Synthetic and Systems Biotechnology 8 (2023) 688–696

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Original Research Article

Reconstruction and metabolic profiling of the genome-scale metabolic network model of *Pseudomonas stutzeri* A1501

Qianqian Yuan [a,c,1], Fan Wei [a,b,c,1], Xiaogui Deng [a,c,d], Aonan Li [a,c,d], Zhenkun Shi [a,c], Zhitao Mao [a,c], Feiran Li [e,**], Hongwu Ma [a,c,*]

[a] *Biodesign Center, Key Laboratory of Engineering Biology for Low-carbon Manufacturing, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China*
[b] *University of Chinese Academy of Sciences, Beijing, 100049, China*
[c] *National Technology Innovation Center of Synthetic Biology, Tianjin, 300308, China*
[d] *School of Biological Engineering, Tianjin University of Science and Technology, Tianjin, China*
[e] *Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China*

A B S T R A C T

Pseudomonas stutzeri A1501 is a non-fluorescent denitrifying bacteria that belongs to the gram-negative bacterial group. As a prominent strain in the fields of agriculture and bioengineering, there is still a lack of comprehensive understanding regarding its metabolic capabilities, specifically in terms of central metabolism and substrate utilization. Therefore, further exploration and extensive studies are required to gain a detailed insight into these aspects. This study reconstructed a genome-scale metabolic network model for *P. stutzeri* A1501 and conducted extensive curations, including correcting energy generation cycles, respiratory chains, and biomass composition. The final model, iQY1018, was successfully developed, covering more genes and reactions and having higher prediction accuracy compared with the previously published model iPB890. The substrate utilization ability of 71 carbon sources was investigated by BIOLOG experiment and was utilized to validate the model quality. The model prediction accuracy of substrate utilization for *P. stutzeri* A1501 reached 90 %. The model analysis revealed its new ability in central metabolism and predicted that the strain is a suitable chassis for the production of Acetyl CoA-derived products. This work provides an updated, high-quality model of *P. stutzeri* A1501for further research and will further enhance our understanding of the metabolic capabilities.

1. Introduction

γ-Proteus *Pseudomonas stutzeri* is a gram-negative, non-fluorescent denitrifying bacterium. It encompasses a diverse range of strains and exhibits a widespread distribution in various natural ecological environment, such as water bodies, soil, and the surface of animals or plants. *P. stutzeri* A1501 is a nitrogen-fixing bacterium that belongs to the genus Pseudomonas and was the first strain identified as an associative nitrogen-fixing bacteria within this genus [1]. *P. stutzeri* A1501 was originally isolated from the rhizosphere of paddies growing in the southern China in 1980s [2], and was initially classified as *Alcaligenes faecalis*. Later, it was reclassified to Pseudomonas genus through phylogenetic analysis of its 16S rRNA sequences [3]. The publication of the complete genome sequence of *P. stutzeri* A1501 in 2008 marked a significant milestone in the research on this strain, greatly advancing its study and opening up new possibilities for its development and application [4]. *P. stutzeri* A1501 is an important strain used to study the symbiotic nitrogen fixation with rice plants, and it has been recognized as one of the genetic engineering nitrogen fixation bacteria for plants as early as 2000 [5]. Currently, it is commonly utilized not only for inoculating rice, but also maize and other gramineous plants [6]. Hence, it is imperative to undertake further and comprehensive investigations on *Pseudomonas stutzeri* A1501, given its potential as a sustainable alternative to conventional chemical nitrogen fertilizers [7,8]. Such studies hold profound importance not only for agricultural production but also for soil amelioration, environmental conservation, and multifaceted applications.

In the meantime, genome-scale metabolic models (GEM) have been

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author. Biodesign Center, Key Laboratory of Engineering Biology for Low-carbon Manufacturing, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China.
** Corresponding author.
E-mail addresses: yuan_qq@tib.cas.cn (Q. Yuan), weif@tib.cas.cn (F. Wei), dengxg@tib.cas.cn (X. Deng), lian@tib.cas.cn (A. Li), zhenkun.shi@tib.cas.cn (Z. Shi), mao_zt@tib.cas.cn (Z. Mao), feiranli@sz.tsinghua.edu.cn (F. Li), ma_hw@tib.cas.cn (H. Ma).
[1] Equal contribution.

https://doi.org/10.1016/j.synbio.2023.10.001
Received 30 June 2023; Received in revised form 21 September 2023; Accepted 10 October 2023
Available online 20 October 2023

valuable tools for systematically understanding various organisms [9]. A genome-scale metabolic network model of *P. stutzeri* A1501, iPB890, was published in 2015 [10]. However, it was identified that this model contained certain limitations, as it lacks crucial metabolic information and may contain errors within the model structure. These shortcomings highlight the ongoing need for further refinement and improvement of metabolic models to ensure their accuracy and comprehensiveness for comprehensive analysis and simulations of *P. stutzeri* A1501's metabolic capabilities.

Here, we reconstruct the GEM for *P. stutzeri* A1501, iQY1018, which will provide a helpful framework for understanding and exploring the metabolic capacity of *Pseudomonas stutzeri* A1501. The reconstruction began with the rapid annotation system RAST server of A1501 genome sequence [11]. The RAST annotation output was then processed through the Model SEED, resulting in an initial reconstruction network [12]. To further refine and expand this network, additional information was extracted from MetaCyc [13] and KEGG database [14]. The integration of data from these sources contributed to the development of a draft network, which was then underwent further refinement and corrections in terms of removing energy generation cycles, curating respiratory chains, and biomass composition. The final model, iQY1018, was successfully developed. The model quality was validated by conducting BIOLOG substrate utilization experiments result performed in this study. To investigate the capabilities of *P. stutzeri* A1501, we conducted the pathway comparison analysis between iPB890 and iQY1018 that revealed its enhanced performance in simulating central metabolism. The analysis also suggested that the strain is well-suited as a chassis for the production of Acetyl CoA-derived products. This study not only deepens our understanding of the metabolic capabilities of *P. stutzeri* A1501 but also provides an updated and high-quality model for *P. stutzeri* A1501, serving as a valuable resource for future research endeavors.

2. Materials and methods

2.1. Draft model generation for Pseudomonas stutzeri A1501

The genome sequence of *Pseudomonas stutzeri* A1501 was downloaded from NCBI database (ftp://ftp.ncbi.nih.gov/genomes/). The RAST server (http://rast.nmpdr.org/) was used to annotate the genome of *P. stutzeri* A1501. Then, a draft reconstructed metabolic network model was automatically generated using the Model SEED platform (http://modelseed.org/).

2.2. Biomass composition

The detailed biomass composition of *P. stutzeri* A1501, including amino acids, fatty acids, lipids, DNA and RNA was generated to describe its cellular compositions (Supplement file1: Table A). The composition was obtained from the biomass composition of previous GEM of *P. stutzeri* A1501 iPB890 [10]. The proportion of biomass compositions was adjusted according to the literature [15]. The parameter of growth-associated maintenance (GAM) and non-growth associated maintenance (NGAM) was adopted from the iPB890.

