

# SCIENTIFIC REPORTS



OPEN

## Complete genome sequence of the cellulose-producing strain *Komagataeibacter nataicola* RZS01

Heng Zhang<sup>1,2</sup>, Xuran Xu<sup>1,2</sup>, Xiao Chen<sup>1,2</sup>, Fanshu Yuan<sup>1,2</sup>, Bianjing Sun<sup>1,2</sup>, Yunhua Xu<sup>3</sup>, Jiazhi Yang<sup>1,2</sup> & Dongping Sun<sup>1,2</sup>

*Komagataeibacter nataicola* is an acetic acid bacterium (AAB) that can produce abundant bacterial cellulose and tolerate high concentrations of acetic acid. To globally understand its fermentation characteristics, we present a high-quality complete genome sequence of *K. nataicola* RZS01. The genome consists of a 3,485,191-bp chromosome and 6 plasmids, which encode 3,514 proteins and bear three cellulose synthase operons. Phylogenetic analysis at the genome level provides convincing evidence of the evolutionary position of *K. nataicola* with respect to related taxa. Genomic comparisons with other AAB revealed that RZS01 shares 36.1%–75.1% of sequence similarity with other AAB. The sequence data was also used for metabolic analysis of biotechnological substrates. Analysis of the resistance to acetic acid at the genomic level indicated a synergistic mechanism responsible for acetic acid tolerance. The genomic data provide a viable platform that can be used to understand and manipulate the phenotype of *K. nataicola* RZS01 to further improve bacterial cellulose production.

Acetic acid bacteria (AAB) are a group of microorganisms that belong to the family *Acetobacteraceae* of the class *Alphaproteobacteria*. These bacteria are widely found on fruits, flowers, and rotten food<sup>1</sup>. Their metabolic uniqueness has been utilized in the industrial production of sorbose, vitamin C, dihydroxyacetone, D-gluconic acid, and bacterial cellulose (BC)<sup>2–5</sup>. Furthermore, AAB can also act as fermentative organisms, therefore, they are applied in beer, wine, and vinegar production<sup>6,7</sup>. A remarkable feature of these strains is their ability to survive under extreme environments, such as high sugar concentrations and low pH values, which makes AAB suitable for various industrial applications<sup>8</sup>.

Currently, AAB include 14 genera, namely *Acetobacter*, *Gluconobacter*, *Gluconacetobacter*, *Komagataeibacter*, *Granulibacter*, *Asaia*, *Acidomonas*, *Kozakia*, *Swaminathania*, *Saccharibacter*, *Neoasaia*, *Tanticharoenia*, *Ameyamaea*, and *Neokomagataea*<sup>1</sup>. Classification within the *Acetobacteraceae* depends on the ubiquinone type, including the Q-9 and Q-10 types. *Acetobacter* species use the Q-9-type ubiquinone, whereas the AAB genera of *Gluconacetobacter* and *Gluconobacter* contain mainly the Q-10-type ubiquinone<sup>9</sup>. The genus *Komagataeibacter*, a gram-negative, obligately aerobic and rod-shaped acidophilic organism, was initially proposed by Yamada in 2012, with 12 other strains on the basis of their taxonomic characteristics<sup>10</sup>. *Komagataeibacter* gen. nov. originated from the *Gluconacetobacter xylinus* (previously named *Acetobacter xylinus*) group, which is first separated from the *Gluconacetobacter liquefaciens* group at the genus level. These 2 groups are significantly different in phylogenetic and phenotypic characteristics<sup>11</sup>. Thus far, 66 complete genomes of *Acetobacteraceae* have been published in NCBI databases (<http://www.ncbi.nlm.nih.gov/>), most of which just present draft genome sequences. The genomes of 3 species of *Komagataeibacter*, *K. xylinus* E25, *K. hansenii* ATCC 23769, and *K. medellinensis* NBRC3288, have been sequenced completely. However, draft genomes are available for *K. europaeus* LMG18494, *K. intermedius* AF2, *K. rhaeticus* AF1, *K. kakiaceti* JCM 25156, and *K. oboediens* 174Bp<sup>12,13</sup>.

Generally, *Komagataeibacter* species possess the ability to secrete exopolysaccharides (EPS), especially BC<sup>14</sup>. Notably, *K. medellinensis* NBRC3288 is the only cellulose-nonproducing member of *Komagataeibacter* species, isolated from vinegar<sup>15</sup>. BC has been the focus of research owing to its excellent properties, such as high purity, predominant three-dimensional structure, high degree of crystallinity, and superior biocompatibility<sup>16–18</sup>. In view

<sup>1</sup>Chemicobiology and Functional Materials Institute, Nanjing University of Science and Technology, Nanjing, 210094, China. <sup>2</sup>School of Chemical Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China. <sup>3</sup>Department of Life Sciences, Lianyungang Normal College, Lianyungang, 222000, China. Correspondence and requests for materials should be addressed to J.Y. (email: [jiazhiyang@sina.com](mailto:jiazhiyang@sina.com)) or D.S. (email: [sundpe301@163.com](mailto:sundpe301@163.com))

Features	
Size of the chromosome (bp)	3,485,191
Size of the plasmids (bp)	102,282 (pKNA01)
	39,914 (pKNA02)
	38,682 (pKNA03)
	37,912 (pKNA04)
	30,554 (pKNA05)
	25,766 (pKNA06)
DNA scaffolds	7
G + C content (%)	61.49
Total genes	3,609
Protein coding genes	3,514
RNA genes	95
Genes with function prediction	2,485

**Table 1.** General features of the *Komagataeibacter nataicola* RZS01 genome.

of these qualities, it is utilized commercially in the development and production of health food, flexible electrodes, acoustic speakers, and tissue scaffolds<sup>16, 19–21</sup>.

