For publication

**Construction and analysis of a genome-scale metabolic network for
Bacillus licheniformis WX-02**

Jing Guo[#], Hong Zhang[#], Cheng Wang, Ji-Wei Chang, Ling-Ling Chen^{*}

Agricultural Bioinformatics Key Laboratory of Hubei Province, College of Informatics,

Huazhong Agricultural University, Wuhan 430070, P. R. China

[#]These authors contributed equally to this work.

* Correspondence

Ling-Ling Chen

Tel: +0086 27 87280877

Fax: +0086 27 87280877

E-mail: Jing Guo - gj30501@163.com; Hong Zhang - 499946534@qq.com; Cheng Wang - wangchengmandy@126.com;

Ji-Wei Chang - longkaichang@163.com

*Correspondence and reprints: Ling-Ling Chen - llchen@mail.hzau.edu.cn

Abstract

We constructed the genome-scale metabolic network of *Bacillus licheniformis* WX-02 by combining genomic annotation, high-throughput phenotype microarray (PM) experiments and

---

literature-based metabolic information. The accuracy of the metabolic network was assessed by an OmniLog PM experiment. The final metabolic model iWX1009 contains 1,009 genes, 1,141 metabolites and 1,762 reactions, and the predicted metabolic phenotypes showed an agreement rate of 76.8% with experimental PM data. In addition, key metabolic features such as growth yield, utilization of different substrates and essential genes were identified by flux balance analysis. A total of 195 essential genes were predicted from LB medium, among which 149 were verified with the experimental essential gene set of *B. subtilis* 168. With the removal of 5 reactions from the network, pathways for poly- γ -glutamic acid (γ -PGA) synthesis were optimized and the γ -PGA yield reached 83.8 mmol/h. Furthermore, the important metabolites and pathways related to γ -PGA synthesis and bacterium growth were comprehensively analyzed. The present study provides valuable clues for exploring the metabolisms and metabolic regulation of γ -PGA synthesis in *B. licheniformis* WX-02.

Keywords: *Bacillus licheniformis* WX-02; Metabolic network; Flux balance analysis; Phenotype microarray; Poly- γ -glutamic acid

1. Introduction

Bacillus licheniformis is a ubiquitous Gram-positive bacterium capable of surviving in a broad range of natural environments. It can be used to produce a number of important industrial products, such as antibiotics [1], poly- γ -glutamic acid (γ -PGA) [2], various

enzymes, etc. [3-6]. At present, the *B. licheniformis* WX-02 genome has been sequenced [7, 8]. Some studies have reported that *B. licheniformis* WX-02 can produce γ -PGA under normal conditions and has a much higher yield under high salt conditions [8, 9]. Up to now, genes (*pgsB*, *pgsC*, *pgsA* and *pgsE*) related to γ -PGA synthesis have been reported in *Bacillus subtilis* and *B. licheniformis*. Although a series of molecular and cellular studies have been performed on *B. licheniformis*, the synthesis and regulation mechanisms of γ -PGA remain largely unknown. To explore the whole metabolic network, especially for metabolites and metabolic pathways related to γ -PGA synthesis and regulation in *B. licheniformis* WX-02, we constructed its genome-scale metabolic network.

Generally, cellular pathways in vivo are organized through networks. As metabolic networks can help to understand the mechanisms of biological activities and contain information on metabolites, metabolic reactions, chemical formulae, charges and the relationships between genes, proteins and reactions, cellular responses to gene knockout, medium conditions and environmental changes can be rapidly and accurately predicted based on metabolic networks. To date, genome-scale metabolic models have been widely used to study the metabolisms of various organisms, such as *B. subtilis* [10] and *Escherichia coli* (*E. coli*) [11]. In the present study, we first constructed the genome-scale metabolic network model of *B. licheniformis* WX-02 (iWX1009) and then corrected it based on the OmniLog phenotypic microarray (PM) experiment and literature [12]. Finally, the predicted metabolic network contains 1,009 genes, 1,141 metabolites and 1,762 reactions, and can serve as a valuable resource for the study of the metabolism and genetic mechanisms of *B. licheniformis* WX-02.

