

SCIENTIFIC REPORTS



OPEN

Insights into mutualism mechanism and versatile metabolism of *Ketogulonicigenium vulgare* Hbe602 based on comparative genomics and metabolomics studies

Nan Jia^{1,2}, Ming-Zhu Ding^{1,2}, Jin Du^{1,2}, Cai-Hui Pan^{1,2}, Geng Tian³, Ji-Dong Lang³, Jian-Huo Fang³, Feng Gao^{1,2,4} & Ying-Jin Yuan^{1,2}

Ketogulonicigenium vulgare has been widely used in vitamin C two steps fermentation and requires companion strain for optimal growth. However, the understanding of *K. vulgare* as well as its companion strain is still preliminary. Here, the complete genome of *K. vulgare* Hbe602 was deciphered to provide insight into the symbiosis mechanism and the versatile metabolism. *K. vulgare* contains the LuxR family proteins, chemokine proteins, flagellar structure proteins, peptides and transporters for symbiosis consortium. Besides, the growth state and metabolite variation of *K. vulgare* were observed when five carbohydrates (D-sorbitol, L-sorbose, D-glucose, D-fructose and D-mannitol) were used as carbon source. The growth increased by 40.72% and 62.97% respectively when *K. vulgare* was cultured on D-mannitol/D-sorbitol than on L-sorbose. The insufficient metabolism of carbohydrates, amino acids and vitamins is the main reason for the slow growth of *K. vulgare*. The combined analysis of genomics and metabolomics indicated that TCA cycle, amino acid and nucleotide metabolism were significantly up-regulated when *K. vulgare* was cultured on the D-mannitol/D-sorbitol, which facilitated the better growth. The present study would be helpful to further understand its metabolic structure and guide the engineering transformation.

K. vulgare was identified as a member of the *Proteobacteria*¹, which can convert L-sorbose to the precursor of vitamin C, 2-keto-L-gulonic acid (2-KGA)². Currently, the genomes of two typical strains in the genus *K. vulgare* have been sequenced completely^{3,4}. However, the understanding of *K. vulgare* is still preliminary partly because of its unique nature. It was found that the *K. vulgare* mono-culture system grew poorly, even on rich natural media such as yeast extracts, peptones and corn steep liquor. The *Bacillus* species are usually co-cultivated with *K. vulgare* to achieve a better growth rate of *K. vulgare* and an adequate 2-KGA yield⁵. Addition of amino acids³, vitamins⁶ and glutathione⁷ could enhance the growth of *K. vulgare*, implying the defects of sulfur and oxidation metabolism, vitamin and amino acid synthesis. However, further research is needed to know the reason for the weak growth of *K. vulgare* and how to adapt its growth state in symbiosis system.

Using GC-TOF/MS technology, metabolomics was suitable for analyzing the variations in primary metabolites, such as intermediates in central carbon metabolism and amino acid biosynthesis. The metabolomics approach has been demonstrated in our previous studies on the interactions between *K. vulgare* and *Bacillus* strain⁸. Here we combine comparative genomics and metabolomics to explore the mechanism behind the positive effects of different carbon sources. Previous research showed that heterologously expression of folate biosynthesis from *Lactococcus lactis* in *K. vulgare* could enhance the growth and 2-KGA production⁹, and the present study would be helpful for further analysis of the metabolism network and identification of the genetic targets for strain improvement.

¹Key Laboratory of Systems Bioengineering (Ministry of Education), School of Chemical Engineering and Technology, Tianjin University, Tianjin, 300072, PR China. ²SynBio Research Platform, Collaborative Innovation Centre of Chemical Science and Engineering (Tianjin), School of Chemical Engineering and Technology, Tianjin University, Tianjin, 300072, PR China. ³Sequencing platform of Tsinghua University, Beijing, 100084, PR China. ⁴Department of Physics, Tianjin University, Tianjin, 300072, PR China. Correspondence and requests for materials should be addressed to M.-Z.D. (email: mzding@tju.edu.cn) or F.G. (email: fgao@tju.edu.cn)

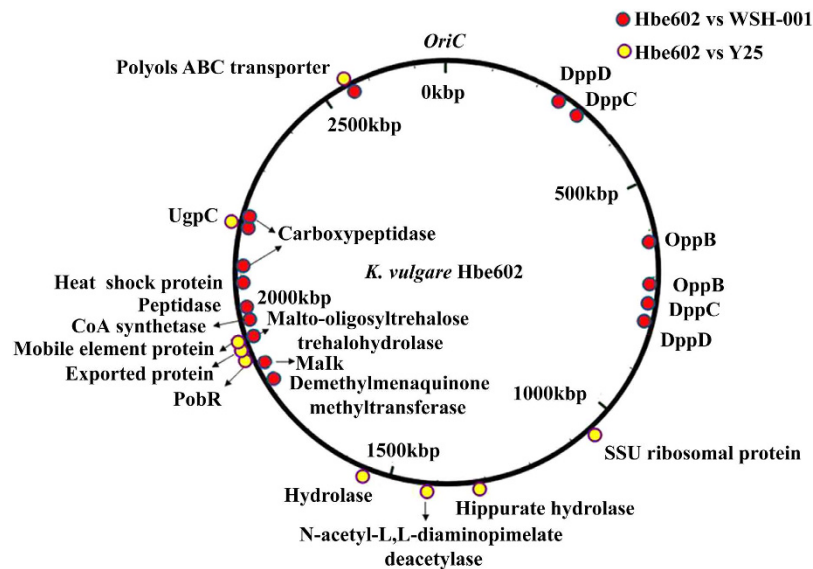


Figure 1. Comparative genomic analysis of the difference among *K. vulgare* Hbe602, WSH-001 and Y25. The chromosome of *K. vulgare* Hbe602 was used in comparison, the differences with WSH-001 and Y25 were labeled in the red dots and yellow dots, respectively.