*K. nataicola* RZS01 originates from rotten apples and is regarded as the model organism for BC production, owing to its high resistance to ethanol and acetic acid, which makes it an ideal strain for industrial applications<sup>22, 23</sup>. In this study, we present the complete genome sequence and sequence analysis of *K. nataicola* RZS01 to obtain genetic information for further insights into its biochemical features. Comparison of *K. nataicola* RZS01 with other AAB strains will address the unique functional characteristics of this strain, as well as the common properties within the *Acetobacteraceae*. Furthermore, the genome data will also provide a rich resource for guiding future research in industrial applications.

## Results

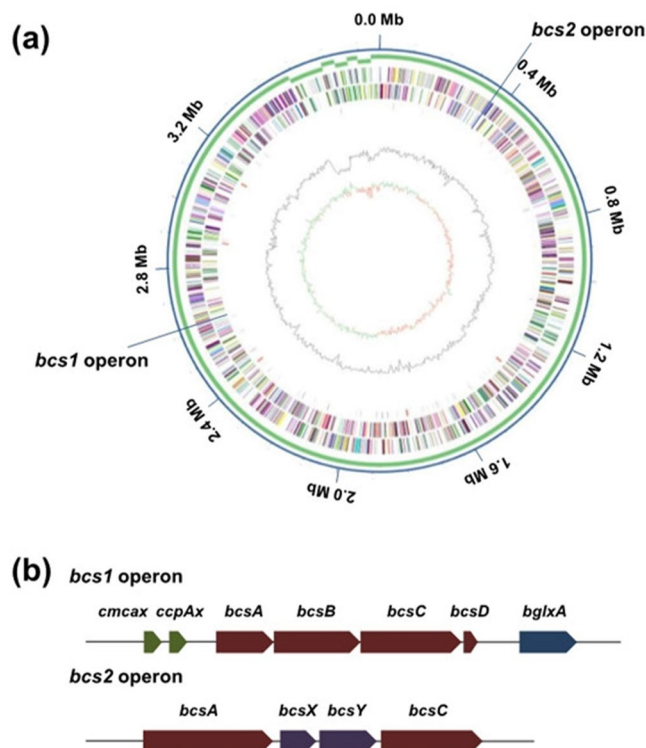
**General features of the *K. nataicola* RZS01 genome.** The complete genome of RZS01 is composed of 1 circular chromosome of 3,485,191 bp with a G + C content of 61.49%, and 6 plasmids ranging in size from 25,766 bp to 102,282 bp (Table 1; Fig. 1). Gene prediction and annotation of the RZS01 genome resulted in 3,609 open reading frames. The number of genes in the plasmids occupies 6.5% of the total genes in the genome, and genes with predicted function are assigned to 68.86% of the genome (2,485 genes). Furthermore, 5 ribosomal RNA operons (16S–23S–5S) were detected and 62 tRNA genes were predicted. From the genome data, we also identified 66 transposase genes, which might be responsible for the genetic instability leading to deficiencies in various physiological properties as reported in other AAB<sup>24</sup>.

The complete genome has a total of 3,514 putative coding sequences, among which 2,485 are assigned a putative function, and 1,029 encode hypothetical proteins. The 6 plasmids have 228 coding DNA sequences (CDSs), of which approximately 47.8% encode hypothetical proteins. The remaining 42 CDSs encode mobile element proteins. Figure 2 summarizes the distribution of RZS01 proteins among 21 functional groups. Statistically, nearly 38% CDSs are completely uncharacterized, which is similar to the proportion of unassigned CDSs in some other sequenced bacterial genomes, such as in *Escherichia coli* (40%), *Haemophilus influenzae* (43%), and *Mycoplasma genitalium* (32%)<sup>25</sup>. The 2 largest functional groups contain 277 and 231 proteins involved in carbohydrate and amino acid metabolism, respectively, which correspond to the physiological functions in RZS01.