2.3. Gap filling

Due to incomplete knowledge and annotation limitations, the reconstructed draft model contains network gaps, namely, the absence of certain reactions and genes required for synthesizing specific metabolites in the biomass. During the gap filling stage, these gaps were reconciled using the weight pFBA methods used in our previous work [16]. Then the missing information regarding reaction-gene associations of *P. stutzeri* A1501 was integrated from the KEGG, MetaCyc and literature (Supplement file1: Table C). The reaction-gene associations of *P. stutzeri* A1501 in KEGG were generated by KEGG API, while for MetaCyc, the Pathway Tools software provided by MetaCyc database (http://bioinformatics.ai.sri.com/ptools/) were utilized.

2.4. BIOLOG experiment

The ability of *P. stutzeri* A1501 to utilize various substrates was tested by performing the BIOLOG phenotypic assay using GENE III microplates on BIOLOG automatic microbial identification system GEN III OmniLog Plus. In this assay, *P. stutzeri* A1501 was tested for its ability to oxidize 71 carbon substrates in minimal medium.

2.5. FBA analysis

FBA is a widely used computational modeling technique used in systems biology to study and predict cellular metabolism for constraint-based metabolic models [17]. FBA analysis in this study was performed in Python using the COBRApy toolbox [18]. All optimizations were performed utilizing the CPLEX solver (IBM, Armonk, NY, USA).

3. Result

3.1. Metabolic network reconstruction

The genome-scale metabolic network model of *P. stutzeri* A1501 was reconstructed following the process depicted in Fig. 1. This process involved in four stages: (1) Reconstruction of the draft model using Model SEED; (2) Generation of the biomass reaction and filling of gaps to create a draft model capable of in silico growth with glucose as the sole carbon source; (3) Modification of the respiratory chain and energy generation processes, and verification of the synthetic pathways for biomass precursors to enhance the accuracy of the model; (4) Conducting BIOLOG substrate utilization experiments and utilizing the obtained data to expand and validate the in silico viability of the model on different substrates.

3.2. Model reconstruction

For the first stage, a draft model containing 949 genes, 1103 metabolites and 1139 reactions was generated using the RAST and Model SEED platform [11,19]. The biomass composition of the iPB890 model of *P. stutzeri* A1501 was adopted with some adjustments. For accurate modeling and satisfying the mass balance constraint, it is necessary to ensure that the sum of biomass components in the GEM equals 1 g [20]. However, the sum of biomass components of iPB890 is 1.19 g DCW (Supplement file1: Table A), which should be refactored. In the new biomass composition of our model, the proportion of lipid composition (phosphatidylethanolamine 65 %, cardiolipin 14 %, phosphatidylglycerol 14 %, others 8 %) was derived from literature [15] and was used to adjust the lipid composition of iPB890 biomass. The amino acid composition of protein is well-conserved among various species [21]. By analyzing proteome composition data from a diverse set of prokaryotes [22], it was observed that the content of L-Alanine was consistently lower than that of L-Leucine. However, in the case of iPB890, the composition of L-Alanine was found to be significantly higher than that of other amino acids. Given the absence of published proteomic data for *P. stutzeri* A1501, we utilized transcriptome data sourced from NCBI (SRR21626847, SRR21626848, SRR21626849, SRR21626850, SRR23519121) to decipher the quantitative relationships within the protein composition of *P. stutzeri* A1501, as protein abundance and mRNA levels exhibit a positive correlation according to the latest research article [23]. We observed that the content of L-alanine in *P. stutzeri* A1501 is lower than that of L-leucine, consistent with previous studies (Supplement file2: figure S1). Furthermore, the compositional ratio of L-alanine to L-leucine (0.127g/0.165g = 0.78) in *P. stutzeri* A1501 is consistent with the compositional ratio of the two amino acids in the biomass of the *E. coli* model iML1515, which is 0.78 (0.51

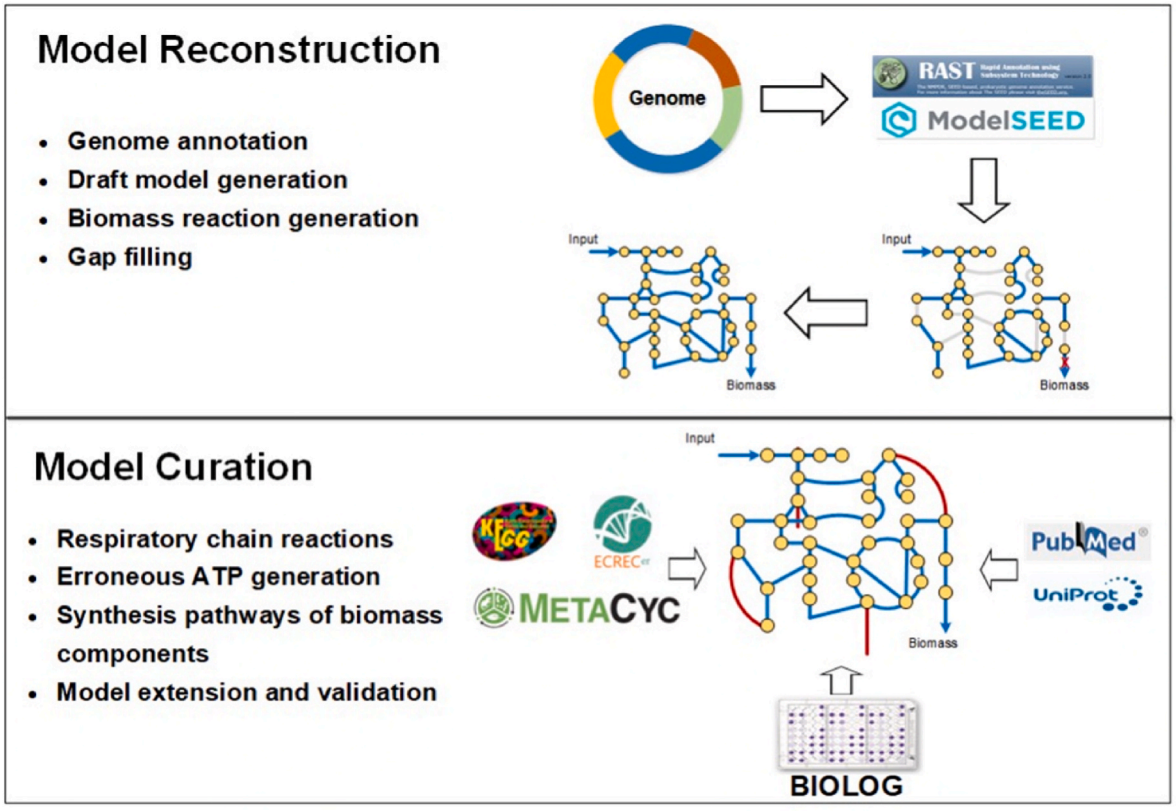


Fig. 1. Workflow of metabolic network reconstruction for *P. stutzeri* A1501. The model development was divided into model generation and curation.

mmol/gDCW*89.094 mg/mol)/(0.45 mmol/gDCW*131.175 mg/mol). However, according to the composition of iPB890, the mass of L-Alanine (1.39mmol/gDCW * 89.094 mg/mmol = 123.84mg/gDCW) is significantly higher than the mass of L-Leucine (0.428mmol/gDCW *131.175 mg/mmol = 56.14 mg/g DCW). This suggests that the accuracy of amino acid content determination in the iPB890 model through HPLC analysis may be compromised. Hence, we adjusted the coefficient of L-alanine from 1.37 to 0.499 mmol/gDCW, in accordance with the ratio relationship between L-Alanine and L-Leucine derived from transcriptome analysis. The compositions of soluble small molecule such as putrescine, heme and spermidine were directly adopted from iPB890. The final biomass composition sum was 1 g DCW (Supplement file1: Table B).