Results

General genomic properties of *K. vulgare* Hbe602. The genome of *K. vulgare* Hbe602 consists of one circular chromosome and two plasmids, which encodes 3,178 predicted proteins, 58 tRNAs and 15 rRNAs in total (Fig. S1 and Table S1). For the circular chromosome, the predicted replication origin is located at 2,765,502–281 nt, which is closely next a *maf* gene (282–887 nt) encoding septum formation protein Maf. The analysis of replication origins for bacteria in DoriC, a database of *oriC* regions in bacterial and archaeal genomes^{10,11}, has also shown the feature that *oriC* region is adjacent to a *maf* gene is conserved in the family *Rhodobacteraceae*. For the plasmids, the putative origins of replication are also at 0 kb, and both contain repeat sequences. Phylogenetic analysis of *K. vulgare* Hbe602 with other species could provide putative evolutionary histories and phenotypic diversity (Fig. S2). To date, *K. vulgare* Hbe602 revealed the high homogeneity with other *K. vulgare* species (Y25 and WSH-001), which was related closely to *Rhodobacter* and *Paracoccus*. Furthermore, the genome-scale sequence comparison by LAST software available at <http://last.cbrc.jp> shows the genome similarity between *K. vulgare* Hbe602 with WSH-001 is higher than that with Y25 (Fig. S3)¹². Through the whole genome comparison, the different genes with the nucleotide identities lower than 90% are obtained, mostly focused on proteolytic enzymes and transporters (Fig. 1). Previously, *K. vulgare* was mistakenly regarded as a member of *Gluconobacter oxydans* due to the similar phenotype^{1,13}. To facilitate the gene function analysis, *G. oxydans* 621H, *Escherichia coli* K-12 and *R. sphaeroides* 2.4.1 are selected to compare the distribution of COG classification with *K. vulgare* Hbe602 (Fig. S4 and Table S2). Consequently, in *K. vulgare*, the number of genes related to amino acid transport and metabolism (E) is similar to those in *E. coli* and *R. sphaeroides* and higher than that in *G. oxydans*. Those functional genes encode many transporters, which absorb the nutrients to compensate its metabolic defect. Besides, the number of transcriptional regulation related genes (K) is even more than that in *E. coli*. The strong transcriptional regulation ability may be activated to form a symbiotic relationship with other microorganisms. With the aid of KEGG analysis, metabolic network is obtained, including the carbohydrate-active enzymes, nitrogen and central carbon metabolism, amino acid and cofactor metabolism.

Symbiosis mechanism related genes. The quorum sensing phenomenon has been well established in Gram-negative bacteria, where N-acyl homoserine lactones are the diffusible communication molecules that modulate cell-density-dependent phenotypes¹⁴. Genomic analysis of 265 *Proteobacteria* species shows that at least 68 species contain LuxI (AHL synthase) and LuxR (transcriptional activator), while 45 species only contain LuxR¹⁵. *K. vulgare* only contains LuxR and lacks LuxI, so it may produce AHL by other means, or obtains the AHL signaling molecules from other microorganisms¹⁶. The LuxR family protein is the hub of quorum sensing system, which regulates many kinds of proteins and thus affects the global quorum sensing. *K. vulgare* also contains phosphorylated protein CtrA taking part in flagella mobile and cell division, which is widely present in the alpha-*Proteobacteria*¹⁷. Besides, *K. vulgare* contains two HK97 family proteases, which are generally encoded by phage. They were also found in *R. capsulatus* as gene transfer agent, which can mediate gene transfer¹⁸. Since *K. vulgare* and *R. capsulatus* are the neighbor strains, the HK97 family protease in *K. vulgare* may not be used for the degradation of protein, but the gene transfer agent. Furthermore, *K. vulgare* Hbe602 contains several chemotaxis proteins (Table S3), including a Che cluster of CheA, CheB, CheD, CheR, CheW and CheX, to respond to the environmental variation. We also found a large cluster with 37 flagella-related genes (Table S4), which operated flagella structure synthesis and flagella motility. The ability of chemotaxis and motility in *K. vulgare* Hbe602 is connected with its natural soil