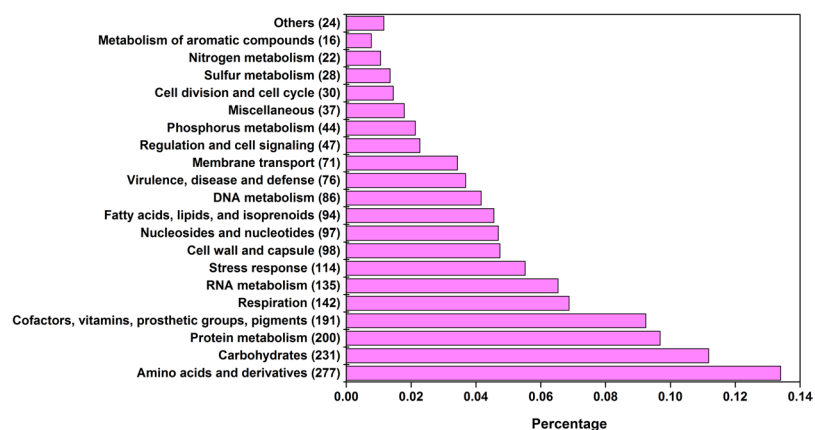
**Phylogenetic analysis and comparative genomics.** Analysis of the most conservative sequence in the process of evolution, the 16S rRNA gene, is widely used in taxonomic resolution in bacteria. To ascertain the phylogenetic positions of RZS01 in *Acetobacteraceae*, a phylogenetic tree was constructed using 16S rRNA gene sequences (Fig. 3). The phylogenetic tree summarizes 13 other *Acetobacteraceae*, the genomes of which have been completely sequenced. The tree suggests that RZS01 is most closely related to *Komagataeibacter xylinus* E25, *Komagataeibacter medellinensis* LMG 1693, and *Gluconacetobacter hansenii* ATCC 23769. Furthermore, comparative analysis of RZS01 in relation to other members of the *Acetobacteraceae* family reveals that this strain has 36.1% shared genes with the human pathogen *Granulibacter bethesdaensis* CGDNIH1 (Table 2). Approximately 40% of the genes are in common with members of the genus *Acidiphilium*. Thus, RZS01 has more genes in common with *Komagataeibacter* species than with *Gluconacetobacter*, which corresponds with the phylogenetic relationship.

**Cellulose production.** *Acetobacteraceae* are characterized by their ability to synthesize high yields of cellulose. Maria José Valera *et al.*<sup>26</sup> investigated the cellulose-producing ability of 77 strains of 35 species of AAB, of which 19 demonstrated high yields of cellulose after cultivation. Species of the genus *Komagataeibacter* are considered to be the main producers of BC, particularly strains of the species *K. xylinus*, *K. hansenii*, *K. swingsii*, *K. rhaeticus*, and *K. nataicola*<sup>27, 28</sup>. In fact, *K. xylinus* is considered a model microorganism for BC production<sup>29, 30</sup>.

On the RZS01 chromosome, we found the full set of cellulose synthase genes in the *bcsI* operon, which comprised the cellulose synthase genes *bcsA1* (B0W47\_12635), *bcsB1* (B0W47\_12640), *bcsC1* (B0W47\_12645), and *bcsD1* (B0W47\_12650), as well as *cmcaX* (B0W47\_12625), *ccpAx* (B0W47\_12630) upstream, and *bglxA*



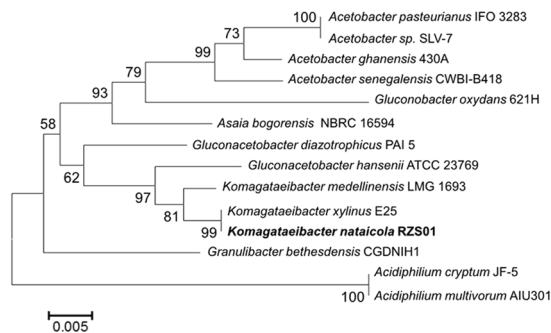
**Figure 1.** Overview of *K. nataicola* RZS01 genome. **(a)** The circles represent (from the outside to the inside): circle 1, DNA base position (bp); circle 2, contig components; circle 3, protein-coding regions transcribed clockwise; circle 4, protein-coding regions transcribed anti-clockwise; circle 5, distribution of genes encoding ncRNA (black), tRNA (blue) and rRNA (red); circle 6, G + C content; circle 7, GC skew. **(b)** The compositions of 2 *bcs* operons, which differ from each other in gene content.



**Figure 2.** Subsystem category distributions of RZS01 proteins.

(B0W47\_12655) downstream. *cmcax* encodes endo- $\beta$ -1,4-glucanase, and *bglxA* encodes  $\beta$ -glucosidase, both of which have been suggested to assist cellulose biosynthesis by hydrolysing tangled glucan chains when there is a failure in chain arrangement, and all these genes are necessary for cellulose production<sup>31</sup>. In addition to the *bcs1* operon, we found an additional *bcs* operon (*bcs2*) at the genomic position 363284–373685, and it differs in structure from the *bcs1* operon. The *bcs2* operon is composed of *bcsA* and *bcsC*, with 2 additional genes homologous to *bcsX* and *bcsY* in the middle of *bcsA* and *bcsC*. The protein encoded by *bcsA* contains the catalytically active subunit with a PilZ domain, which is responsive to c-di-GMP<sup>32</sup>. An additional *bcsA* (*bcsA3*) is also located on the reverse strand. Unlike the *bcsA1* gene, the *bcsA2* and *bcsA3* genes do not contain a PilZ domain.

