are numbered as: 1, *A. brasilense* [26, 27]; 2, *A. lipoferum* [26]; 3, *A. halopraeferens* (RefSeq ID: NZ_AUCF00000000.1); 4, *A. thiophilum* [3]; 5, *Azospirillum* sp. [6]; 6, *A. humicireducens*
### Table 7 Genes of *A. humicireducens* SgZ-5\(^T\) involved in nitrogen fixation

| Locus Tag   | Size/aa | Gene   | Gene product                                               |
|-------------|---------|--------|------------------------------------------------------------|
| A6A40_02185 | 852     | fixA   | Electron transfer flavoprotein beta subunit                |
| A6A40_02190 | 1080    | fixB   | Electron transfer flavoprotein alpha chain                 |
| A6A40_02195 | 1302    | fixC   | Flavoprotein-ubiquinone oxidoreductase                     |
| A6A40_09085 | 210     | fixU   | Nitrogen fixation protein                                  |
| A6A40_02200 | 285     | fixX   | Ferredoxin-like protein                                    |
| A6A40_09040 | 1866    | nifA   | Nif-specific transcriptional activator                     |
| A6A40_09050 | 1518    | nifB   | Nitrogenase FeMo cofactor biosynthesis protein              |
| A6A40_02900 | 1440    | nifD   | Nitrogenase molybdenum-iron protein alpha chain            |
| A6A40_02875 | 1407    | nifE   | Nitrogenase molybdenum-cofactor biosynthesis protein       |
| A6A40_02905 | 897     | nifH   | Nitrogenase iron protein                                   |
| A6A40_02895 | 1560    | nifK   | Nitrogenase molybdenum-iron protein subunit beta           |
| A6A40_02870 | 1371    | nifN   | Nitrogenase molybdenum-cofactor biosynthesis protein       |
| A6A40_02230 | 1206    | nifS   | Nitrogenase metalloclusters biosynthesis protein           |
| A6A40_02235 | 924     | nifU   | Iron-sulfur cluster assembly scaffold protein               |
| A6A40_02225 | 1122    | nifV   | Homocitrate synthase                                       |
| A6A40_02215 | 336     | nifW   | Nitrogenase-stabilizing/protective protein                 |
| A6A40_02865 | 399     | nifX   | Nitrogenase molybdenum-iron protein                        |
| A6A40_09070 | 333     | nifZ   | Nitrogenase P-cluster assembly                             |
| A6A40_09075 | 306     | nifZ   | Nitrogenase P-cluster assembly                             |
| A6A40_02220 | 852     | cysE   | Serine acetyltransferase                                   |
| A6A40_02925 | 909     | draG   | ADP-ribosyl-[dinitrogen reductase] hydrolase               |
| A6A40_02920 | 891     | draT   | ADP-ribosyl-[dinitrogenase reductase] transferase          |
| A6A40_07245 | 2847    | glnD   | [Protein-Pii] uridylyltransferase                          |
| A6A40_07685 | 339     | glnB   | Nitrogen regulatory protein P-II                           |
| A6A40_05220 | 1200    | ntrB   | Nitrogen regulation sensor histidine kinase                |
| A6A40_05215 | 1146    | ntrC   | Nitrogen regulation response regulator                     |
| A6A40_05205 | 1401    | ntrX   | Sigma-54-dependent transcriptional regulator              |
| A6A40_05210 | 2319    | ntrY   | Nitrogen regulation sensor histidine kinase                |

**Fig. 4** Organization of the nitrogen fixation gene cluster in *A. humicireducens* SgZ-5\(^T\). Arrows represent genes and their respective direction of transcription. Genes are colored as depicted in the lower box.
B510 [6, 26, 28, 29], suggesting that nitrogen fixation process demands the systematic action of various genes.