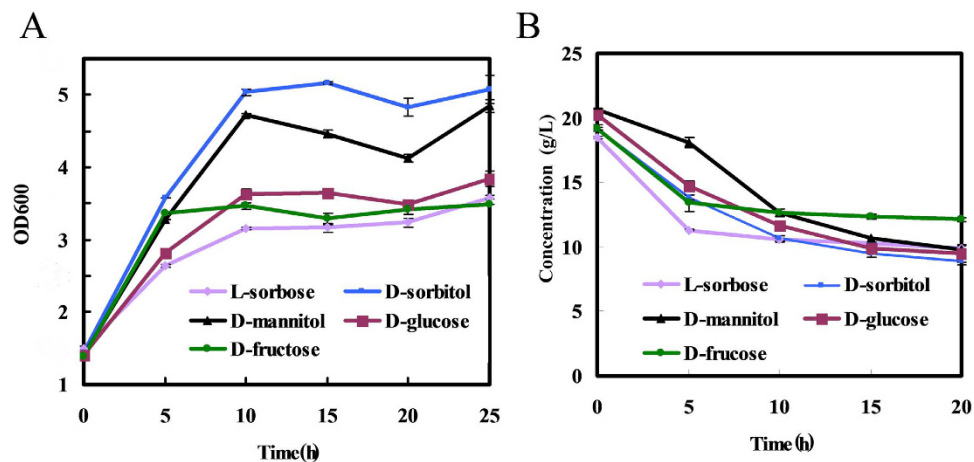


Figure 2. Growth state of *K. vulgare* Hbe602 in different carbon sources. (A) Growth curve the *K. vulgare* Hbe602 strain grown in seed medium with different carbon source (L-sorbose, D-sorbitol, D-mannitol, D-glucose and D-fructose). The Y axis represents the average OD_{600nm} at each time point; (B) Extracellular concentration of different carbohydrates.

environment, and is used for inducing the optimal symbiotic relationship. Moreover, *K. vulgare* Hbe602 encodes 18 predicted proteases, 48 peptides and almost 380 transporter related proteins, which can effectively decompose the related substances from the cultured environment or companion strain¹⁹.

Metabolomics analysis of the versatile metabolism in *K. vulgare* Hbe602. Due to the limited understanding of genome and metabolic network, the role of L-sorbose in the bio-conversion for *K. vulgare* is not clear. We try to add different carbon sources (L-sorbose, D-sorbitol, D-mannitol, D-fructose and D-glucose) for observing the growth phenotype and analyzing the key genes. The growth conditions of *K. vulgare* varied dramatically on five carbon sources and the highest growth value with D-mannitol was about 1.63 times than that with L-sorbose (Fig. 2). Metabolomics analysis was carried out to compare the metabolome of *K. vulgare* cultured on five carbon sources. A total of 54 metabolites were identified, involving the central carbon metabolism, amino acid and lipid biosynthesis. Multivariate data analysis was used to identify the bio-markers in response to the environment changes. The cells cultured in sorbitol and mannitol were clustered together at a short distance from those cultured in sorbose (Fig. 3A,C). They may have the similar way to improve the growth of *K. vulgare*. From the distances of the metabolites away from the origin in PCA loading plot, many metabolites affecting primary metabolism were responsible for the differences in the samples, including serine, lysine, nicotinic acid, 2-keto-D-gluconic acid, xylitol and mannonic acid (Fig. 3B). Fifty-four metabolites were categorized into 4 clusters based on the variable similarity patterns (Fig. S5). Twenty-two metabolites (clusters 1 and 4) maintained at a higher level in the cells cultured in sorbitol/manntiol than those in sorbose. Ten metabolites (cluster 2) accumulated mostly in sorbose and 7 metabolites (cluster 3) accumulated mostly in sorbitol.

It is interesting to remark that *K. vulgare* Hbe602 preformed a better growth on seed medium supplied with D-sorbitol/D-mannitol compared with L-sorbose. Actually, sorbose and 2-KGA were detected in the cells when sorbitol was used as substrate, and mannoic acid was detected when mannitol was used as substrate (Fig. 4A). *G. oxydans* could convert D-glucose to 2-/5-keto-gluconic acid²⁰. We observed the related acid product when *K. vulgare* was cultured in glucose, fructose and mannitol as well (Fig. 4A). *K. vulgare* lacks many amino acid synthesis pathways (Fig. S6) and we only found 10 amino acids in the cells (Fig. 4B). Among them, valine, leucine, proline, glycine, threonine and 5-oxo-proline were biomass relevant, which maintained at a higher level in the cells cultured in sorbitol/manntiol compared with those in sorbose. Serine was maintained at a significant high level when cells cultivated in sorbitol. Glycine, proline, serine and threonine could be converted into intermediates of the tricarboxylic acid (TCA) cycle and used for generating energy or other components. Besides, the levels of valine, proline, glycine and serine were extremely higher in the cells receiving GSH²¹. Supplement of glycine, proline and threonine led to 29.1%, 24.9% and 28.1% increased in the growth of *K. vulgare*, respectively³. In addition, proline has acted as osmolarity protective agent, which greatly accumulated in sorbitol/manntiol for anti-stress protection²². Four important intermediates in TCA, including citric acid, octadecanoic acid, succinic acid and fumaric acid were identified, and maintained higher levels in the cells cultured in sorbitol/manntiol than those in sorbose (Fig. 4C). The urea concentration in the cells cultured in sorbitol was 1.5 times of those in sorbose, and nicotinic acid, pyrimidine and adenosine represented higher concentration when the cells cultured in sorbitol/manntiol than in sorbose (Fig. 4D).

Discussion

It has always been an important subject for enhancing the growth ability of *K. vulgare*. In the present study, the metabolic network of *K. vulgare* was obtained by KEGG analysis, and the environmental response in different carbon sources (L-sorbose, D-sorbitol, D-mannitol, D-fructose and D-glucose) was analyzed by metabolomics studies. The genomics analysis found that *K. vulgare* lacks the endoenzyme, which transforms D-mannitol, L-sorbose, D-sorbitol and D-glucose to D-fructose (Fig. S7). The different sugars may be obtained through a large