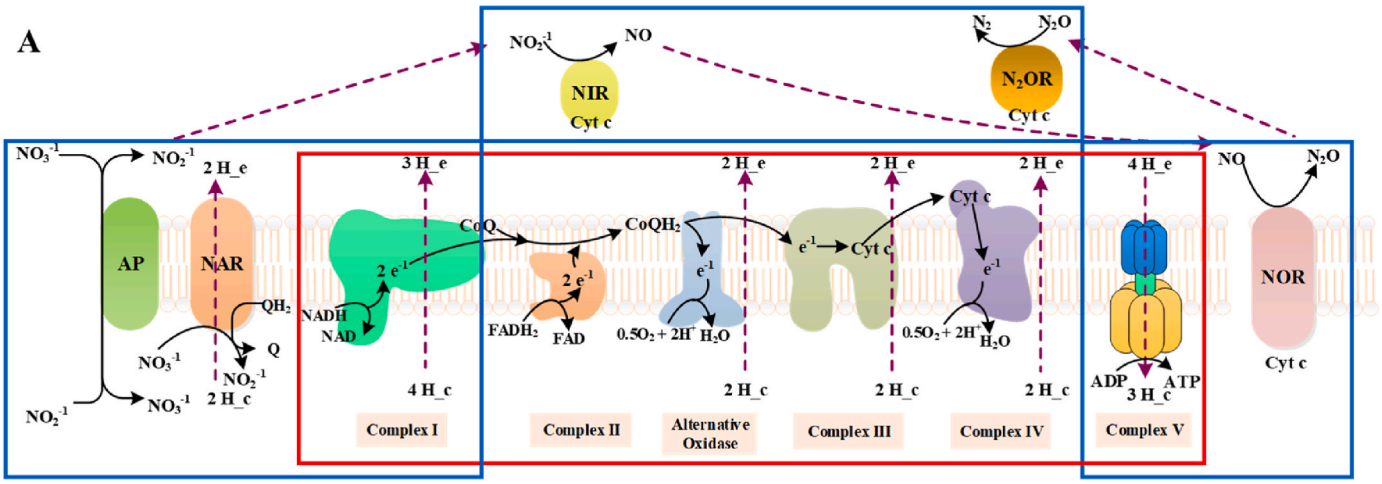
Additionally , we conducted an analysis to assess the impact of alterations in the ratios of biomass components on the growth yield of *P. stutzeri* A1501. To achieve this, we systematically modified the coefficients of individual compounds within the biomass equation, both increasing and decreasing them by 20 % according to previous method [24]. Subsequently, Flux Balance Analysis (FBA) was carried out with the modified biomass equation as the objective for each variation. The resultant values were then compared with the growth rate obtained using the unaltered biomass equation. The outcomes of this investigation revealed that the variation of any single biomass constituent by up to 20 % in either direction had a minimal impact of less than 1.5 % on the growth yield of *P. stutzeri* A1501 (Supplement file2: figure S2). These results are consistent with results of a previous study on the sensitivity of growth yield to biomass composition [25]. This observation underscores the robustness of the overall model. In the gap filling stage, 32 missing reactions were added to enable the synthesis of biomass components if the FBA results indicate that those components cannot be synthesized from glucose. The missing reactions and associated genes were added according to MetaCyc and KEGG. In addition, five reactions specific to *P. stutzeri* A1501 strain were manually extracted from the published literature and experimental data and added into the model (Supplement file1: Table C).

3.3. Model curation

3.3.1. Determination of respiratory chain reactions

P. stutzeri A1501 contains complete aerobic and anaerobic respiratory chains. However, the reactions within both the aerobic and anaerobic respiratory chains were found to be partially absent in the draft model. Specifically, in the anaerobic respiratory chain, the reaction rxn22447 (H_2O + 2.0 Cytochrome *c*3+ + N_2 <=> 2.0 H+ + 2.0 Cytochrome *c*2+ + Nitrous oxide) was missing. Although this reaction is present in the Model SEED database, it lacked annotation by RAST. To address this, we incorporated both the reaction and its corresponding genes based on information from literature and KEGG (https://www.genome.jp/pathway/psa00910+PST_3550). Similarly, in the aerobic respiratory chain, the absent reaction rxn06106 (2.0 Cytochrome *c*3+ + Ubiquinol-8→ 2.0 H+ + 2.0 Cytochrome *c*2+ + Ubiquinone-8) along with its corresponding genes, has also been integrated into the model (https://www.genome.jp/pathway/psa00190). These absent reactions were listed as added reactions in Supplement file 1 Table C.

The aerobic respiratory chain is presumed to be similar to that of *P. aeruginosa* [26], which transfers electrons from NADH/$FADH_2$ to oxygen through a series of respiratory chain complexes (Fig. 2). In this process, intracellular proton is pumped out of the cell and synthesized proton gradient which in turn drives the synthesis of ATP (blue box shown in Fig. 2A). The first step of proton transport in the respiratory chain of iPB890 does not pump out any proton, which resulted in the low P/O ratio of 1 in iPB890. When modifying the first step of the proton transport reaction in the respiratory chain of iPB890 to actively pump out protons and achieve a P/O ratio of 7/4 (according to published metabolic network model of *P. aeruginosa* and *E. coli*), the resulting growth rate was calculated at 0.67 h^{-1}. Therefore, the low P/O ratio may be one of the reasons for the lower calculated growth rate (0.546



B

rxn05638	-796.9	H_e + L-Proline_e <=> H_c + L-Proline_c
rxn10945	796.9	L-Proline_e <=> L-Proline_c
rxn05313	-333.3	(3.0)Na_e + Pi_e <=> (3.0)Na_c + Pi_c
rxn05312	-333.3	H_c + Pi_c <=> H_e + Pi_e
rxn05209	1000.0	H_c + Na_e <=> H_e + Na_c
rxn05209	753.4	H_c + Na_e => H_e + Na_c
rxn05221	-753.4	Na_e + L-Proline_e <=> Na_c + L-Proline_c
rxn10945	753.4	L-Proline_e <=> L-Proline_c

Net reaction:
H_c => H_e

Fig. 2. Respiratory chain of *P. stutzeri* A1501 (A) and the net proton gradient generation reactions in the draft model (B). The red box represents the aerobic oxidative respiratory chain, while the blue box represents the respiratory chain where nitrate serves as the electron acceptor under anaerobic conditions.

h^{-1}) compared to the experimental value (0.569 h^{-1}). However, it is important to note that there are numerous additional constraints and regulatory factors that affect the growth rate of an organism in reality, which may not be fully captured in idealized model simulations. These factors can include limitations in nutrient availability, metabolic regulation, enzyme kinetics, and other biological processes. Therefore, it is not uncommon for the calculated growth rate to be lower than the experimental value due to the simplifications and assumptions made in the model.