Since tryptophan is a main precursor for biosynthesis of IAA, a well-known phytohormone [30], the genes in A. humicireducens SgZ-5\(^\text{T}\) related to the production of this amino acid have been analyzed (Additional file 2). The genome harbors three genes \textit{trpE}, \textit{trpG} and \textit{trpEG} (A6A40_04380, A6A40_04655 and A6A40_05775), each encoding the key enzyme anthranilate synthase in tryptophan biosynthesis. Together with \textit{trpG}, the genes \textit{trpD} (A6A40_04650) and \textit{trpC} (A6A40_04645) form a gene cluster of 2.4 kb. Except for anthranilate synthase, this \textit{trpGDC} gene cluster encodes anthranilate phosphoribosyltransferase and indole-3-glycerol phosphate synthase, which plays a role in synthesis of tryptophan used in multiple biological processes including IAA biosynthesis [31]. The same \textit{trpGDC} cluster was previously found in \textit{A. brasiliense} [32]. Although the \textit{idpC} gene, related to the indole-3-pyruvate pathway for the biosynthesis of IAA [30], was not discovered in the \textit{A. humicireducens} SgZ-5\(^\text{T}\) genome, alternative pathway might exist in SgZ-5\(^\text{T}\). In the genome, A6A40_22745 and A6A40_22755 were assigned as candidates for \textit{iaaM} and \textit{iaaH} genes, respectively. These two genes were also found in the \textit{Azospirillum} sp. B510 genome, and are known to be involved in the IAM pathway for IAA biosynthesis by catalyzing the decarboxylation of tryptophan into IAM and the hydrolysis of IAM to produce IAA [6, 30].

The \textit{A. humicireducens} SgZ-5\(^\text{T}\) genome also contains a \textit{terpene} gene cluster of 24.0 kb consisting of 23 genes (A6A40_04945, A6A40_04950, A6A40_04955, ..., A6A40_05055) (Additional file 3). This gene cluster encodes a series of proteins, which are involved in the biosynthesis of secondary metabolite production of terpenoid. Thereinto, A6A40_05010 was indentified as the \textit{crtB} gene, encoding phytoene synthase involved in the biosynthesis of carotenoid. Similar genes in this gene cluster were previously observed in the \textit{A. lipoforum} 4B genome [7, 26]. Furthermore, some phytohormones including gibberelmins and abscisic acid with over 120 types found in plants, fungi, and bacteria, are synthesized through the terpenoid pathway [2]. Therefore, \textit{A. humicireducens} SgZ-5 exhibits an attractive application as a PGPR likely harboring multiple PGPP in agriculture.

### Conclusion

We report here an inventory of the genomic features of the nitrogen-fixing bacterium \textit{A. humicireducens} SgZ-5\(^\text{T}\). The genome sequence of strain SgZ-5\(^\text{T}\) revealed further genetic elements involved in nitrogen fixation and its regulation, as well as in the production of phytohormones. We anticipate that knowledge of this genome will contribute to new insights into the mechanisms of plant growth stimulation through genomic comparisons among available complete genomes of \textit{Azospirillum} strains.

### Additional files

**Additional file 1:** Phylogenetic tree based on the partial \textit{nifH} gene sequences showing the position of \textit{A. humicireducens} SgZ-5\(^\text{T}\) relative to other species within the genus \textit{Azospirillum} and related genus. The strains and their corresponding GenBank accession numbers of \textit{nifH} gene were indicated in parentheses. The sequences were aligned using Clustal W and the neighbor-joining tree was constructed based on kimura 2-parameter distance model by using MEGA 5. Bootstrap values above 50% were obtained from 1000 bootstrap replications. Bar, 0.01 substitutions per nucleotide position. \textit{Leptospirillum ferriphilum} YSKT was used as an outgroup. (DOCX 64 kb)

**Additional file 2:** Genes of \textit{A. humicireducens} SgZ-5\(^\text{T}\) involved in biosynthesis of tryptophan. (DOCX 16 kb)

**Additional file 3:** Genes of \textit{A. humicireducens} SgZ-5\(^\text{T}\) located in a terpene gene cluster. (DOCX 16 kb)

### Abbreviations

- A6A40: Genes of \textit{A. humicireducens} SgZ-5\(^\text{T}\) gene cluster.
- A6A40_04945, A6A40_04950, A6A40_04955, ..., A6A40_05055 (Additional file 3). This gene cluster encodes a series of proteins, which are involved in the biosynthesis of secondary metabolite production of terpenoid. Thereinto, A6A40_05010 is indentified as the \textit{crtB} gene, encoding phytoene synthase involved in the biosynthesis of carotenoid. Similar genes in this gene cluster were previously observed in the \textit{A. lipoforum} 4B genome [7, 26]. Furthermore, some phytohormones including gibberelmins and abscisic acid with over 120 types found in plants, fungi, and bacteria, are synthesized through the terpenoid pathway [2]. Therefore, \textit{A. humicireducens} SgZ-5 exhibits an attractive application as a PGPR likely harboring multiple PGPP in agriculture.