2. Materials and methods

2.1. Construction of the genome-scale metabolic network

The initial draft metabolic network of *B. licheniformis* WX-02 was constructed based on its genomic annotation (reference information was derived from GenBank: CP012110) and the metabolic model of *B. subtilis* 168 (iBsu1103). First, we searched orthologous proteins between *B. subtilis* 168 and *B. licheniformis* WX-02 using BLASTP (E values $\leq 1 \times 10^{-5}$, sequence identity $\geq 35\%$, alignment coverage of both query and subject sequence $\geq 70\%$) and retained the best hits. To obtain more comprehensive information, especially for the genes with no orthologs in *B. subtilis* 168, we retrieved additional genome annotation of *B. licheniformis* from KEGG database using KAAS [13] and BASys [14]. As genome annotation can provide information on gene-protein-reaction (GPR) association, we added the genome annotation into the initial draft metabolic model. To improve the accuracy of the predicted model, we retrieved information on relevant genes and reactions from PubChem [15], BRENDA [16], KEGG [17] and BsubCyc [18] databases and added it to the network. In addition, transport reactions from TransportDB [19] and TCDB [20] databases and organism-specific reactions from the literature were also included in the metabolic network. All additional reactions were charge-balanced to improve the reliability of the metabolic network.

2.2. Correction of the predicted metabolic network model

To ensure the simulation accuracy of the metabolic network model, we assessed the network model and made some corrections. This process includes the following steps. (1)

Detecting the quality and charge balance for all biochemical reactions. Generally, missing and redundant H^+ and H_2O can lead to an imbalance, so we corrected the reactions by adding or removing these two substrates. (2) Checking the reversibility of the reaction direction. The direction was determined according to the eleven rules proposed by Ma and Zeng [21]. As the metabolic network was constructed based on an iBsu1103 model, we also referenced it with the BsubCyc database of *B. subtilis* 168 [18]. For those reactions without reference information, we defaulted them as reversible reactions. (3) Detecting and filling in the metabolic gap. To complete this process, we searched the corresponding genes and reactions by analyzing the KEGG pathway map and checked biochemical reactions in references and books. (4) Testing the effectiveness of the metabolic network model. We determined whether the network model could predict the synthesis of all biomass precursors during growth in standard medium. If all biomass precursors could be synthesized, the biomass flux was simulated on standard medium, and the correlation of the simulation and experimental results was checked. On the contrary, if not all biomass precursors could be synthesized, the simulation of growth rate was considered zero. In this case, we needed to add some reactions to promote the synthesis of all biomass precursors. In general, the correction of the metabolic network model was an iterative process.

2.3. Generation of the biomass objective function

The biomass objective function (BOF), which plays a key role in a metabolic network and has a direct impact on the calculation of the growth rate in a simulated process [22], is the basis for further simulation analysis. The biomass equation was generated by defining major and essential components in target species that compose the cellular biological

macromolecules such as DNA, RNA, protein, etc. Each macromolecule was divided into its corresponding building blocks, i.e. amino acids, fatty acids and nucleotides [22]. Since the experimental biomass composition of *B. licheniformis* WX-02 was not available, we employed the equation of *B. subtilis* 168 metabolic model iBsu1103 instead. The reaction related to biomass production in the iBsu1103 metabolic model is divided into seven building blocks: DNA synthesis, RNA synthesis, protein synthesis, lipid content, lipoteichoic acid synthesis, cell wall synthesis and biomass synthesis [10].

2.4. OmniLog PM experiment

We performed the OmniLog PM experiment to analyze substrate utilization in *B. licheniformis* WX-02 [12]. The PM experiment included 190 carbon sources, 95 nitrogen sources, 59 phosphate sources and 35 sulfur sources. All procedures were performed following the manufacturer's instruction. The change of color caused by tetrazolium dye was used to detect whether *B. licheniformis* WX-02 cell could respire actively, and then the utilization of different substrates could be measured by the active respiration in minimal medium.