The corrected P/O in the model was set to 7/4 (according to published metabolic network model of *P. aeruginosa* [27] and *E. coli* [28]). Then the optimal ATP production rate was calculated by setting the ATP maintenance reaction (rxn00062 ATP + H_2O => H^+ + Pi + ADP) as the objective function [20]. However, the calculated result of ATP production rate was 1000 mmol $gDCW^{-1}$ h^{-1}, which is the upper boundary set for reactions, even when the glucose uptake rate was set to 0. After analyzing, we found that proton gradient can be generated from reactions shown in Fig. 2B, without utilizing respiratory chain. After identifying and correcting reaction direction of rxn05638, rxn05313 and rxn05221 to be irreversible, the optimal yield of ATP production was determined to be 24.125 mmol per mol of glucose. Under anaerobic conditions, denitrification occurs where nitrate serves as the final electron acceptor, producing nitrogen gas [29]. This process also generates a proton gradient used for ATP synthesis, known as chemiosmotic ATP synthesis (red box shown in Fig. 2A).

When the P/O ratio was set 7/4, the predicted growth rates for aerobic and anaerobic respiratory by iQY1018 were 0.81 h^{-1} and 0.54 h^{-1} with glucose uptake rate of 10 mmol $gDCW^{-1}h^{-1}$, respectively. Similarly, upon adjusting the P/O ratio to 7/4, the predicted growth rates for aerobic and anaerobic respiratory by iPB890 were 0.67 h^{-1} and 0.398 h^{-1} with glucose uptake rate of 10 mmol $gDCW^{-1}h^{-1}$. In this context, the notion of iPB890 exhibiting aerobic growth greater than the experimental data (0.569 h^{-1}), simultaneously displaying anaerobic growth lower than the experimental data (0.453 h^{-1}). Given that simulation predictions reflect theoretical maximum values, it's reasonable for simulated data to surpass experimental data. Consequently, when comparing the predicted growth rates between iPB890 and iQY1018, the growth rate predictions by iQY1018 appear more reasonable.

3.3.2. *Manual correction of metabolic pathways of biomass precursors*

The synthetic pathways of biomass components were checked manually. FBA simulations were performed with glucose as the substrate and the precursor as the target product. The simulated results were manually checked and corrected according to pathways provided by MetaCyc database. There were two main types of mistakes: (1) errors in reaction direction, e.g. in the draft model, threonine was synthesized by reverse reaction rxn00541 (L-Threonine <=> Acetaldehyde + Glycine), which is catalyzed by threonine aldolases (Fig. 3). Threonine aldolases have the ability to catalyze the decomposition of threonine into glycine and acetaldehyde as part of secondary glycine biosynthetic pathway [30]. However, they are unable to carry out the reverse reaction for threonine synthesis. (2) redundant reactions existed in draft model, e.g. the normal sulfate metabolism process in *P. stutzeri* A1501 is sulfate => adenylyl sulfate => adenosine-5-phosphosulfate => sulfite (https://www.genome.jp/pathway/pas00920) [31]. There is a lumped one-step reaction (rxn05256) from adenylyl sulfate to sulfite with one ATP saving in the draft model. However, this reaction lacked annotation information for *P. stutzeri* A1501 and was subsequently removed from the model. Similar errors also exist in the iPB890 model, as shown in Fig. 3, where cysteine is directly synthesized from pyruvate. After correcting the errors in the initial model, all the biomass synthesis pathways now follow the correct biological rules, as indicated by the green arrows in Fig. 3.

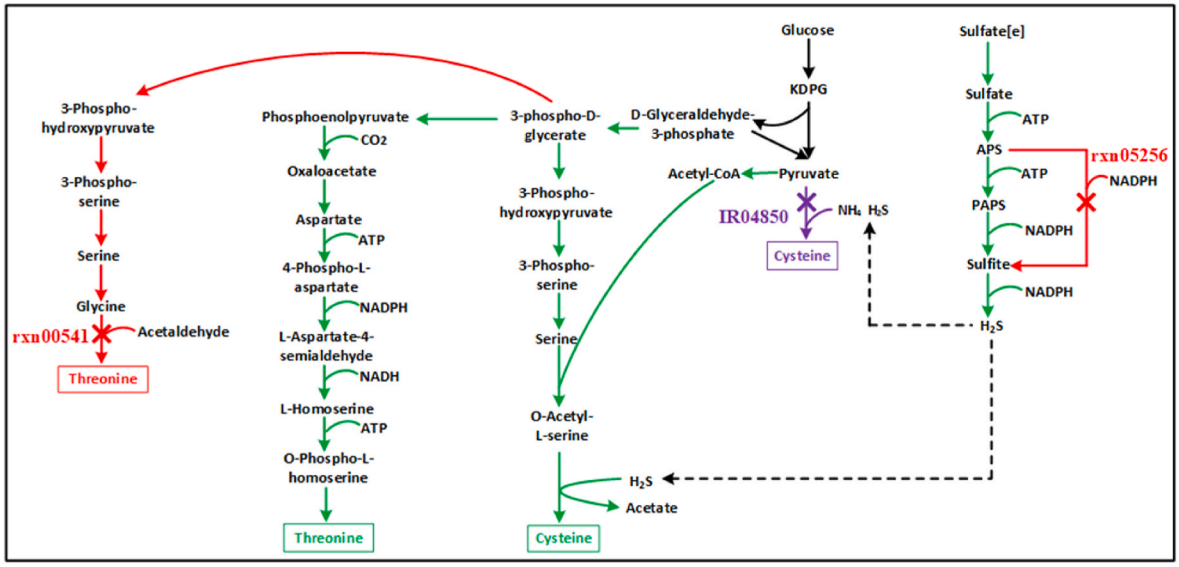


Fig. 3. Correcting the erroneous pathways in the initial model, the red lines represent the erroneous pathways in the initial model, the green lines represent the correct pathways in the final model, and the purple lines represent the erroneous pathways in the iPB890 model.