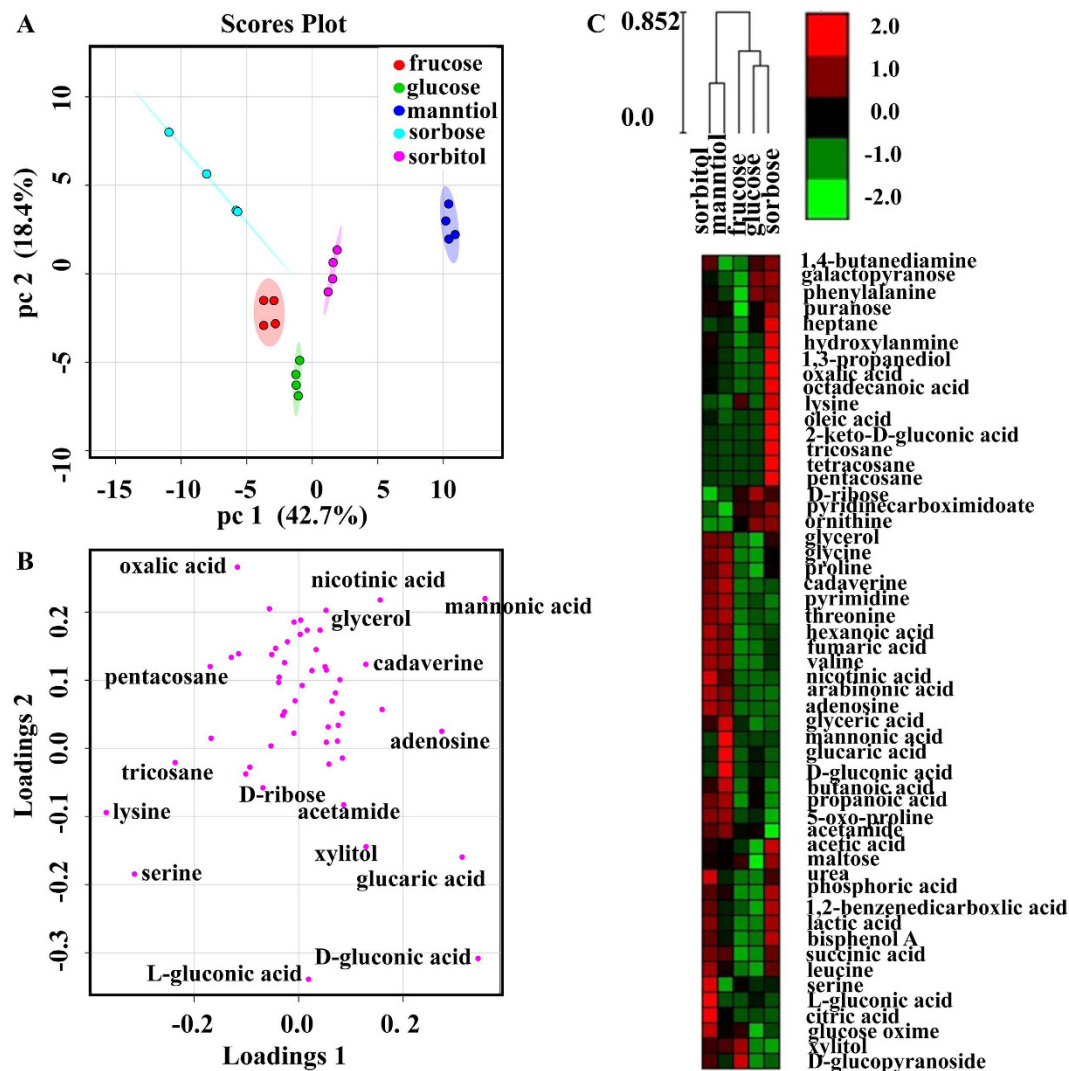


Figure 3. Statistics analysis of the metabolite distribution among different samples. (A) Scores plot of the samples; (B) Loading plot of the samples; (C) Heat map of metabolite expressions in different samples.

number of trans-membrane proteins, and different combinations of L-sorbose dehydrogenases and L-sorbosone dehydrogenases from *K. vulgare* were introduced into *G. oxydans* to converse L-sorbose to 2-KGA²³. Additionally, the digestive enzymes of glycoside were found, such as glycoside hydrolase, beta-glucosidase and glycosyl hydrolase, involving in the substance utilization.

In the nitrogen metabolism of *K. vulgare*, we found the urease and four urea transporters, which can hydrolyze urea and release ammonia for re-utilizing. Besides, we also found the ammonia transporter and carbamoyl-phosphate synthase for absorbing foreign ammonia, and nitrogenase for nitrogen fixation. The nitrogen metabolism regulated protein NtrC, NtrY and NtrX were encoded by a gene cluster. NtrC combined with the promoter region of *glnB* to activate *glnB/glnA* under nitrogen-deficient conditions²⁴ and regulated the *ntrY* and *ntrX*²⁵. The urea concentration within the cells cultured in sorbitol was 1.5 times of those in sorbose medium, providing enough substrate for amino synthesis and pH adjustment²⁶.

In the central carbon metabolism of *K. vulgare*, we identified the complete TCA cycle pathway and pentose phosphate pathway. Twelve C4-dicarboxylate transporters used ionic electrochemical potential gradient to absorb C4-dicarboxylic acid substances to compensate the TCA cycle²⁷. The key gene encoding 6-phosphofruco-2-kinase is absent in the glycolytic pathway. Therefore, the pentose phosphate pathway (PPP) is used as the major central carbohydrate metabolism pathway to link up the TCA cycle.