2.5. Simulation of the metabolic network model

After obtaining the accurate metabolic network model of *B. licheniformis* WX-02, we simulated it with stoichiometric matrix formalism S ($S = m \times n$, here m and n represent the number of metabolites and reactions of the network, respectively, $m < n$). The metabolic capabilities of *B. licheniformis* WX-02 were calculated with flux balance analysis (FBA) performed using the COBRA Toolbox with glpk solver in a Matlab environment [23]. For growth stimulation, biomass synthesis was selected as the objective function to be

maximized, and optimization was performed using linear programming techniques. The essential genes were predicted by FBA in glucose M9 minimal medium and LB medium, with a glucose uptake rate of 5 mmol/gDW/h and an oxygen uptake rate of 18.5 mmol/gDW/h, respectively. If removal of gene A could make the flux of BOF become zero, this gene was defined as an essential gene. In addition, possible carbon, nitrogen, phosphorus and sulfur sources supporting growth of *B. licheniformis* WX-02 were also identified using FBA.

2.6. Optimization of the pathway for γ -PGA production

The OptKnock program of the COBRA toolbox in Matlab was applied to optimize the metabolic pathways of γ -PGA synthesis [24]. The OptKnock program includes two optimization strategies, i.e., the maximum value of the objective function in vivo and in vitro. In this study, we regarded the maximization of γ -PGA yield as the objective function in vitro and chose the biomass equation as target function in vivo to optimize the metabolic pathways of γ -PGA synthesis under aerobic conditions.

3. Results and discussion

3.1. Metabolic network construction for *B. licheniformis* WX-02

Since *B. licheniformis* WX-02 and *B. subtilis* 168 are highly similar in their genome (Fig. 1A), and the genome-scale metabolic network model iBsu1103 of *B. subtilis* 168 has been well established, we constructed the metabolic network for *B. licheniformis* WX-02 based on its genomic annotation and the iBsu1103 model. In this way, the draft metabolic network contained 945 orthologous genes and 1,586 reactions. To make the network more credible and perfect, genes annotated as metabolic enzymes by KAAS and/or BASys, transport

reactions obtained from TransportDB and TCDB and reactions obtained from high-throughput biological phenotyping experiments were all added to the constructed metabolic network. Given that some gaps in the network might result in the existence of unconnected regions, we combined GapFind [25] and CytoHubba (a plugin in Cytoscape) [26] with information from the literature and genome to fill the gaps. Finally, metabolic network model iWX1009 of *B. licheniformis* WX-02 consists of 1,009 genes, 1,762 reactions and 1,141 metabolites (Fig. 1B, Table 1, Supplementary File S1).

All reactions and metabolites were transformed into Systems Biology Markup Language (SBML) format. Cytoscape [26] was applied to visualize and analyze the topological properties of the metabolic network, and CytoHubba was used to remove the isolated metabolites and reactions. Based on this process, the network finally contained 3,432 metabolites and 7,508 reactions (Fig. 1C). Almost all the biological networks in model organisms are characterized as scale-free networks [27], and the topological properties of the metabolic network for *B. licheniformis* WX-02 indicate that it is also a scale-free network (Fig. 1D). We used clusters of orthologous groups (COG) to classify the 1,009 genes in iWX1009 (Fig. 2A) [28]. These genes mainly fell into the functional categories of ‘energy production and metabolism (C)’, ‘amino acid transport and metabolism (E)’, ‘carbohydrate transport and metabolism (G)’, ‘coenzyme transport and metabolism (H)’ and ‘inorganic ion transport and metabolism (P)’, which account for 67% of the total genes, demonstrating that the iWX1009 model contains the basic metabolites and metabolic pathways for sustaining the growth. Nevertheless, there are only a few genes belonging to the category of ‘secondary metabolite biosynthesis, transport and catabolism (Q)’, which may be due to the fact that

studies of secondary metabolites are rather limited, even for the model organism *E. coli*. Fig. 2B depicts the classification of metabolic pathways and the corresponding number of reactions in iWX1009. The number of reactions related to membrane transport were the largest, followed by those of carbohydrates, fatty acids and lipids.