3.3.3. *Phenotyping data analysis*

The model was validated and curated using carbon source utilization data, which was produced using the Microbial Identification Test Panel (GEN III Microplate). Among all 71 carbon sources in BIOLOG experiment, 41 of them was shown that can be utilized by *P. stutzeri* A1501. To assess the accuracy of the model's predictions for specific carbon sources, we conducted a comparison between the experimental results and the model's predictions. This was achieved by performing Flux Balance Analysis (FBA) using a minimal medium supplemented with the corresponding carbon source (Materials and Methods). Disagreements between the in silico simulations and the BIOLOG experimental results were used as the hypotheses to refine the model reconstruction. It was found that out of the 41 substrates experimentally utilized by *P. stutzeri* A1501, only 9 of them were capable of supporting in silico growth as a sole carbon source in initial model. 8 of them were present as intracellular metabolites but there were no transporters assigned in the *P. stutzeri* A1501. To support in silico growth, transport reactions for each carbon source were implemented. 17 of them were investigated, and missing reactions were added to fill the gaps according to literature (Supplement file1: Table D). Finally, according to the predicted results, 64 out of 71 substrates were consistent with the experimental data, and the prediction accuracy of substrates utilization of *P. stutzeri* A1501 reached 90 %, which is higher than the 55 % accuracy of iPB890 model (Fig. 4). However, seven carbon sources could be used as the sole carbon source for cell growth in BIOLOG experiments, but not in FBA prediction in *P. stutzeri* A1501. Among the seven carbon sources (Dextrin, gentiobiose, 3-methyl-D-glucose, Glucuronamide, Methyl pyruvate, Bromosuccinic acid, Tween 40), dextrin (formula: $(C_{12}H_{20}O_{10})n$) is a polymer, and its exact composition (as indicated by 'n') remains unknown. The metabolic pathways for the catabolic utilization of the other six carbon sources have not been elucidated at present by querying literature and biochemical metabolism databases such as KEGG and MetaCyc. As a

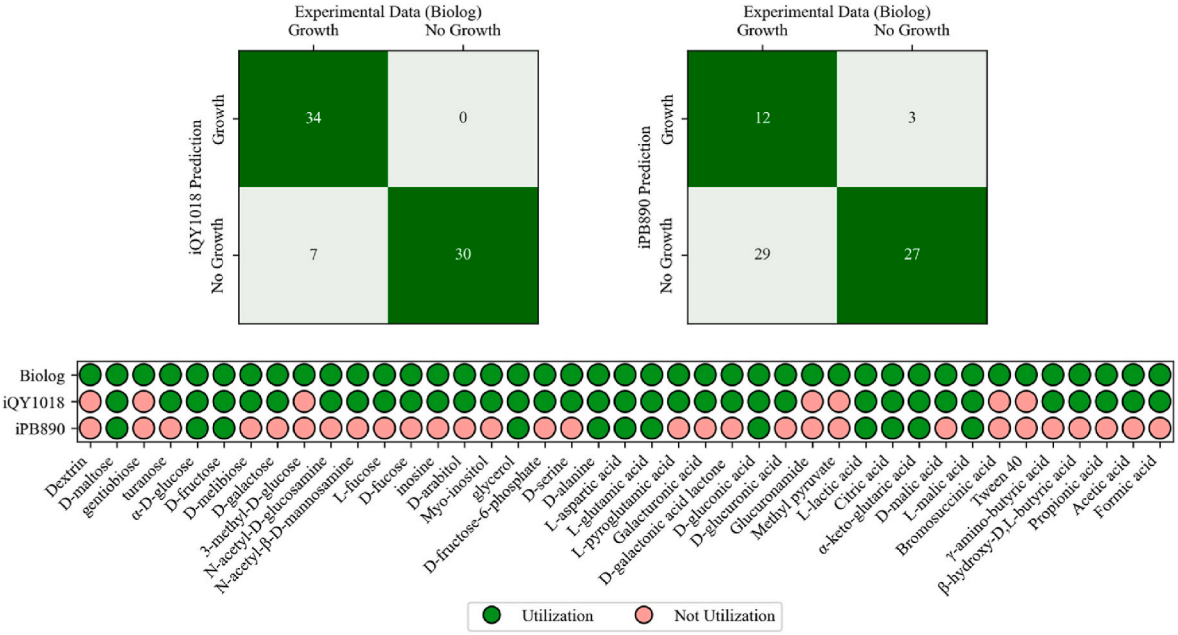


Fig. 4. Comparison of substrate utilization prediction accuracy between iQY1018 and the iPB890 model.

consequence of the current limitations in our comprehension of the associated metabolism, the specific catabolic reactions corresponding to these carbon sources have not yet been ascertained.

In the iPB890 study, the model was validated using carbon source utilization data generated by BIOLOG MicroPlate™, which included 95 carbon sources. However, only 25 of these carbon sources were selected for simulation due to the presence of corresponding transport reactions in iPB890. Therefore, to facilitate comparison, we also assessed the carbon source utilization of iQY1018 using the same set of 25 carbon sources. Among these, 88 % (22 out of 25) of the modeling results matched the experimental data in iQY1018, while 80 % (20 out of 25) aligned with the experimental data in iPB890 (refer to Supplement file 1 Table H and Supplement file 2 Figure S3). These results from the BIOLOG assays highlight that iQY1018 exhibits a higher degree of accuracy in simulating substrate utilization when compared to the iPB890 model.

3.3.4. Enzyme annotation of genes using ECRECer

Previous studies have demonstrated that enzymes exhibit promiscuity in their substrate recognition, even within extensively studied model organisms such as *Escherichia coli* [32]. To compensate the missing and omitted gene for SEED annotations in the initial model, the genes of the organism were annotated using the state-of-the-art machine learning-based method ECRECer (https://ecrecer.biodesign.ac.cn). ECRECer is designed based on a gated recurrent unit framework and is capable of predicting EC number in a multi-objective hierarchy and multitasking manner [33]. ECRECer supplemented missing genes for 55 reactions and expanded their GPR (gene-protein-reaction) associations (Supplement file1: Table E). For example, gene PST_3564 is annotated as an aromatic amino acid aminotransferase (EC number 2.6.1.57), exhibiting aspartate and 4-hydroxy-L-glutamate aminotransferase activity in the initial model. ECRECer have uncovered that the enzyme also has arogenate aminotransferase activity, thereby the GPR of the reaction rxn01270 is expanded.

3.4. Comparison to existing reconstruction of P. stutzeri A1501

Following the above process, a genome-scale metabolic model iQY1018 for *P. stutzeri* A1501 was reconstructed (Table 1). These models in various formats (XML, JSON, XLSX, MAT) are available for access (https://github.com/tibbdc/psa-GEM/tree/develop/model). iQY1018 consists of 1018 genes, 1420 reactions, and 1203 metabolites. As depicted in Table 1, iQY1018 exhibits a higher genome coverage compared to iPB890, indicating that a larger portion of the genome is accounted for in iQY1018. The most prominent metabolic pathways in iQY1018 are associated with metabolism of amino acids and carbohydrates (Fig. 5A). Compared to iPB890, a total of 128 additional genes were incorporated and distributed across 11 pathways (Fig. 5A and Supplement file1: Table F). For example, in the metabolism of other amino acids, the newly added gene PST_2761 catalyzes the reaction ($2.0H_2O$ + ATP + 5-oxo-proline– > ADP + Phosphate + L-Glutamate), which is crucial for the utilization of the substrate L-pyroglutamic acid. Therefore, the incorporation of these genes into the model expands its metabolic capabilities, resulting in more accurate predictions.