Generally, AAB possess the ability to secrete other polysaccharides, including capsular polysaccharides, lipopolysaccharides, and water-soluble EPS. One of the most studied polysaccharides is a xanthan gum-like EPS, named 'acetan', the structure of which is shown in Supplementary Information. The chemical repeat unit consists of a cellobiose unit solubilized by attachment of a charged pentasaccharide sidechain to one of the glucose



**Figure 3.** A phylogenetic tree based on 16S rRNA sequences constructed by the neighbour-joining method of MEGA5.1.

Acetic acid bacterium strain	Number of shared genes	Percentage of shared genes
<i>Gluconacetobacter diazotrophicus</i> Pal5	5031	72.0
<i>Komagataeibacter medellinensis</i> NBRC 3288	4742	75.0
<i>Acetobacter senegalensis</i> 108B	4339	61.7
<i>Komagataeibacter xylinus</i> E25	4335	65.0
<i>Komagataeibacter hansenii</i> ATCC 23769	3592	52.4
<i>Acetobacter pasteurianus</i> IFO 3283	3216	51.1
<i>Acidiphilium multivorum</i> AIU301	3098	44.5
<i>Gluconobacter oxydans</i> 621H	2993	50.3
<i>Acetobacter sp.</i> SLV-7	2893	48.1
<i>Acetobacter ghanensis</i> 430A	2774	46.0
<i>Asaia bogorensis</i> NBRC 16594	2742	43.7
<i>Acidiphilium cryptum</i> JF-5	2656	39.4
<i>Granulibacter bethesdensis</i> CGDNIH1	2146	36.1

**Table 2.** Comparative analysis of *Komagataeibacter nataicola* RZS01. with members of the *Acetobacteraceae* family.

residues<sup>33,34</sup>. Genetic analysis of RZS01 shows that the biosynthesis of acetan occurs via a pathway similar to that for BC biosynthesis. Because it is the composite of different monosaccharides, 2 CDSs are responsible for encoding UDP-glycosyl transferase (B0W47\_01855, B0W47\_03955, and B0W47\_12635), GDP-mannosyl transferase (B0W47\_09075, and B0W47\_11925), and phosphomannose isomerase (B0W47\_06245). The enzymes UDP-glycosyl transferase and GDP-mannosyl transferase transfer a glucose and mannose residues from UDP-Glc and GDP-Man, respectively, to one of the intermediates of acetan<sup>35</sup>. However, not all genes from the gum operon are present in RZS01. Only 2 CDSs (B0W47\_14415 and B0W47\_14460) represent the genes of *gumK* and *gumH*, respectively. The genes for *gumB*, *C*, *D*, *E*, *G*, *I*, *J*, and *M* are absent in the RZS01 genome.

**Regulation of cellulose biosynthesis.** BC biosynthesis is regulated by the expression of *bcs* genes, which appear to be expressed in response to the second messenger cyclic-di-GMP (c-di-GMP)<sup>36</sup>. An earlier study revealed that cellulose biosynthesis could be stimulated almost 100-fold by c-di-GMP<sup>37</sup>. c-di-GMP is produced from 2 molecules of GTP by diguanylate cyclases (DGCs) and is broken down into 5'-phosphoguananylyl-(3'-5')-guanosine by specific phosphodiesterases (PDEs). DGC activity is associated with the GGDEF domain, whereas PDE activity is associated with EAL or HD-GYP domains; all these domains are essential for enzymatic activity<sup>38</sup>. We found 3 *cdg* operons (*cdg1*, *cdg2*, and *cdg3*) containing a c-di-GMP PDE gene followed by a DGC gene. Besides, 4 standalone c-di-GMP PDEs (*pde1-4*) and 1 standalone DGC are present in the genome, which share 32–40% amino acid sequence identity with each other. These multiple genes together control the intracellular c-di-GMP levels.

**Genes for substrate metabolism.** On the basis of the analysis of gene annotation for the complete genome, *K. nataicola* RZS01 possesses various transporters for the uptake of substrates and ions. We identified 49 ABC transporters, 10 symporters, and 47 permeases, which are responsible for the transportation of sugars (e.g. glucose and ribose), polyols (e.g. mannitol and sorbitol), sugar acids (e.g. gluconate and acetate), amino acids, purines, pyrimidines, phosphate, sulphate,  $\text{NH}_4^+$ , and metal ions. In addition, some components of the phosphotransferase system are also found. As the most widely used substrate for industrial production, glucose could be taken up by a sugar symporter (B0W47\_14370). Owing to the lack of the gene encoding phosphofructokinase (EC 2.7.1.11), RZS01 has an incomplete Embden-Meyerhof-Parnas pathway, as has been reported in *Acetobacter*

Family	Cofactor	Genes
Glucose dehydrogenase	PQQ	B0W47_01230, B0W47_02520, B0W47_10950
Glycerol dehydrogenase	PQQ	B0W47_01005, B0W47_01010
Alcohol dehydrogenase	PQQ	B0W47_13410
Aldehyde dehydrogenase	PQQ	B0W47_16410
2-Keto-D-gluconate dehydrogenase	FAD	B0W47_11030, B0W47_11035, B0W47_11040
Gluconate 2-dehydrogenase	FAD	B0W47_05395, B0W47_05400, B0W47_05405, B0W47_13405

**Table 3.** Primary membrane-bound dehydrogenases in *K. nataicola* RZS01.

*pasteurianus* 386B<sup>6</sup>, *Gluconacetobacter diazotrophicus* Pal5<sup>39</sup>, and *Gluconobacter oxydans* 621H<sup>2</sup>. However, the genes encoding the enzymes of the pentose-phosphate (PPP) pathway are all found, suggesting that glucose might be degraded via the PPP pathway. Carbohydrates are generally metabolized by dehydrogenases and isomerases after phosphorylation by specific kinases. In addition, RZS01 possesses an orthologous cluster for gluconate utilization similar to the *GntU*, *GntK*, and *GntR* cluster found in *E.coli*<sup>40</sup>. However, RZS01 does not contain a *GntK* gene, which encodes gluconate kinase. As for disaccharides, such as sucrose and lactose, substrates can be utilized after being decomposed to monosaccharides. RZS01 contains levansucrase (B0W47\_08430), which converts sucrose into glucose and fructose for further oxidation, suggesting that this strain can take up sucrose from the environment to use it as a carbon source. Sucrose is able to adjust the osmotic pressure of the medium, which has been verified by Velasco-Bedrán<sup>41</sup>.