3.2 Analysis of *in silico* simulation and high-throughput substrate utilization

With the constructed iWX1009 model, the aerobic growth rate of *B. licheniformis* WX-02 in various nutrients (carbon, nitrogen, phosphorus and sulfur sources) could be calculated with BOF and simulated by setting each kind of source as the unique source with an uptake rate of 5 mmol/g cell/h. To verify the iWX1009 model, we determined additional carbon, nitrogen, phosphorus and sulfur sources that supported simulated growth of *B. licheniformis* WX-02 in minimal medium by a PM experiment (Supplementary Table S1). Finally, 220 of the 379 substrates (190 for carbon sources, 95 for nitrogen sources, 59 for phosphorus sources and 35 for sulfur sources) were tested under aerobic conditions based on M9 glucose minimal medium. The comparison showed an inconsistency between simulated results and PM results (Table 2). For instance, PM results showed that *B. licheniformis* WX-02 could utilize L-fucose, L-rhamnose or glycogen as the unique carbon source, while *in silico* growth stimulation results indicated that these substrates could not be used as carbon sources. To make the network model more comprehensive, we added a number of missing reactions supported by the literature and KEGG database. In this way, the results of *in silico* computation were then identical to those of the PM experiment. After reconciling with biological experimental data, iWX1009 is able to predict the metabolic phenotypes of *B. licheniformis* WX-02 with an overall agreement rate of 76.8%, which is

slightly higher than the 74% for iBsu1103 [10] and 76% for iAF1260 (*E. coli* K12 MG1655) [11]. These results demonstrate that the metabolic network can accurately simulate metabolism in *B. licheniformis* WX-02.

Although we made great efforts to refine the network model, the utilization of 51 substrates (21 for carbon sources, 21 for nitrogen sources, 5 for phosphorus and 4 for sulfur sources) among the 220 substrates was still inconsistent. These inconsistencies can be divided into two cases: 1) Some compounds could be utilized in PM experiments, whereas they were not predicted by in silico simulation, such as 2, 3-butanediol, which may be due to the lack of transport and metabolic reactions of these compounds. Therefore, with the accumulation of biochemical metabolism data of *B. licheniformis* WX-02, these reactions can be added to the network to make it more accurate. 2) The utilization of some compounds could be predicted by in silico simulation, while they could not be detected in PM experiments, indicating that there might be a deficiency in the metabolic model. As construction was mainly based on genome annotation and iBsu1103 model, the predicted metabolic network may contain all possible metabolic reactions under different environments, nutritional conditions and physiological cycles, and thus is static and oversaturated.

3.3. Comparison of substrate utilization between *B. licheniformis* WX-02 and other organisms

With more and more whole genomes being sequenced and annotated, corresponding metabolic network models have been continuously constructed, such as the iBsu1103 model for *B. subtilis* 168 and iAF1260 for *E. coli* K12 MG1655. Here we compared and evaluated similarities and differences in PM data in *B. licheniformis* WX-02, *B. subtilis* 168 and *E.*

coli K-12 MG1655. PM data of *B. subtilis* 168 and *E. coli* K-12 MG1655 were obtained from the study of Feist et al. [11] and Oh et al. [29], respectively. The comparison showed that there are 60 different substrates in these three species (Table 3). For these substrates, 19 substrates are only utilized by *B. licheniformis* WX-02, while 5 substrates are only utilized by *B. subtilis* 168 and *E. coli* K-12 MG1655. The different abilities of substrate utilization may be due to the specific genes or operons in different organisms, which can regulate the metabolic process of the corresponding compounds. For example, the PM experiments showed that D-allose can be utilized as a carbon source in *B. subtilis* 168 and *E. coli* K-12 MG1655, but not in *B. licheniformis* WX-02, which may be caused by the lack of genes related to D-allose in the *B. licheniformis* WX-02 genome. Nevertheless, some phenomena could not be explained by this. For example, PM experiments showed that *B. licheniformis* WX-02 could not use L-serine as the unique nitrogen source, but its genome has associated enzymes and metabolic genes. These phenomena may due to the fact that organisms tend to choose particular metabolic reactions under different culture conditions and external environments.