Table 1
General features of iQY1018 and compared with iPB890.

Genome features			Note
Genome size (bp)	4,085,337		
Total open reading frames (ORFs)	3995		
Model features	iQY1018	iPB890	
Total reactions	1420	1034	
Metabolites	1203	813	
ORFs associated in model	1018	890	Common: 782
ORF coverage	25.5 %	22.2 %	Unique of iQY1018: 236 Unique of iPB890: 108

MEMOTE was utilized to evaluate the quality of iQY1018 and iPB890 (iQY1018_memote.html, iPB890_memote.html). Apart from the notable enhancements in consistency and the annotation of reactions and metabolites, iQY1018 exhibits explicit annotations for genes and SBO-terms (Systems Biology Ontology) (Fig. 5B). This grants iQY1018 a distinct advantage in terms of the overall score. All the validation outcomes corroborated the high reliability of iQY1018.

We also conducted a comparison between iQY1018 and two available models Pseudomonas_ stutzeri_CGMCC_1_1803 and Pseudomonas_stutzeri_28a24 in CarveMe under identical growth conditions (Table 2). (https://github.com/cdanielmachado/embl_gems/blob/master/models/p/pseudomonas/Pseudomonas_stutzeri_CGMCC_1_1803.xml.gz; https://github.com/cdanielmachado/embl_gems/blob/master/models/p/pseudomonas/Pseudomonas_stutzeri_28a24.xml.gz)

When glucose was employed as the substrate, the growth rate predicted by the model Pseudomonas_ stutzeri_CGMCC_1_1803 was determined to be zero. To assess the accuracy of substrate utilization in different models, we utilized 71 carbon sources from the BIOLOG experiment. In this regard, the accuracy of substrate utilization for iQY1018 and Pseudomonas_stutzeri_28a24 reached 90 % and 56 %, respectively. However, for Pseudomonas_stutzeri_CGMCC_1_1803, the growth rate remained zero across all 71 carbon sources, suggesting a potential gap in biomass precursor synthesis within this model. Furthermore, both models in CarveMe lacked corresponding reactions in the non-oxidative glycolysis (NOG) pathway. This could potentially result in the absence of carbon-conserving pathways for acetyl-CoA derivatives such as acetate and L-glutamate. In addition, the two models in CarveMe exhibit incomplete aerobic and anaerobic respiratory chains. Specifically, they both lack the anaerobic respiratory chain where nitrate serves as the final electron acceptor, leading to the production of nitrogen gas. The quality evaluation of the three models was performed using MEMOTE, revealing that iQY1018 achieved higher scores than the two models in CarveMe. The comparative analysis of iQY1018 and the two models in CarveMe strongly indicates that iQY1018 stands as a high-quality model of *P. stutzeri* A1501. This model holds promise for further research and is expected to significantly advance our comprehension of the metabolic capabilities of the organism.

3.5. Analysis of metabolic capacity of P. stutzeri A1501 based on iQY1018

In order to reveal the metabolic characteristics of *P. stutzeri* A1501, we conducted an analysis and comparison of the metabolic pathways related to biomass composition in iQY1018 and iPB890. The pathway-consensus approach proposed in our earlier studies was used to compared these two metabolic model reconstructions [20]. Firstly, we lumped a series of respiratory chain reactions of iQY1018 and iPB890 into one overall reaction (Eq (1) and Eq (2)). The respiratory chain of iQY1018 is complete, thus iQY1018 (P/O = 7/4) has the higher P/O ratios than iPB890 (P/O = 1). We replaced the respiratory chain reactions in iPB890 by the one in iQY1018 to perform fair comparisons. Then we simulated the optimal production rate of every precursor of iQY1018 and iPB890 with fixing glucose input of 10 mmol $gDCW^{-1}$ h^{-1} (Supplement file1: Table G).

$$4NADH+11H+2O_2+7ADP+7PI \rightarrow 7ATP+4NAD+11H_2O \quad (1)$$

$$2NADPH+2H + O_2+2ADP+2PI \rightarrow 2ATP+2NADP+2H_2O \quad (2)$$

Acetate, acetyl phosphate, acetyl CoA (AcCoA) and L-glutamate were found to have higher optimal formation rate in iQY1018 compared with iPB890. Through comparative analysis, it was discovered due to the existence of that phosphoketolase (Fxpk) and its corresponding reaction in iQY1018, encoded by gene PST_1512 through RAST annotation. Fxpk is capable of catalyzing the conversion of Fructose-6-phosphate (F6P) or

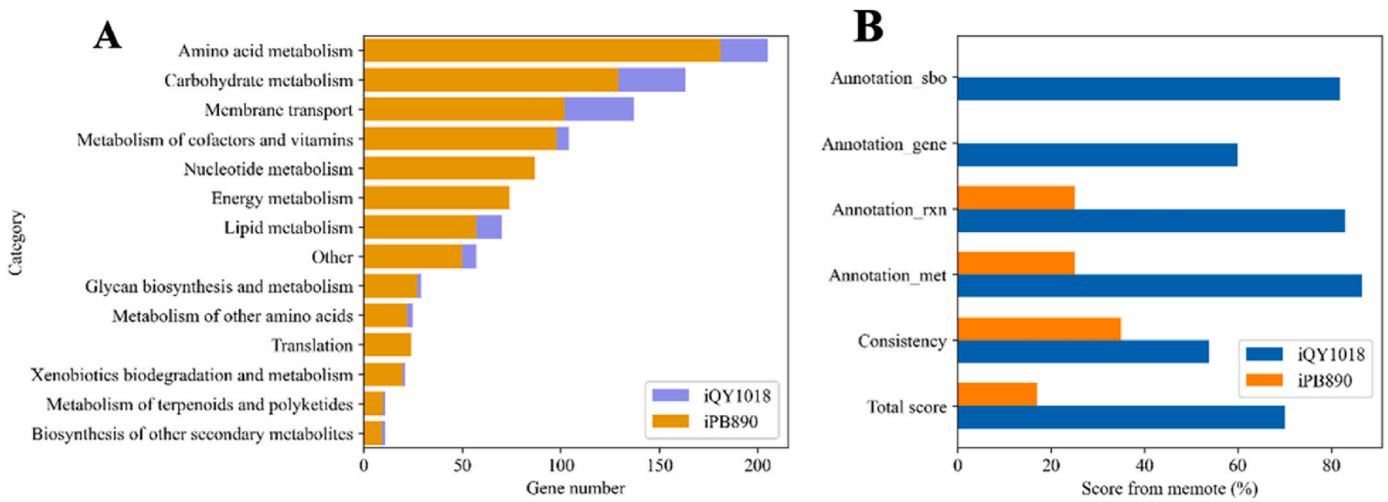


Fig. 5. Comparative analysis of genes in metabolic pathways. (A) and assessment of model quality using MEMOTE (B) for in iQY1018 and iPB890.