Genome analysis illustrated that RZS01 possesses all the genes catalysing metabolic pathways involved in the *de novo* synthesis of all nucleotides, amino acids, phospholipids and many vitamins. Ammonia, which participates in the activity of glutamate synthase (B0W47\_13175 and B0W47\_13180) and glutamine synthase (B0W47\_07980), can be taken up by a specific ammonia transporter (B0W47\_07675).

**Membrane-associated primary dehydrogenases and the respiratory chain.** *K. nataicola* RZS01 contains several genes that encode membrane-associated primary quinone reductases coupling substrate oxidation with quinone reduction, which can be classified into 2 groups (Table 3)<sup>2</sup>. The first group of dehydrogenases depends on the cofactor pyrroloquinoline quinone (PQQ). The most representative dehydrogenases are PQQ-dependent alcohol dehydrogenase (PQQ-ADH) and aldehyde dehydrogenase (PQQ-ALDH), which convert alcohol into acetaldehyde and acetaldehyde into acetate, respectively. The genes encoding PQQ-ADHs are present as 8 copies distributed throughout the chromosomes, and just is localized in the membrane. Moreover, RZS01 possesses 1 membrane-bound PQQ-ALDH, whereas 5 other copies are dispersed in the cytoplasm. Furthermore, the genome of RZS01 contains glucose dehydrogenase and glycerol dehydrogenase.

**Adaptation to extreme conditions.** *K. nataicola* RZS01 possesses the ability to tolerate low pH and high concentrations of organic acid in the environment<sup>42</sup>. One strategy for this tolerance is that RZS01 has a cytosolic acetate-assimilating detoxification pathway, whereby acetate is first converted into acetyl-CoA, which is catalysed by acetyl-CoA synthetase (*acn*; B0W47\_06505). The product is further oxidized by the citric acid cycle to water and carbon dioxide. Overoxidation occurs in *Acetobacter* and *Gluconacetobacter* but not in *Gluconobacter*, which exhibits relatively weak acetic acid resistance<sup>8, 43</sup>. The second mechanism for acid tolerance may involve an ABC transporter gene, *aata* (B0W47\_06270), which functions as an efflux pump of acetic acid<sup>44</sup>. In another similar efflux mechanism, a proton motive force-dependent efflux system occurs in the cytoplasmic membrane, which is capable of exporting acetic acid and is dependent on the proton motive force but not on ATP<sup>45</sup>. Besides, there is also an operon involved in high resistance to acetic acid, a trait that can be a suitable target for enhancement by breeding. This operon is composed of the genes *groE* (B0W47\_13360 and B0W47\_13365), *groS* (B0W47\_00055), and *groL* (B0W47\_02405, B0W47\_03785, and B0W47\_09650), and the overexpression of the *groESL* operon increases acetic acid resistance in *Acetobacter*<sup>46</sup>. A previous thorough genome analysis revealed the presence of a gene cluster encoding a urease (*ureDABCEFG*; B0W47\_16420, B0W47\_16425, B0W47\_16430, B0W47\_16435, B0W47\_16440, B0W47\_16445, and B0W47\_16450), a urea transporter (*urtABCDE*; B0W47\_16475, B0W47\_16480, B0W47\_16485, B0W47\_16490, and B0W47\_16495), and an allophanate hydrolase (B0W47\_14240, B0W47\_14245, B0W47\_16295, and B0W47\_16300). These proteins transport urea and catalyse it to ammonia, which enables the survival of RZS01 in an acidic environment, such as in vinegar fermentation. This mechanism is also present in *A. pasteurianus* 386B and *G. bethesdensis* CGDNIH1 but is not widespread among AAB strains<sup>6, 47</sup>.

## Discussion

*K. nataicola* is widely distributed in nature and displays strong abilities of producing BC and tolerating acetic acid. In this study, we sequenced, annotated, and analysed the complete genome sequence of *K. nataicola* RZS01 and found that RZS01 possesses a 3.5-Mb chromosome and 6 plasmids. The global overview of all genes provided comprehensive insights into the metabolic features, including the uptake of different substrates and intolerance of acidic stress during BC production.