In the cofactor and nucleotide metabolism of *K. vulgare*, metabolism pathway of thiamine, biotin and porphyrin, biosynthesis pathway of folate and ubiquinone are not complete. The companion strain has complete B-family vitamin synthesis pathways, which can complement *K. vulgare*¹⁹. The cluster *pqqABCDEN* of the pyrroloquinoline quinone (PQQ) biosynthesis pathway was isolated and over-expressed in *K. vulgare* Hbe602²⁸. The *K. vulgare* mutant harboring a plasmid with the complete *pqqABCDE* cluster achieved 3-folds higher level of PQQ biosynthesis than the wild-type. PqqN was annotated as a channel protein based on the best blast hit with the bestrophin in *Shinella* strain²⁹, and co-expression of *pqqN* and *pqqABCDE* decreased the PQQ level²⁸. Furthermore, the cell

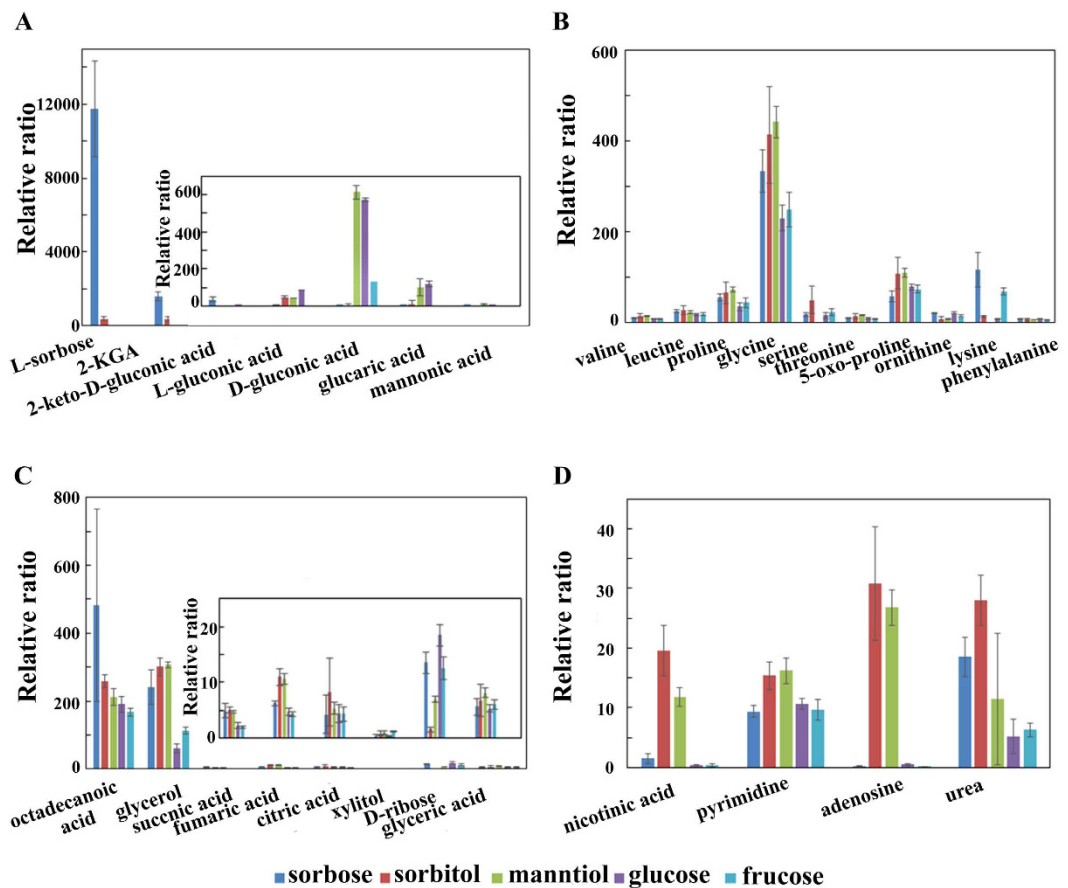


Figure 4. Changes of detected intermediates in different carbon sources. (A) Related acid production; (B) Glycolysis and tricarboxylic acid cycle; (C) Amino acids; (D) Others. The relative abundance was calculated by normalization of peak area of each metabolite to internal standard, and the error bars showed the standard deviations of four replicates.

lysis of companion strain provided purine for *K. vulgare*, and the addition of purine improved growth and 2-KGA production of *K. vulgare*³⁰. From the proteomic study of experimental evolution in previous study, we found the biosynthesis of purine and pyrimidine was improved in the evolved *K. vulgare* in the co-culture³¹. Nicotinic acid, pyrimidine and adenosine were really high concentration in the cultured cells of sorbitol/mannitol than those of sorbose, which could be used as an alternative source to participate or adjust the related metabolic reactions³².

In the amino acid metabolism of *K. vulgare*, three reactions during the conversion of succinate to glutamic acid are defective. In the methionine cycle, *K. vulgare* converts methionine to S-Adenosylmethionine (SAM), and then transforms to S-adenosyl homocysteine cysteine and homocysteine, again into methionine. The last step of the methionine cycle reaction need N5-methyl-tetrahydrofolate as the methyl transfer agent, while tetrahydrofolic acid synthetic pathway is defective in *K. vulgare*, which may stop the normal circulation³³. In the tryptophan metabolism pathway, *K. vulgare* only contains some isolated reactions and the conversion steps for tryptophan to acetyl CoA/FAD/FADH₂ are absent. Besides, *K. vulgare* is absent of the fatty acid metabolism pathway so that it can't obtain acetyl CoA directly. The concentration of acetyl CoA significantly contributed to the formation of citric acid in TCA cycle, which can be used not only as an electron acceptor, but also as a precursor for amino acid biosynthesis and ATP production. Hence, the lack of tryptophan metabolism pathway severely affected the growth of *K. vulgare*. Furthermore, *K. vulgare* harbors the complete valine, leucine and isoleucine biosynthesis pathways, but only contains several isolated reactions in the related degradation pathway. The lack of degradation ability may lead to the amino acid accumulation in *K. vulgare*, providing extracellular amino acids to companion strain in consortium³⁴.