Table 2
The comparison between iQY1018 and two available models in CarveMe.

	iQY1018	Pseudomonas_stutzeri_28a24	Pseudomonas_stutzeri_CGMCC_1_1803
Grwth rate(h^{-1}) (10 mmol Glucose)	0.81	0.85	0
Accuracy of substrate utilization	90 %	56 %	/
Non-oxidative glycolysis (NOG)	✓	×	×
Integrity of the respiratory chains	complete	absent	absent
MEMOTE score	70	20	20

xylulose-5-phosphate (Xu5P) into acetyl phosphate (AcP) and erythrose-4-phosphate (E4P) or glyceraldehyde-3-phosphate (G3P) (Eq (3) and Eq (4)). The pathway was named non-oxidative glycolysis (NOG) for carbon-conserving (Fig. 6) [34].

rxn01187 H+ + Pi + Xu5p =>G3p + Acp + H_2O (3)

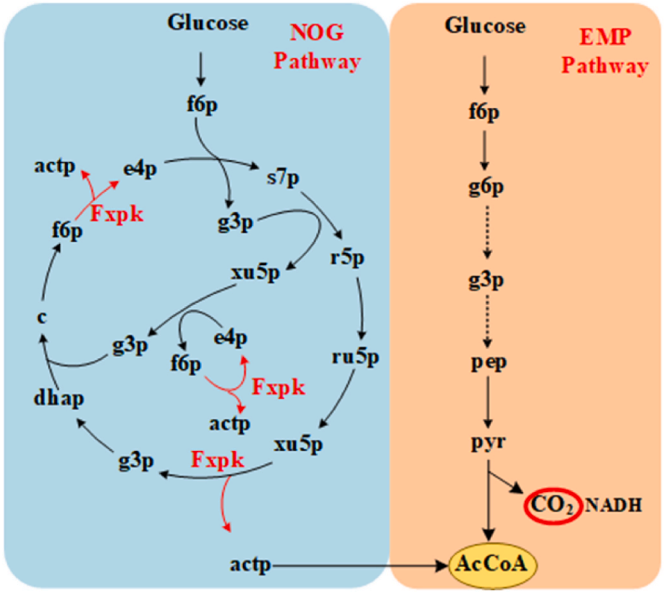


Fig. 6. The non-oxidative glycolysis (NOG) pathway and Embden-Meyerhof-Parnas pathway (EMP) pathway in iQY1018. f6p, D-fructose 6-phosphate; fdp, fructose 1,6-bisphosphate; dhap, Dihydroxyacetone phosphate; g3p, D-glyceraldehyde 3-phosphate; xu5p, D-xylulose 5-phosphate; r5p, D-ribose 5-phosphate; ru5p, D-ribulose 5-phosphate; e4p, D-erythrose 4-phosphate; s7p, D-sedoheptulose 7-phosphate; pep, phosphoenolpyruvate; pyr, pyruvate; and AcCoA, acetyl-coenzyme A.

rxn00548 H+ + F6P + Pi => E4P + Acp + H_2O (4)

In NOG pathway, E4P and G3P can enter the central metabolic pathway, while AcP can produce AcCoA through a single-step reaction (Fig. 6). This enzymatic process using Fxpk enables the ultimate conversion of 1 mol of glucose into 3 mol of AcCoA, achieving 100 % carbon conversion. The inclusion of Fxpk in the metabolic pathway offers advantages over traditional fermentation processes as it helps avoid the loss of carbon [35]. By bypassing the production of pyruvate and directly producing AcCoA, Fxpk prevents the loss of 1/3 of the carbon content. This can increase the yield of AcCoA and its derivatives, including acetate, L-glutamate, and fatty acids. However, Fxpk and its coding gene PST_1512 and the corresponding reaction are missing in iPB890.

The expression of PST_1512 gene and its enzymatic activity was verified. Through Blast comparison, it was found that the gene PST_1512 shares the highest homology with gene BADO_0732 from *Bifidobacterium adolescentis* 22L (*B. adolescentis* 22L). BADO_0732 also encodes Fxpk enzyme, suggesting a potential conservation of the Fxpk enzyme and its associated metabolic function across different bacterial species. According to the gene sequence provided by NCBI, the corresponding primers were designed, and the gene sequences were obtained by PCR using the genomes of the two strains as templates. Finally, two Fxpk genes were cloned into the expression vector of pET28a and transformed into *E. coli* to induce expression. The target protein was purified by nickel column. The purity and size of the protein were verified by polyacrylamide gel electrophoresis (Supplement file2: figure S3). Fxpk has the activity of catalyzing two substrates. The specific activities of fructose 6-phosphate and xylose-phosphate catalyzed by Fxpk were detected (Supplement file2: Table S1).

When comparing the pathways of iQY1501 and iPB890, we found that there were errors in the metabolic pathway of iPB890. For example, pyruvate could produce from cysteine directly using reaction IR04850 (Pyruvate + NH_4^+ + Hydrogen sulfide <=> H_2O + Cysteine), whereas

cysteine is synthesized from serine through acetylation and sulfide incorporation in bacteria and could not be directly produced by pyruvate [36] (Fig. 3). Another example is glycine synthesis, which is produced by the reverse reaction of RR00772 (Glycine + Tetrahydrofolate + NAD^+ <=> CO_2 + NH_4^+ + 5,10-Methylenetetrahydrofolate + NADH). However, reaction RR00772 should be irreversible and it can only be carried out on the direction of decarbonization. These reversibility errors of the reaction in iPB890 resulted in completely incorrect prediction results. From the comparison of the computational results, we can see that the synthesis rates of Cysteine and Glycine in iPB890 are both erroneously higher than those in iQY1018 (Supplement file1: Table G).

In addition to errors of reaction reversibility, there are a lot of errors in enzymatic annotation information in the iPB890. For example, IR00178 (Oxaloacetate + ATP => ADP + CO2 + Phosphoenolpyruvate) was catalyzed by the enzyme 1.4.3.5 (pyridoxal 5′-phosphate synthase) in iPB890. However, IR00178 cannot be catalyzed by 1.4.3.5 but 4.1.1.49 (phosphoenolpyruvate carboxykinase). Enzymes belong to EC1.4.3.5 group can catalyze reactions such as R00277 (pyridoxamine 5′-phosphate + H_2 + O_2 => pyridoxal 5′-phosphate + NH_3 + H_2O_2) and R00278 (pyridoxine 5′-phosphate + O_2 => pyridoxal 5′-phosphate + H_2).