### Additional files

**Additional file 1:** Phylogenetic tree based on the partial \textit{nifH} gene sequences showing the position of \textit{A. humicireducens} SgZ-5\(^\text{T}\) relative to other species within the genus \textit{Azospirillum} and related genus. The strains and their corresponding GenBank accession numbers of \textit{nifH} gene were indicated in parentheses. The sequences were aligned using Clustal W and the neighbor-joining tree was constructed based on kimura 2-parameter distance model by using MEGA 5. Bootstrap values above 50% were obtained from 1000 bootstrap replications. Bar, 0.01 substitutions per nucleotide position. \textit{Leptospirillum ferriphilum} YSKT was used as an outgroup. (DOCX 64 kb)

**Additional file 2:** Genes of \textit{A. humicireducens} SgZ-5\(^\text{T}\) involved in biosynthesis of tryptophan. (DOCX 16 kb)

**Additional file 3:** Genes of \textit{A. humicireducens} SgZ-5\(^\text{T}\) located in a terpene gene cluster. (DOCX 16 kb)

### Abbreviations

- AQDS: Anthraquinone-2, 6-disulfonate; IAA: Indole-3-actic acid; IAM: Indole-3-acetamide; MFC: Microbial fuel cell; NA: Nutrient Agar; NB: Nutrient Broth; PGPP: Plant growth promoting properties; PGPR: Plant growth-promoting rhizobacteria

### Acknowledgments

This work was supported by the Guangdong Academy of Sciences Funds for Innovation Driven Development, China (2017GDAASCX-0409), the National Natural Science Foundation of China (41501546), the Guangdong Natural Science Foundation, China (2016A030313779), and the Science and Technology Planning Project of Guangdong, China (2017A030303057).

### Authors’ contributions

LZ and SZ conceived and designed the experiments. GY, YW and XL performed the experiments. ZY assembled and analysed genome. ZY and LZ drafted the manuscript. GY and SZ revised the manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

1Guangdong Key Laboratory of Agricultural Environment Pollution Integrated Control, Guangdong Institute of Eco-Environmental Science & Technology, Guangzhou 510650, China. 2School of Environment, Guangdong Key Laboratory of Environmental Pollution and Health, Jinan University, College of Resources and Environment, Fujian Agricultural and Forestry University, Fuzhou 350002, China.

### Received

26 February 2017 Accepted: 24 September 2018 Published online: 16 October 2018

### References

1. Okon Y, Vanderleyden J. Root-associated Azospirillum species can stimulate plants. ASM News. 1997;63:366–70.
2. Bashan Y, De-Bashan LE. How the plant growth-promoting bacterium \textit{Azospirillum} promotes plant growth-a critical assessment. Adv Agron. 2010; 108:77–136.
3. Kwak Y, Shin JH. First \textit{Azospirillum} genome from aquatic environments: whole-genome sequence of \textit{Azospirillum thiooxidans} Bv-ST, a novel dioxidotrophic harboring a capacity of sulfur-chemolithotrophy from a sulfide spring. Mar Genomics. 2016;25:21–4.
4. Okon Y, Heytler PG, Hardy RWF. \textit{N}2 fixation by \textit{Azospirillum brasiliense} and its incorporation into host \textit{Setaria italica}. Appl Environ Microbiol. 1983;46:694–7.

Yu et al. Standards in Genomic Sciences (2018) 13:28
5. Steenhoudt O, Vanderleyden J. Azospirillum, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. FEMS Microbiol. Rev. 2002;24:487–506.

6. Kaneko T, Minamiya K, Iwasa T, Nakatsuka H, Mitsui H, Kawahara Y, et al. Complete genomic structure of the cultivated rice endophyte Azospirillum sp. BS10. DNA Res. 2010;17:37–50.

7. Droge B, Sanguin H, Borland S. Genome wide profiling of Azospirillum lipoferum 4B gene expression during interaction with rice roots. FEMS Microbiol. Ecol. 2014;87:543–55.

8. Tsavkelova EA, Klimova SY, Cherdyntseva TA, Netrusov AI. Microbial producers of plant growth stimulators and their practical use: a review. Appl. Biochem. Microbiol. 2006;42:117–26.

9. Sant’Anna FH, Almeida LGPL, Cecagno R, Reolon LA, Siqueira FM, Machado MRS, et al. Genomic insights into the versatility of the plant growth-promoting bacterium Azospirillum amazonense. BMC Genomics. 2011;1:2409.

10. Lin SY, Liu YC, Hameed A, Hsu YH, Huang H, Lai WA, et al. Azospirillum aggrega sp. nov., a nitorgen-fixing species isolated from cultivated soil. Int J Syst Evol Microbiol. 2016;66:1453–8.