Methods

Cultivation and DNA extraction. The *K. vulgare* Hbe602 strain was cultured in 250 mL flasks with 50 mL of seed medium (30 °C, 250 rpm) supplied with 2% carbon source for 35 h. The seed medium contains 3 g/L beef extract, 3 g/L yeast powder, 3 g/L corn steep liquor, 0.2 g/L MgSO₄, 1 g/L KH₂PO₄, 1 g/L urea and 10 g/L peptone. The genome sample was carried out by CTAB/NaCl extraction³⁵. The quality of the DNA was assessed by spectrophotometer and gel electrophoresis. DNA samples with a 260/280 nm absorbance ratio of 1.8–2.0 and a 260/230 nm absorbance ratio of 2.0–2.2 were considered pure and then used for the library construction and sequencing.

Determination of 2-KGA and different substrates. 2-KGA and different substrates were detected by HPLC using an Aminex HPX-87H column (Bio-Rad, Hercules, CA, USA) coupled with a refractive index (RI) detector. 5 mM H₂SO₄ was used as the mobile phase at a speed of 0.6 mL/min and the RI temperature was kept at 65 °C.

Sequencing and assembly. Using the 454 GS FLX system, single-end libraries with 15-fold coverage and mate paired-end libraries with 10-fold coverage were constructed. The genome was sequenced by the Sanger shotgun approach. The reads were assembled into 6 contigs by using the 454 Newbler assembler and the gaps between the contigs were closed by PCR amplification.

Genome annotation and bioinformatics analysis. The genome analysis was carried out by the Rapid Annotation using Subsystems Technology (RAST) analysis platform³⁶. The function of genes was also annotated by using BLAST³⁷ against Kyoto Encyclopedia of Genes and Genomes (KEGG) database³⁸ and Clusters of Orthologous Groups of proteins (COG) database³⁹. The tRNAs and rRNAs were predicted by tRNAscan-SE⁴⁰ and RNAmmer⁴¹. The subcellular location of proteins and the signal peptides were predicted by PSORT⁴² and SignalP 4.0⁴³. The origin of replication (*oriC*) and putative DnaA boxes were identified by Ori-Finder⁴⁴. CVTree was performed for the phylogenetic analysis⁴⁵ and the phylogenetic tree was generated using the MEGA program⁴⁶. The GC-Profile was used to compute the GC content variation in DNA sequences and to predict the horizontal gene transfer⁴⁷. CGView Server was used for the visualization of circular genomes⁴⁸ and the metabolic network was constructed by KEGG automatic annotation server KAAS⁴⁹.

Nucleotide sequence accession numbers. The sequence of the *K. vulgare* Hbe602 genome has been deposited at DDBJ/EMBL/GenBank under the accession numbers CP012908, CP012909 and CP012910.

Metabolites extraction and derivatization. Cells cultured in different carbon source of logarithmic growth phase were quenched and extracted as intracellular metabolites according to our previous method⁵⁰. An extra group of quenched cells was washed and dried to calculate the dry weight of the sampled cells. The 10 µL succinic d₄ acid (0.1 mg/mL) was used as an internal standard to correct for minor variations occurring during sample preparation and analysis. The extracts of intracellular were lyophilized and four independent experiments were performed for each sample. Two-stage chemical derivatization was performed as described previously⁵⁰. Firstly, methoximation of the carbonyl groups was carried out by dissolving sample in 50 µL methoxamine hydrochloride (20 mg/mL in pyridine) and incubating it at 40 °C for 60 min. Then, 80 µL N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) was added and it was incubated at 37 °C for 30 min for trimethylsilylation.

Metabolomic analysis by GC-TOF/MS. Metabolites were analyzed by GC-TOF/MS (Waters Corp., USA) as described previously⁵⁰. The 1 µL derivatized sample was injected by Agilent 7683 autosampler into GC (Agilent 6890) which was equipped with DB-5MS column (30 m × 0.25 mm × 0.25 µm, J&W Scientific, Folsom, CA). The oven temperature was programmed as: 70 °C for 2 min, then increased to 290 °C (5 °C/min), holding for 3 min. The ion source temperature and ionization current were 250 °C and 40 µA, respectively. The mass scan range was 50–800 m/z. Peak detection, deconvolution, and peak quantification were performed using Masslynx software 4.1⁵¹. Metabolites were identified by comparing their mass fragmentation patterns with NIST mass spectral library⁵². The area of each acquired peak was normalized against the internal standard and dry cell weight for calculating the relative abundance. Multivariate data analysis was performed by principal-components analysis (PCA)⁵³ and hierarchical cluster analysis (HCA)⁵⁴ to view the relative differences in the metabolites concentrations among diverse conditions.