Comparative genome analysis of 13 AAB revealed that the genome of *K. nataicola* RZS01 is most closely related to that of *K. medellinensis* NBRC 3288, with 75.0% shared genes, which is congruent with the phylogenetic

relationship. On the basis of the genomic sequence of RZS01, we can now more extensively describe the process of cellulose biosynthesis and the physiological basis of the underlying pathways in this organism. For other AAB species whose genomes have been sequenced, 2 *bcs* operons are present in RZS01, which may explain the high cellulose productivity observed. However, the regulation of AcsAB activity by c-di-GMP may also play important roles. We identified 3 *cdg* operons containing a DGC and a c-di-GMP PDE. Four standalone c-di-GMP PDEs and 1 standalone DGC are also present in the genome. This regulatory mechanism has also been found in other bacteria<sup>38</sup>.

*K. nataicola* RZS01 contains many membrane-bound dehydrogenases, which are responsible for the assimilation of substrates and contribution to acetic acid resistance, resisting the presence of high concentrations of acetic acid. The metabolism of several amino acids, such as threonine, glycine, and ornithine, produces a large amount of ammonia, which decreases the intracellular pH value. Certain other mechanisms also participate in acetic acid tolerance, including alcohol overoxidation, an acetate assimilation-related mechanism, an ABC transporter, and a proton motive force-dependent efflux system.

In summary, we uncovered global insights into BC production and acidic resistance mechanisms by genome analysis. These results provide useful information for further studies on evolution and genetic variation in *Komagataeibacter* and provide a valuable resource for biological research. Furthermore, comparative genomics analysis and functional genomics analysis can also be performed to trace the origin and evolution of this organism.

## Methods

**Bacterial strain and isolation of chromosomal DNA.** *K. nataicola* RZS01 (CGMCC number 10961) used in this study was isolated from rotten apples<sup>48</sup>. For genomic DNA extraction, RZS01 was grown from a single colony in HZ medium, containing 20 g·L<sup>-1</sup> glucose, 6 g·L<sup>-1</sup> ammonium sulphate, 1 g·L<sup>-1</sup> monopotassium phosphate, 0.4 g·L<sup>-1</sup> magnesium sulphate, 3 g·L<sup>-1</sup> peptone, 2.25 g·L<sup>-1</sup> yeast extract, and 0.4 g·L<sup>-1</sup> carboxymethyl-cellulose sodium at 30 °C for 24 h. The produced cellulose was digested by adding 0.2% cellulase and maintained at 30 °C for 2 h. After the cells were harvested, chromosomal DNA was isolated using the QiaAmp DNA Mini kit (Qiagen, Valencia, Germany) following the manufacturer's instructions. The quality of the DNA was evaluated using a Nanodrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) after isolation.

**Genome sequencing and assembly.** The whole genome of *K. nataicola* RZS01 was sequenced at Hangzhou GeneRui Biotechnology Co., Ltd. (Hangzhou, China) using a PacBio RS DNA sequencer. The average insert size was 550 bp. Data from 515-Mb paired-end reads were delivered by Baseclear as 2 fastq files. The FastQC procedure was used to verify the raw read data, and QC metrics including insert-sizes, mapped reads, unmapped reads and reads that align with a deviated pattern were examined<sup>49</sup>. The sequencing errors were discarded using the error-correction module of Allpaths-LG. For single-molecule real-time sequencing (SMRT), a SMRT bell DNA template library with an insert size of 2~40 kb was prepared. Subsequently, the fragmented DNA was end-repaired and ligated into hairpin adapters. Sequence reads have been deposited in the ENA Sequence Read Archive (EMBL: ERS550016). Sequencing reads were corrected using the HGAP pipeline<sup>50</sup>. Genomic sequences were assembled *de novo* using PacBio analysis RS\_HGAP\_Assembly.3 (Pacific Biosciences, Menlo Park, CA, USA).

**Bioinformatics analysis.** Automatic gene prediction and annotation were performed using Glimmer3.0. The genome was visualized using Circos (<http://circos.ca/tutorials/lessons/>). 16S rRNA phylogeny was constructed by the neighbour-joining method of MEGA5.1 at default settings. Metabolic enzymes were identified from UniProt (<http://www.uniprot.org/>) and BLASTp analysis (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to identify orthologues in the genomes.

**Nucleotide sequence accession numbers.** The accession numbers of the complete sequences of *K. nataicola* RZS01 determined in this study and of the 6 plasmids from NCBI can be found in GenBank (<http://www.ncbi.nlm.nih.gov>) under the accession no. CP019875 to CP019881.