4. Conclusion

In conclusion, we have reconstructed a genome-scale metabolic model, iQY1018, for *P. stutzeri* A1501 through an iterative and systematic approach. This model incorporates a larger portion of the genome compared to the previous model, iPB890, resulting in improved genome coverage and metabolic capabilities. The manual correction of metabolic pathways, annotation of genes using ECRECer, and validation using Microbial Identification Test Panel (GEN III Microplate) have contributed to the refinement and accuracy of the model. The comparative analysis of iQY1018 and iPB890 revealed metabolic differences, such as the presence of the non-oxidative glycolysis pathway (NOG) in iQY1018, which enhances the production rates of acetate, acetyl phosphate, acetyl CoA, and L-glutamate. Overall, iQY1018 demonstrates higher prediction accuracy for substrate utilization and exhibits explicit annotations for genes and SBO-terms, ensuring its reliability. This comprehensive metabolic model provides valuable insights into the metabolic characteristics of *P. stutzeri* A1501 and can serve as a valuable tool for further investigation and optimization of its metabolic capabilities.

Data availability

The datasets used in this study can be found in the Supplementary Material online or requested to the corresponding author. The model, data, and code have been published on GitHub (https://github.com/tibbdc/psa-GEM).

Author contributions

HM and FL conceived and designed the experiments with inputs from all coauthors. QY and FW built the model iQY1018 with help from XD, AL and ZS. QY and FW completed GEN III Microplate experiments. QY and FW finished the model simulation. FL, QY, FW and ZM prepared the original draft. All authors contributed to data analysis, read, revised and approved the final manuscript.

Funding

This research was funded by the National Key Research and Development Program of China (2018YFA0901400), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB0480000), Tianjin Synthetic Biotechnology Innovation Capacity Improvement Projects (TSBICIP-PTJS-001), Ministry of Science of China and Youth Innovation Promotion Association CAS (292023000018).

Declaration of competing interest

The authors declare no competing financial interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2023.10.001.

References

[1] Yan Y, et al. Global transcriptional analysis of nitrogen fixation and ammonium repression in root-associated Pseudomonas stutzeri A1501. BMC Genom 2010;11:1–13.
[2] Vermeiren H, et al. The rice inoculant strain Alcaligenes faecalis A15 is a nitrogen-fixing Pseudomonas stutzeri. Syst Appl Microbiol 1999;22(2):215–24.
[3] Lalucat J, et al. Biology of Pseudomonas stutzeri. Microbiol Mol Biol Rev 2006;70(2):510–47.
[4] Yan Y, et al. Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated Pseudomonas stutzeri A1501. Proc Natl Acad Sci USA 2008;105(21):7564–9.
[5] Venieraki A, et al. Characterization of nitrogen-fixing bacteria isolated from field-grown barley, oat, and wheat. J Microbiol 2011;49:525–34.
[6] Gilvanova E, et al. Potential of the bacterium Pseudomonas Stutzeri as a plant growth stimulator and a destructor of technogenic pollutants. In: AIP conference proceedings. AIP Publishing LLC; 2022.
[7] Sanow S, et al. Molecular mechanisms of Pseudomonas assisted plant nitrogen uptake-opportunities for modern agriculture. Mol Plant Microbe Interact 2023;(ja).
[8] Mehmood N, et al. Multifaceted impacts of plant-beneficial Pseudomonas spp. in managing various plant diseases and crop yield improvement. ACS Omega; 2023.
[9] Mao Z, et al. CAVE: a cloud-based platform for analysis and visualization of metabolic pathways. Nucleic Acids Res 2023.
[10] Babaei P, Marashi S-A, Asad S. Genome-scale reconstruction of the metabolic network in Pseudomonas stutzeri A1501. Mol Biosyst 2015;11(11):3022–32.
[11] Aziz RK, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genom 2008;9(1):1–15.
[12] Devoid S, et al. Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. Syst Metabol Eng: Method Protocol 2013:17–45.
[13] Caspi R, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 2010;38(suppl_1):D473–9.
[14] Kanehisa M, et al. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 2014;42(D1):D199–205.
[15] KaAWAI Y, et al. Ornithine-containing lipids of some Pseudomonas species. Eur J Biochem 1988;175(3):633–41.
[16] Luo J, et al. Reconstruction of a genome-scale metabolic network for shewanella oneidensis MR-1 and analysis of its metabolic potential for bioelectrochemical systems. Front Bioeng Biotechnol 2022;10:913077.
[17] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol 2010;28(3):245–8.
[18] Schellenberger J, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. Nat Protoc 2011;6(9):1290–307.
[19] Seaver SM, et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. Nucleic Acids Res 2021;49(D1):D575–88.
[20] Yuan Q, et al. Pathway-consensus approach to metabolic network reconstruction for Pseudomonas putida KT2440 by systematic comparison of published models. PLoS One 2017;12(1):e0169437.
[21] Gilis D, et al. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. Genome Biol 2001;2:1–12.
[22] Hormoz S. Amino acid composition of proteins reduces deleterious impact of mutations. Sci Rep 2013;3(1):2919.
[23] Balakrishnan R, et al. Principles of gene regulation quantitatively connect DNA to RNA and proteins in bacteria. Science 2022;378(6624):eabk2066.
[24] Puchałka J, et al. Genome-scale reconstruction and analysis of the Pseudomonas putid a KT2440 metabolic network facilitates applications in biotechnology. PLoS Comput Biol 2008;4(10):e1000210.
[25] Pramanik J, Keasling J. Effect of Escherichia coli biomass composition on central metabolic fluxes predicted by a stoichiometric model. Biotechnol Bioeng 1998;60(2):230–8.
[26] Zumft WG. Cell biology and molecular basis of denitrification. Microbiol Mol Biol Rev 1997;61(4):533–616.
[27] Oberhardt MA, et al. Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. Am Soc Microbiol; 2008.
[28] Monk JM, et al. i ML1515, a knowledgebase that computes Escherichia coli traits. Nat Biotechnol 2017;35(10):904–8.

[29] Jüngst A, Braun C, Zumft WG. Close linkage in Pseudomonas stutzeri of the structural genes for respiratory nitrite reductase and nitrous oxide reductase, and other essential genes for denitrification. Mol Gen Genet MGG 1991;225(2):241–8.
[30] Dückers N, et al. Threonine aldolases—screening, properties and applications in the synthesis of non-proteinogenic β-hydroxy-α-amino acids. Appl Microbiol Biotechnol 2010;88:409–24.
[31] Kertesz MA. Riding the sulfur cycle – metabolism of sulfonates and sulfate esters in Gram-negative bacteria. FEMS (Fed Eur Microbiol Soc) Microbiol Rev 2000;24(2):135–75.
[32] Guzmán GI, et al. Model-driven discovery of underground metabolic functions in Escherichia coli. Proc Natl Acad Sci USA 2015;112(3):929–34.
[33] Shi Z, et al. Enzyme commission number prediction and benchmarking with hierarchical dual-core multitask learning framework. Research 2023;6. 0153.
[34] Bogorad IW, et al. Building carbon-carbon bonds using a biocatalytic methanol condensation cycle. Proc Natl Acad Sci U S A 2014;111(45):15928–33.
[35] Yang X, et al. Systematic design and in vitro validation of novel one-carbon assimilation pathways. Metab Eng 2019;56:142–53.
[36] Chambers L, Trudinger P. Cysteine and S-sulphocysteine biosynthesis in bacteria. Archiv fuer Mikrobiol 1971;77:165–84.