3.4. Prediction of essential genes

Essential genes control the central metabolic pathways in cells and play a significant role in cell growth [30]. Given that *B. licheniformis* WX-02 is an important industrial strain, we attempted to predict its essential gene set. In the present study, 235 and 195 essential genes were detected in M9 glucose minimal medium and LB-rich medium respectively (Supplementary Table S2), among which 186 genes were shared by both mediums, indicating that if a gene is essential in LB medium, it is very likely to be an essential gene in M9

medium as well. The predicted essential genes were classified based on COG category and they were mainly enriched in macromolecular synthesis (Fig. 2C).

To further analyze the credibility of the essential genes identified by in silico simulation, we compared them with those essential genes that had been verified by experiments. As no essential gene set has been reported in *B. licheniformis* WX-02 and the essential genes in *B. subtilis* 168 have been comprehensively studied [30], we compared the 195 essential genes obtained from LB medium with the 271 experimental essential genes of *B. subtilis* 168. Finally, 149 essential genes were verified and the rest of the 46 predicted essential genes may be specific to *B. licheniformis* WX-02, which needs to be further validated.

3.5. Optimization of the pathway for γ -PGA yield maximization

We used the OptKnock program of the COBRA toolbox in the Matlab to optimize the metabolic pathways of γ -PGA synthesis, and found that the biomass could be maximized and the γ -PGA yield could reach up to 83.8 mmol/h with removal of 5 reactions from the metabolic network (Table 4). Among these reactions, rxn00800 and rxn00097 are involved in nucleoside and nucleotide metabolism, while the other two reactions, rxn05164 and rxn05299, are related to histidine transport. These reactions play an important role in the cellular process of microbial metabolism. The arginine ammonia enzyme (EC: 3.5.3.6) related to the rxn00047 reaction can work under anaerobic conditions to help *B. licheniformis* WX-02 utilize arginine as an energy source. Each of the above five reactions is associated with one gene, which might play a key role in γ -PGA metabolism. The associated genes are mainly related to adenylosuccinate lyase (rxn00800), adenylate kinase (rxn00097), ABC transporter (rxn05164), amino acid permease (rxn05299) and arginine deiminase (rxn00047).

3.6. Important metabolites and metabolic pathways related to γ -PGA synthesis and cellular growth

A typical metabolic network always contains some small molecule metabolites related to electronics or functional groups (methyl group, phosphate group, one carbon unit, etc.). These metabolites are involved in many reactions within the network and are located in the center, but they cannot provide real and meaningful information for further study. Thus, before analyzing the important metabolites of the predicted network, we removed the general metabolites identified with reference [31]. Table 5 lists 16 important metabolites involved in the largest number of metabolic reactions, including amino acids, pyruvate, phosphoenolpyruvate (PEP), α -oxoglutarate, formate and tetrahydrofaolate. These metabolites play important roles in sustaining the fundamental metabolism and producing a high yeild of γ -PGA in *B. licheniformis* WX-02.

Here we take the metabolisms of three important metabolites as examples to illustrate their roles: (1) The metabolism of glucose-6-phosphate can release a large amount of ATP to maintain the normal growth of *B. licheniformis* WX-02. (2) Pyruvate can be transformed into various metabolites under aerobic condition, such as Ala, acetyl-coA and oxaloacetate, and these metabolites enter the tricarboxylic acid (TCA) cycle to form L-glutamate, which is subsequently synthesized into γ -PGA. These processes are also quite vital for maintaining the growth of bacterium and producing γ -PGA. (3) α -Oxoglutarate is highly necessary for γ -PGA synthesis, since it is an important material for the formation of L-glutamate (Fig. 2D). The above three metabolites are interconnected and are quite important for γ -PGA metabolism. The contentration of glucose-6-phosphate will directly affect the formation of pyruvate, and

the increase in pyruvate will promote the synthesis of α -oxoglutarate through the TCA cycle; then, a large amount of L-glutamate will be formed to produce γ -PGA.