## References

1. Saichana, N., Matsushita, K., Adachi, O., Frébert, I. & Frébertova, J. Acetic acid bacteria: A group of bacteria with versatile biotechnological applications. *Biotechnol. Adv.* **33**, 1260–1271 (2014).
2. Prust, C. *et al.* Complete genome sequence of the acetic acid bacterium *Gluconobacter oxydans*. *Nat. Biotechnol.* **23**, 195–200 (2005).
3. Adachi, O. *et al.* New developments in oxidative fermentation. *Appl. Microbiol. Biotechnol.* **60**, 643–653 (2003).
4. Lazarini, S. C. *et al.* Characterization of bilayer bacterial cellulose membranes with different fiber densities: a promising system for controlled release of the antibiotic ceftriaxone. *Cellulose* **23**, 737–748 (2016).
5. Liu, M. *et al.* Metabolomic profiling coupled with metabolic network reveals differences in *Gluconacetobacter xylinus* from static and agitated cultures. *Biochem. Eng. J.* **101**, 85–98 (2015).
6. Illegheems, K., Vuyst, L. D. & Weckx, S. Complete genome sequence and comparative analysis of *Acetobacter pasteurianus* 386B, a strain well-adapted to the cocoa bean fermentation ecosystem. *BMC Genomics* **14**, 42–45 (2013).
7. Battey, A. S. & Schaffner, D. W. Modelling bacterial spoilage in cold-filled ready to drink beverages by *Acinetobacter calcoaceticus* and *Gluconobacter oxydans*. *J. Appl. Microbiol.* **91**, 237–247 (2001).
8. Kersters, K., Lisdiyanti, P., Komagata, K. & Swings, J. The Family *Acetobacteraceae*: The genera *Acetobacter*, *Acidomonas*, *Asaia*, *Gluconacetobacter*, *Gluconobacter*, and *Kozakia*. (Springer New York, 2006).
9. Yamada, Y., Hoshino, K. & Ishikawa, T. The phylogeny of acetic acid bacteria based on the partial sequences of 16S ribosomal RNA: The elevation of the subgenus to the generic level. *Biosci. Biotechnol. Biochem.* **61**, 1244–1251 (1997).
10. Yamada, Y. *et al.* Subdivision of the genus *Gluconacetobacter* Yamada, Hoshino and Ishikawa 1998: the proposal of *Komagataeibacter* gen. nov., for strains accommodated to the *Gluconacetobacter xylinus* group in the  $\alpha$ -Proteobacteria. *Ann. Microbiol.* **62**, 849–859 (2012).