11. Zhou SG, Han LC, Wang YQ, Yang GQ, Zhuang L, Hu P. Azospirillum humicola sp. nov., a nitorgen-fixing bacterium isolated from a microbial fuel cell. Int J Syst Evol Microbiol. 2013;63:2618–24.

12. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetic analysis (MEGA) software version 4.0. Mol Ecol. 2007;16:279–90.

13. Nakade DB. Halotolerant Azospirillum lipoferum sp. nov. as a biofertilizer for saline soils. J Pure Appl Microbiol. 2013;7:491–7.

14. Tamitielvi S, Dutta A, Sindhu M. Azospirillum lipoferum and Pseudomonas fluorescens as effective biological agents for enhanced agro-productivity. BioSci. 2016;32:670–3.

15. Raymond J, Siefert JL, Staples CR, Blankenship RE. The natural history of nitrogen fixation. Mol Biol Evol. 2004;21:541–54.

16. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26:5:541–7.

17. Luo R, Liu B, Xie Y, Li J, Huang W, Yuan J, et al. SOAPdenovo2: an improved memory-efficient short-read de novo assembler. GigaScience. 2012;1:18.

18. Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene start in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 2001;29:2607–18.

19. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.

20. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8.

21. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, et al. Complete genome sequence of the cultivated rice endophyte Azospirillum amazonense. BMC Genomics. 2011;1:2409.

22. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: the problems of phylogeny. Syst. Zool. 1990;39:406–31.

23. Drogue B, Sanguin H, Borland S. Genome wide profiling of Azospirillum lipoferum 4B gene expression during interaction with rice roots. FEMS Microbiol. Ecol. 2014;87:543–55.

24. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting signal peptides from transmembrane regions. Nat Methods. 2011;8:785–9.

25. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, et al. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 2015;43:W237–43.

26. Wisniewski-Dyé F, Borziak K, Khalsa-Moyers G, Alexandre G, Sukharnikov LO, Wuchet K, et al. Azospirillum genomes reveal transition of Bacteria from aquatic to terrestrial environments. PLoS Genet. 2011;7:e1002430.

27. Rivera D, Revale S, Molina R, Gualda J, Puente M, Maroniche G, et al. Complete genome sequence of the model rhizosphere strain of Azospirillum brasilense aZ39, successfully applied in agriculture. Genome Announc. 2014;2(4):e00683–14.

28. Fu HA, Fitzmaurice WP, Roberts GP, Burris RH. Cloning and expression of draT genes from Azospirillum lipoferum. Gene. 1990;86:95–8.

29. Zhang Y, Burris RH, Roberts GP. Cloning, sequencing, mutagenesis, and functional characterization of draT and draG genes from Azospirillum brasilense. J Bacteriol. 1992;174:3364–9.

30. Spaepen S, Vanderleyden J, Remans R. Indole-3-acetic acid in microbial and microorganism-plant signaling. FEMS Microbiol. Rev. 2007;31:425–48.

31. Kageyama S, Asai S, Kim SJ, Yun BW, Lee U. Complete genome sequence of plant growth-promoting bacterium Leifsonia xyli SE134, a possible gibbonellin and auxin producer. J Biotechnol. 2016;239:34–8.

32. Zimmer W, Aparicio C, Elmerich C. Relationship between tryptophan biosynthesis and indole-3-acetic acid production in Azospirillum identification and sequencing of a tryptGDC cluster. Mol Gen Genet. 1991;229:41–51.

33. Garrity GM, Bell JA, Liburn T. Phylum XIV. Proteobacteria phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey’s manual of systematic bacteriology, volume 2, part B. 2nd ed. New York: Springer; 2005. p. 1–14.

34. Gogarten JP, Goldberg MD. Bacteria and archaebacteria. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, editors. Bergey’s Manual of systematic bacteriology, volume 2, part C. 2nd ed. New York: Springer; 2005. p. 1–14.

35. Pfennig N, Trüper HG. Higher taxa of the phototropic bacteria. Int J Syst Bacteriol. 1971;21:17–8.

36. Skerman VBD, McGowan V, Sheath PHA. Approved lists of bacterial names. Int J Syst Bacteriol. 1980;30:225–420.

37. Tanand JJ, Krieg NR, Dobberer SS, A.J. Taxonomic study of the Spirillum lipoferum group, with descriptions of a new genus, Azospirillum gen. nov. and two species, Azospirillum lipoferum (Beijerinck) comb. nov. and Azospirillum brasiliense sp. nov. Can J Microbiol. 1978;24:967–80.

38. List Editor. Validation list no. 2. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. Int J Syst Bacteriol. 1979;29:79–80.

39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.