Fig. 2D shows that γ -PGA is formed from the polymerization of L-glutamate and D-glutamate. As is well known, D-glutamate cannot be directly absorbed by cells and it can only be converted from L-glutamate in two ways. One is by L-glutamate being converted to D-glutamate by glutamate racemase; the other way is by L-alanine first being converted to D-alanine, which is then transformed into D-glutamate through D-alanine aminotransferase. For intracellular L-glutamate, it is mainly produced through the following processes; (1) absorption of exogenous L-glutamate from the culture medium; (2) de novo synthesization from glucose by TCA cycle; in this pathway, L-glutamine and α -oxoglutarate are synthesized into L-glutamate under the catalysis of glutamate synthase (corresponding to the reaction of rxn00085); (3) transformation from other amino acids; for example, the reaction of aspartate and α -oxoglutarate can produce L-glutamate and oxaloacetic acid under catalysis of glutamic-oxalacetic transaminase; arginine is hydrolyzed into urea and ornithine by arginase and then transformed into glutamic acid- γ -semialdehyde by ornithine aminotransferase, which is finally converted to L-glutamate by the glutamic acid- γ -semialdehyde dehydrogenase; proline can also be transformed into glutamate and α -oxoglutarate. It has been reported that proline plays an important roleb under salt and drought conditions in most bacteria, and it may partially inhibit b expression of γ -PGA-degrading enzymes [32].

In conclusion, we constructed genome-scale metabolic network model iWX1009

---

containing 1,009 genes, 1,141 metabolites and 1,762 reactions in *B. licheniformis* WX-02. The metabolic model was validated with a PM experiment, which shows an overall agreement rate of 76.8% with phenotypic microarray data. Although the predicted model is far from perfect, it provides a platform for understanding *B. licheniformis* WX-02 metabolism and pathways related to γ -PGA synthesis.

The differences in PM data in *B. licheniformis* WX-02, *E. coli* K-12 MG1655 and *B. subtilis* 168 show phenotype discrepancies among them; 19 substrates can only be used by *B. licheniformis* WX-02, while 5 substrates are only utilized by *B. subtilis* 168 and *E. coli* K-12 MG1655. Through in silico simulation of gene essentiality, we predicted 195 essential genes from the LB medium which were then verified with the experimental essential gene set from *B. subtilis* 168. Finally, 149 essential genes were verified and the other 46 may be specific to *B. licheniformis* WX-02. Based on metabolic model iWX1009, we analyzed metabolisms related to γ -PGA synthesis and bacterium growth, and found that glucose-6-phosphate, pyruvate and α -oxoglutarate are highly important, and can sustain the fundamental metabolism and ensure a high yield of γ -PGA in *B. licheniformis* WX-02. The metabolic network described here will provide valuable information for research on metabolism and γ -PGA synthesis in *B. licheniformis* WX-02.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (31271406 and 31071659), the Program for New Century Excellent Talents in University (NCET-13-0807) and the National Science Foundation of Hubei Province (2015CFA044).

Reference

1. McInerney MJ, Javaheri M, Nagle DPJr. Properties of the biosurfactant produced by *Bacillus licheniformis* strain JF-2. J Ind Microbiol 1990;5:95-101.
2. Konglom N, Chuensangjun C, Pechyen C, Sirisansaneeyakul S. Production of Poly- γ -glutamic acid by *Bacillus licheniformis*, Synthesis and Characterization. Journal of Metals, Materials and Mineral 2012;22:7-11.
3. Veith B, Herzberg C, Steckel S, Feesche J, Maurer KH, Ehrenreich P, et al. The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential. J Mol Microbiol Biotechnol 2004;7:204-211.
4. Pötter M, Oppermann-Sanio FB, Steinbüchel A. Cultivation of bacteria producing polyamino acids with liquid manure as carbon and nitrogen source. Appl Environ Microbiol 2001;67:617-622.
5. Liu YF, Zhang SL, Yong YC, Ji ZX, Ma X, Xu ZH, et al. Efficient production of acetoin by the newly isolated *Bacillus licheniformis* strain MEL09. Process Biochem 2011;46:390.
6. Qi G, Kang Y, Li L, Xiao A, Zhang S, Wen Z, et al. Deletion of meso-2,3-butanediol dehydrogenase gene budC for enhanced D-2,3-butanediol production in *Bacillus*