References

1. Urbance, J., Bratina, B., Stoddard, S. & Schmidt, T. Taxonomic characterization of *Ketogulonigenium vulgare* gen. nov., sp. nov. and *Ketogulonigenium robustum* sp. nov., which oxidize L-sorbose to 2-keto-L-gulonic acid. *Int. J. Syst. Evol. Microbiol.* **51**, 1059–1070 (2001).
2. Yin, G. *et al.* Production of vitamin C precursor-2-keto-L-gulonic acid from L-sorbose by a novel bacterial component system of SCB329-SCB933. *Ind. Microbiol.* **27**, 1–7 (1997).
3. Liu, L. M., Chen, K. J., Zhang, J., Liu, J. & Chen, J. Gelatin enhances 2-keto-L-gulonic acid production based on *Ketogulonigenium vulgare* genome annotation. *J. Biotechnol.* **156**, 182–187 (2011).
4. Xiong, X. H. *et al.* Complete genome sequence of the bacterium *Ketogulonigenium vulgare* Y25. *J. Bacteriol.* **193**, 315–316 (2011).
5. Takagi, Y., Sugisawa, T. & Hoshino, T. Continuous 2-Keto-L-gulonic acid fermentation by mixed culture of *Ketogulonigenium vulgare* DSM 4025 and *Bacillus megaterium* or *Xanthomonas maltophilia*. *Appl. Microbiol. Biotechnol.* **86**, 469–480 (2010).
6. Fan, S. *et al.* Development of a minimal chemically defined medium for *Ketogulonigenium vulgare* WSH001 based on its genome-scale metabolic model. *J. Biotechnol.* **169**, 15–22 (2014).
7. Ma, Q. *et al.* Proteomic analysis of *Ketogulonigenium vulgare* under glutathione reveals high demand for thiamin transport and antioxidant protection. *PLoS one.* **7**, e32156 (2012).
8. Du, J., Zhou, J., Xue, J., Song, H. & Yuan, Y. J. Metabolomic profiling elucidates community dynamics of the *Ketogulonigenium vulgare*-*Bacillus megaterium* consortium. *Metabolomics.* **8**, 960–973 (2012).
9. Cai, L. *et al.* Genetic engineering of *Ketogulonigenium vulgare* for enhanced production of 2-keto-L-gulonic acid. *J. Biotechnol.* **157**, 320–325 (2012).
10. Gao, F. & Zhang, C. T. DoriC: a database of *oriC* regions in bacterial genomes. *Bioinformatics.* **23**, 1866–1867 (2007).
11. Gao, F., Luo, H. & Zhang, C. T. DoriC 5.0: an updated database of *oriC* regions in both bacterial and archaeal genomes. *Nucl. Acid. Res.* **41**, D90–D93 (2013).
12. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
13. Yin, G. *et al.* Fermentation process. US Patent 4,935,359 (1990).