11. Yamada, Y. *et al.* Description of *Komagataeibacter* gen. nov., with proposals of new combinations (*Acetobacteraceae*). *J. Gen. Appl. Microbiol.* **58**, 397–404 (2012).
12. Santos, R. A. C. D. *et al.* Draft genome sequence of *Komagataeibacter intermedius* strain AF2, a producer of cellulose, isolated from kombucha tea. *Genome Announcements* **3**, 1–2 (2015).
13. Santos, R. A. C. D. *et al.* Draft genome sequence of *Komagataeibacter rhaeticus* strain AF1, a high producer of cellulose, isolated from kombucha tea. *Genome Announcements* **2**, 1–2 (2014).
14. Mamlouk, D. & Gullo, M. Acetic acid bacteria: Physiology and carbon sources oxidation. *Indian J. Microbiol.* **53**, 377–384 (2013).
15. Ogino, H. *et al.* Complete genome sequence of NBRC 3288, a unique cellulose-nonproducing strain of *Gluconacetobacter xylinus* isolated from vinegar. *J. Bacteriol.* **193**, 6997–6998 (2011).
16. Lee, K. Y., Buldum, G., Mantalaris, A. & Bismarck, A. More Than Meets the Eye in Bacterial Cellulose: Biosynthesis, Bioprocessing, and Applications in Advanced Fiber Composites †. *Macromol. Biosci.* **14**, 10–32 (2014).
17. Sun, S. *et al.* Comparison of the mechanical properties of cellulose and starch films. *Biomacromolecules* **11**, 126–132 (2009).
18. Na, Y., Stilwell, M. D., Santos, T. M. A., Wang, H. & Weibel, D. B. Agarose particle-templated porous bacterial cellulose and its application in cartilage growth *in vitro*. *Acta Biomater.* **12**, 129–138 (2014).
19. Klemm, D., Schumann, D., Udhardt, U. & Marsch, S. Bacterial synthesized cellulose—artificial blood vessels for microsurgery. *Prog Polym Sci. Prog. Polym. Sci.* **26**, 1561–1603 (2001).
20. Yadav, V. *et al.* Novel *in vivo*-degradable cellulose-chitin copolymer from metabolically engineered *Gluconacetobacter xylinus*. *Appl. Environ. Microbiol.* **76**, 6257–6265 (2010).
21. Ummartyotin, S., Juntaro, J., Sain, M. & Manuspiya, H. Development of transparent bacterial cellulose nanocomposite film as substrate for flexible organic light emitting diode (OLED) display. *Ind. Crops Prod.* **35**, 92–97 (2012).
22. Mehta, K., Pfeffer, S. & Malcolm Brown, R. Jr. Characterization of an *acsD* disruption mutant provides additional evidence for the hierarchical cell-directed self-assembly of cellulose in *Gluconacetobacter xylinus*. *Cellulose* **22**, 119–137 (2015).
23. Zhong, C. *et al.* Revealing differences in metabolic flux distributions between a mutant strain and its parent strain *Gluconacetobacter xylinus* CGMCC 2955. *PLoS One* **9**, e98772 (2014).
24. Kondo, K. & Horinouchi, S. Characterization of an insertion sequence, IS12528, from *Gluconobacter suboxydans*. *Appl. Environ. Microbiol.* **63**, 1139–1142 (1997).
25. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Sci* **5331**, 1453–1462 (2010).
26. Valera, M. J., Torija, M. J., Mas, A. & Mateo, E. Cellulose production and cellulose synthase gene detection in acetic acid bacteria. *Appl. Microbiol. Biotechnol.* **99**, 1–13 (2015).
27. Aydın, Y. A. & Aksoy, N. D. Isolation and characterization of an efficient bacterial cellulose producer strain in agitated culture: *Gluconacetobacter hansenii* P2A. *Appl. Microbiol. Biotechnol.* **98**, 1065–1075 (2014).
28. Castro, C. *et al.* *Gluconacetobacter medellinensis* sp. nov., cellulose- and non-cellulose-producing acetic acid bacteria isolated from vinegar. *Int. J. Syst. Evol. Microbiol.* **63**, 1119–1125 (2013).
29. Chawla, P. R., Bajaj, I. B., Survase, S. A. & Singhal, R. S. Microbial cellulose: fermentative production and applications. *Food Technol. Biotechnol.* **47**, 107–124 (2009).
30. Römling, U. Molecular biology of cellulose production in bacteria. *Res. Microbiol.* **153**, 205–212 (2002).
31. Nakai, T. *et al.* Formation of highly twisted ribbons in a carboxymethylcellulase gene- disrupted strain of a cellulose-producing bacterium. *J. Bacteriol.* **195**, 958–964 (2013).
32. Morgan, J. L., Strumillo, J. & Zimmer, J. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nat.* **493**, 181–186 (2012).
33. Ojinnaka, C. *et al.* Structure and conformation of acetan polysaccharide. *Int. J. Biol. Macromol.* **19**, 149–156 (1996).
34. Griffin, A. M., Morris, V. J. & Gasson, M. J. Identification, cloning and sequencing the *aceA* gene involved in acetan biosynthesis in *Acetobacter xylinum*. *FEMS Microbiol. Lett.* **137**, 115–121 (1996).
35. Colquhoun, I. J. *et al.* Structure and conformation of a novel genetically engineered polysaccharide P2. *Carbohydrate Research* **330**, 325–333 (2001).
36. Ross, P. *et al.* The cyclic diguanylic acid regulatory system of cellulose synthesis in *Acetobacter xylinum*. Chemical synthesis and biological activity of cyclic nucleotide dimer, trimer, and phosphothioate derivatives. *J. Biol. Chem.* **265**, 18933–18943 (1990).
37. Ross, P., Mayer, R. & Benziman, M. Cellulose biosynthesis and function in bacteria. *Microbiol. Rev.* **55**, 35–58 (1991).
38. Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nat. Rev. Microbiol.* **7**, 263–273 (2009).
39. Marcelo, B. *et al.* Complete genome sequence of the sugarcane nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus* Pa15. *BMC Genomics* **10**, 43–50 (2009).
40. Tong, S., Porco, A., Isturiz, T. & Conway, T. Cloning and molecular genetic characterization of the *Escherichia coli* *gntR*, *gntK*, and *gntU* genes of *GntI*, the main system for gluconate metabolism. *J. Bacteriol.* **178**, 3260–3269 (1996).
41. Velasco-Bedrán, H. & López-Isunza, F. The unified metabolism of *Gluconacetobacter entanii* in continuous and batch processes. *Process Biochem.* **42**, 1180–1190 (2007).
42. Matsushita, K. *et al.* Acetic acid production in acetic acid bacteria leading to their 'death' and survival. In: Yamada M (ed) *Survival and death in bacteria* 169–181 (Research Signpost, 2005).
43. Garrity, G. M. *Bergey's manual of systematic bacteriology* (Springer, 2011).
44. Nakano, S., Fukaya, M. & Horinouchi, S. Putative ABC transporter responsible for acetic acid resistance in *Acetobacter aceti*. *Appl. Environ. Microbiol.* **72**, 497–505 (2006).
45. Matsushita, K., Inoue, T., Adachi, O. & Toyama, H. *Acetobacter aceti* possesses a proton motive force-dependent efflux system for acetic acid. *J. Bacteriol.* **187**, 4346–4352 (2005).
46. Nakano, S. & Fukaya, M. Analysis of proteins responsive to acetic acid in *Acetobacter*: molecular mechanisms conferring acetic acid resistance in acetic acid bacteria. *Int. J. Food Microbiol.* **125**, 54–59 (2008).
47. Greenberg, D. *et al.* Genome sequence analysis of the emerging human pathogenic acetic acid bacterium *Granulibacter bethesdensis*. *J. Bacteriol.* **189**, 8727–8736 (2008).
48. Zhou, L. L., Sun, D. P., Hu, L. Y., Li, Y. W. & Yang, J. Z. Effect of addition of sodium alginate on bacterial cellulose production by *Acetobacter xylinum*. *J. Ind. Microbiol. Biotechnol.* **34**, 483–489 (2007).
49. Andrews, S. FQC: A quality control tool for high throughput sequence data. <http://www.Bioinformaticsabraham.ac.uk/projects/fastqc/> (2010).
50. Chin, C. S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 21206076), Natural Science Foundation of Jiangsu Province (No. BK2012401), the Fundamental Research Funds for the Central Universities (No. 30920130121001), the Science and Technology Innovation Fund of QiXia District (201528) and a project funded by the Surface Program of Natural Science Research of Jiangsu Higher Education Institutions (No. 16KJB180034).

### Author Contributions

X.X. and B.S. contributed to the annotation of the genome; X.C. and F.Y. helped to analyse the genomic information; H.Z. and Y.X. wrote the first draft of the manuscript; and J.Y. and D.S. contributed substantially to revisions.

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-04589-6](https://doi.org/10.1038/s41598-017-04589-6)

**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.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017