-
- licheniformis*. Biotechnol Biofuels 2014;7:16.
7. Yangtse W, Zhou Y, Lei Y, Qiu Y, Wei X, Ji Z, et al. Genome sequence of *Bacillus licheniformis* WX-02. J Bacteriol 2012;194:3561-3562.
  8. Guo J, Cheng G, Gou XY, Xing F, Li S, Han YC, et al. Comprehensive transcriptome and improved genome annotation of *Bacillus licheniformis* WX-02. FEBS Lett 2015;589:2372-2381.
  9. Wei X, Ji Z, Chen S. Isolation of halotolerant *Bacillus licheniformis* WX-02 and regulatory effects of sodium chloride on yield and molecular sizes of poly-gamma-glutamic acid. Appl Biochem Biotechnol 2010;160:1332-1340.
  10. Henry CS, Zinner JF, Cohoon MP, Stevens RL. iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. Genome Biol 2009;10:R69.
  11. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 2007;3:121.
  12. Bochner BR, Gadzinski P, Panomitros E. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res 2001;11:1246-1255.
  13. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 2007; 35:W182-185.
  14. Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, et al. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res 2005;33:W455-459.
  15. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: a public information

-
- system for analyzing bioactivities of small molecules. *Nucleic Acids Res* 2009;37:W623-633.
16. Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, et al. BRENDA, the enzyme information system in 2011. *Nucleic Acids Res* 2011;39:D670-676.
17. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 2008;36:D480-484.
18. Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 2014; 42:D459-D471.
19. Ren Q, Chen K, Paulsen IT. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res* 2007;35:D274-279.
20. Saier MHJr, Tran CV, Barabote RD. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* 2006;34:D181-186.
21. Ma H, Zeng AP. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 2003;19:270-277.
22. Feist AM, Palsson BO. The biomass objective function. *Curr Opin Microbiol* 2010;13:344-349.
23. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol* 2010;28:245-248.
24. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for

-
- identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 2003;84:647-657.
25. Satish Kumar V, Dasika MS, Maranas CD. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* 2007;8:212.
26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-2504.
27. Albert R. Scale-free networks in cell biology. *J Cell Sci* 2004;118:4947-4957.
28. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 2001;29:22-28.
29. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J Biol Chem* 2007;282:28791-28799.
30. Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, et al. Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci U S A* 2003;100:4678–4683.
31. Durek P, Walther D. The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles. *BMC Syst Biol* 2008;2:100.
32. Hahne H, Mäder U, Otto A, Bonn F, Steil L, Bremer E, et al. Comprehensive proteomics and transcriptomics analysis of *Bacillus subtilis* salt stress adaptation. *J Bacteriol* 2010;192:870-882.

Figure legends

Fig. 1. Overview of the prediction of the metabolic network of *B. licheniformis* WX-02.

(A) Two-dimensional similarity plots to compare the distribution of orthologs on the chromosomes of *B. licheniformis* WX-02 and *B. subtilis* 168. A straight line of dots suggests co-linearity of chromosomes between the two species. (B) The flow chart for constructing the metabolic network of *B. licheniformis* WX-02. (C) Metabolites of the network are colored according to their classification of metabolic pathways; therefore, metabolites with the same color belong to the same pathway. (D) Degree distribution of the metabolic network.

Fig. 2. Classification of genes, reactions included in iWX1009, important metabolites and metabolic pathways related to γ -PGA synthesis and cell growth.

(A) COG functional categories of genes in iWX1009. (B) The number of reactions (both gene-associated and non-gene-associated) associated with each functional class. (C) Distribution of the essential genes according to their classification of metabolic pathways. (D) Important metabolites and metabolic pathways related to γ -PGA synthesis and cellular growth.

Table 1. Properties of the metabolic network model iWX1009

Features	Number
Genes	1009 (22%)
Reactions	1762
Metabolic reactions	1105
Cytoplasmic	1080
Extracellular	25
Transport reactions	296
Exchange reactions	263
Gene-associated reactions	1248
Unique metabolites	1141
Cytoplasmic	1136
Extracellular	262

Table 2. Simulation of the utilization of different substrates in *B. licheniformis* WX-02

Sources	Compounds	Consistency		Inconsistency		Consistency
	with exchange					rate (%)
	reactions					
		E-G	E-NG	E-NG	E-G	
		C-G	C-NG	C-G	C-NG	
Carbon	111	85	5	2	19	81.0
Nitrogen	61	30	10	8	13	67.2
Phosphor	36	31	0	0	5	86.1
Sulfur	12	7	1	1	3	66.7
Total	220	153	16	11	40	76.8

Note: E, experimental; C, computational; G, growth; NG, no growth.