14. Kleerebezem, M., Quadri, L. N., Kuipers, O. P. & De Vos, W. M. Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. *Mol. Microbiol.* **24**, 895–904 (1997).
15. Case, R. J., Labbate, M. & Kjelleberg, S. AHL-driven quorum-sensing circuits: their frequency and function among the *Proteobacteria*. *Isme J.* **2**, 345 (2008).
16. Subramoni, S. & Venturi, V. LuxR-family: bachelor sensors/regulators of signalling molecules. *Microbiol.* **155**, 1377–1385 (2009).
17. Mercer, R. G. *et al.* Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *Fems. Microbiol. Lett.* **331**, 53–62 (2012).
18. Lang, A. S. & Beatty, J. T. Importance of widespread gene transfer agent genes in α -proteobacteria. *Trends. Microbiol.* **15**, 54–62 (2007).
19. Jia, N., Du, J., Ding, M. Z., Gao, F. & Yuan, Y. J. Genome sequence of *Bacillus endophyticus* and analysis of its companion mechanism in the *Ketogulonigenium vulgare*-*Bacillus* strain consortium. *PLoS one.* **10**, e0135104 (2015).
20. Olijve, W. & Kok, J. Analysis of growth of *Gluconobacter oxydans* in glucose containing media. *Arch. Microbiol.* **121**, 283–290 (1979).
21. Zhou, J., Yi, H., Wang, L. L., Zhang, W. W. & Yuan, Y. J. Metabolic analysis of the positive effects on *Ketogulonigenium vulgare* growth and 2-keto-L-gulononic acid production by reduced glutathione. *Omic.* **16**, 387–396 (2012).
22. Chambers, S. T. & Kunin, C. M. Isolation of glycine betaine and proline betaine from human urine. Assessment of their role as osmoprotective agents for bacteria and the kidney. *J. Clin. Invest.* **79**, 731 (1987).
23. Gao, L. L. *et al.* Stepwise metabolic engineering of *Gluconobacter oxydans* WSH-003 for the direct production of 2-keto-L-gulononic acid from D-sorbitol. *Metab. Eng.* **24**, 30–37 (2014).
24. Huergo, L. F. *et al.* Regulation of *glnB* gene promoter expression in *Azospirillum brasilense* by the NtrC protein. *Fems. Microbiol. Lett.* **223**, 33–40 (2003).
25. Pawlowski, K., Klosse, U. & De Bruijn, F. Characterization of a novel *Azorhizobium caulinodans* ORS571 two-component regulatory system, NtrY/NtrX, involved in nitrogen fixation and metabolism. *Mol. Gen. Genet.* **231**, 124–138 (1991).
26. Li, Q., Diao, J. Y., Xiang, B. T. & Cao, Z. Studies on metabolism of nitrogen source in fermentation of nitrogen source in fermentation of 2-keto-L-gulononic acid. *Acta Microbiologica Sinica.* **1**, 002 (1996).
27. Forward, J. A., Behrendt, M. C., Wyborn, N. R., Cross, R. & Kelly, D. J. TRAP transporters: a new family of periplasmic solute transport systems encoded by the *dctPQM* genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. *J. Bacteriol.* **179**, 5482–5493 (1997).
28. Du, J., Bai, W., Song, H. & Yuan, Y. J. Combinational expression of sorbose/sorbosone dehydrogenases and cofactor pyrroloquinoline quinone increases 2-keto-L-gulononic acid production in *Ketogulonigenium vulgare*-*Bacillus cereus* consortium. *Metab. Eng.* **19**, 50–56 (2013).
29. Tsunenari, T. *et al.* Structure-function analysis of the bestrophin family of anion channels. *J. Biol. Chem.* **278**, 41114–41125 (2003).
30. Ma, Q. *et al.* Integrated proteomic and metabolomic analysis of an artificial microbial community for two-step production of vitamin C. *PLoS one.* **6**, e26108 (2011).
31. Ma, Q., Zou, Y., Lv, Y. J., Song, H. & Yuan, Y. J. Comparative proteomic analysis of experimental evolution of the *Bacillus cereus*-*Ketogulonigenium vulgare* co-culture. *PLoS one.* **9**, e91789 (2014).
32. Hou, S. *et al.* Genome sequence of the deep-sea γ -proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *PNAS.* **101**, 18036–18041 (2004).
33. Leduc, S., De Troostembergh, J. C. & Lebeault, J. M. Folate requirements of the 2-keto-L-gulononic acid-producing strain *Ketogulonigenium vulgare* LMP P-20356 in L-sorbosone/CSL medium. *Appl. Microbiol. Biotechnol.* **65**, 163–167 (2004).
34. Ding, M. Z., Zou, Y., Song, H. & Yuan, Y. J. Metabolomic analysis of cooperative adaptation between co-cultured *Bacillus cereus* and *Ketogulonigenium vulgare*. *PLoS one.* **9**, e94889 (2014).
35. Wilson, K. Preparation of genomic DNA from bacteria. *Current protocols in molecular biology*: 241–245 (1987).
36. Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genet.* **9**, 75 (2008).
37. Altschul, S. F. & Gish, W. Local alignment statistics. *Meth. Enzymology.* **266**, 460–480 (1996).
38. Kanehisa, M. *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucl. Acid. Res.* **34**, D354–D357 (2006).
39. Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC bioinformatics.* **4**, 41 (2003).
40. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl. Acid. Res.* **25**, 0955–0964 (1997).
41. Lagesen, K. *et al.* RNAmmer: consistent annotation of rRNA genes in genomic sequences. *Nucl. Acid. Res.* **35**, 3100–3108 (2007).
42. Nakai, K. & Horton, P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends. Biochem. Sci.* **24**, 34–35 (1999).
43. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods.* **8**, 785–786 (2011).
44. Gao, F. & Zhang, C. T. Ori-Finder: a web-based system for finding *oriCs* in unannotated bacterial genomes. *BMC bioinformatics.* **9**, 79 (2008).
45. Xu, Z. & Hao, B. L. CVTree update: a newly designed phylogenetic study platform using composition vectors and whole genomes. *Nucl. Acid. Res.* **37**, W174–W178 (2009).
46. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
47. Gao, F. & Zhang, C. T. GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. *Nucl. Acid. Res.* **34**, W686–W691 (2006).
48. Grant, J. R. & Stothard, P. The CGView Server: a comparative genomics tool for circular genomes. *Nucl. Acid. Res.* **36**, W181–W184 (2008).
49. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucl. Acid. Res.* **35**, W182–W185 (2007).
50. Ding, M. Z., Wang, X., Yang, Y. & Yuan, Y. J. Comparative metabolic profiling of parental and inhibitors-tolerant yeasts during lignocellulosic ethanol fermentation. *Metabolomics.* **8**, 232–243 (2012).
51. Ding, M. Z., Zhou, X. & Yuan, Y. J. Metabolome profiling reveals adaptive evolution of *Saccharomyces cerevisiae* during repeated vacuum fermentations. *Metabolomics.* **6**, 42–55 (2010).
52. Stein, S. E. & Scott, D. R. Optimization and testing of mass spectral library search algorithms for compound identification. *J. Amer. Chem. Soc. Mass Spectrom.* **5**, 859–866 (1994).
53. Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemometr. Intell. Lab. Syst.* **2**, 37–52 (1987).
54. Hubert, L. J. Hierarchical cluster analysis. *Encyclopedia of statistical sciences* (2006).

Acknowledgements

This work was funded by the Ministry of Science and Technology of China (“973” Program: 2014CB745102, 2015AA020101), and the National Natural Science Foundation of China (Major Program: 21390203, general program: 31171238 and 31571358). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

F.G., M.Z.D. and Y.J.Y. designed the project and experiments; N.J., J.D., C.H.P., J.D.L., J.H.F., G.T. and Y.J.Y. performed the experiments; F.G. and Y.J.Y. contributed reagents/materials/ analysis tools; N.J., M.Z.D. and F.G. analyzed the final data and wrote the manuscript. All the authors read and approved the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Jia, N. *et al.* Insights into mutualism mechanism and versatile metabolism of *Ketogulonicigenium vulgare* Hbe602 based on comparative genomics and metabolomics studies. *Sci. Rep.* **6**, 23068; doi: 10.1038/srep23068 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>