Table 3. Differences in metabolic phenotypes among *E. coli* K-12 MG1655, *B. subtilis* 168 and *B. licheniformis* WX-02

Source	Compound	Present in iWX1009	WX-02	Bsu168	Eco	Note:
C source	g-Amino-N-Butyric acid	EX_4abut_e	G	NG	NG	G, grow th; NG, no grow th.
C source	L-Methionine	EX_met_L_e	G	NG	NG	
C source	L-Ornithine	EX_orn_e	G	NG	NG	
C source	Putrescine	EX_ptrc_e	G	NG	NG	
C source	L-Tartaric acid	EX_tartr_L_e	G	NG	NG	
C source	Tyramine	EX_tym_e	G	NG	NG	
C source	Acetamide	EX_ad_e	G	NG	NG	
C source	N-Acetyl-L-glutamate	EX_acglu_e	G	NG	NG	
C source	L-Isoleucine	EX_ile_L_e	G	NG	NG	
C source	L-Leucine	EX_leu_L_e	G	NG	NG	
C source	L-Lysine	EX_lys_L_e	G	NG	NG	
C source	L-Phenylalanine	EX_phe_L_e	G	NG	NG	
C source	L-Valine	EX_val_L_e	G	NG	NG	
C source	D,L-Carnitine	EX_crn_e	G	NG	NG	
C source	2,3-Butanediol	EX_btd_RR_e	G	NG	NG	
C source	2,3-Butanone	EX_Diacetyl_e	G	NG	NG	
C source	3-Hydroxy-2-butanone	EX_actn_R_e	G	NG	NG	
N source	N-Acetyl-D-Galactosamine	EX_acgal_e	G	NG	NG	
N source	Inosine	EX_ins_e	G	NG	NG	
C source	2'-Deoxyadenosine	EX_dad_2_e	NG	G	G	
C source	b-D-Allose	EX_all_D_e	NG	G	G	
N source	L-Serine	EX_ser_L_e	NG	G	G	
N source	L-Citrulline	EX_citr_L_e	NG	G	G	
N source	D-Glucosamine	EX_gam_e	NG	G	G	

Table 4. Yield of γ -PGA calculated with OptKnock under the glucose medium

Target	Rxn-list	Enzyme (EC)	Product (mmol/hr)	Biomass (l/hr)
	rxn00800	4.3.2.2 (N ₆ -(1,2-Dicarboxyethyl) AMP AMP_lyase		
	rxn05164	L_Histidine_ABC transport		
γ -PGA	rxn00097	2.7.4.3 (adenylate kinase)	83.8	6.71
	rxn05299	HIS _t 2		
	rxn00047	3.5.3.6 (L-arginine deiminase)		

Table 5. Metabolites involved in the largest number of metabolic reactions in *B. licheniformis*

WX-02 metabolic network		
Name	Abbreviation in network	Degrees
Pyruvate	M_pyr_c	54
L_Glutamate	M_glu_L_c	52
2_Oxoglutarate	M_akg_c	28
L_Serine	M_ser_L_c	20
L_Glutamine	M_gln_L_c	17
Phosphoenolpyruvate	M_pep_c	17
L_Aspartate	M_asp_L_c	16
D_fructose_6_phosphate	M_f6p_c	15
Glycine	M_gly_c	14
L_Lysine	M_lys_L_c	14
Formate	M_for_c	13
L_Alanine	M_ala_L_c	12
D_Glucose_6_phosphate	M_g6p_c	12
L_Methionine	M_met_L_c	11
Tetrahydrofolate	M_thf_c	11
L_Cysteine	M_cys_L_c	10

A **B** **C** **D**

A

B

C

D