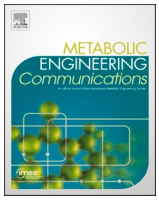
Metabolic Engineering Communications 19 (2024) e00251

Contents lists available at ScienceDirect

Metabolic Engineering Communications

journal homepage: www.elsevier.com/locate/mec



Reconstruction and analyses of genome-scale *halomonas* metabolic network yield a highly efficient PHA production

Luhui Zhang [a,1], Xinpei Sun [a,b,1], Jianwen Ye [c,1], QianQian Yuan [d,1], Xin Zhang [e], Fei Sun [f], Yongpan An [a], Yutong Chen [a], Yuehui Qian [a], Daqian Yang [a], Qian Wang [a], Miaomiao Gao [a], Tao Chen [e,****], Hongwu Ma [d,***], Guoqiang Chen [c,**], Zhengwei Xie [a,b,g,*]

[a] *Peking University International Cancer Institute, School of Basic Medical Sciences, Peking University, Beijing, 100191, China*
[b] *Peking University - Yunnan Baiyao International Medical Center, School of Pharmaceutical Science, Peking University, Beijing, 100191, China*
[c] *MOE Key Lab of Bioinformatics, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, 100084, China*
[d] *Biodesign Center, Key Laboratory of Engineering Biology for Low-carbon Manufacturing, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China*
[e] *Key Laboratory of Systems Bioengineering (Ministry of Education), School of Chemical Engineering and Technology, Tianjin University, Tianjin, China*
[f] *School of Pharmacy, University of Wisconsin Madison, WI, 53705, USA*
[g] *Gigaceuticals Co., Ltd, Beijing, 102206, China*

A R T I C L E I N F O

Keywords:
PHA
Halomonas
Metabolic network
HCSA
Batch fermentation

A B S T R A C T

In pursuit of reliable and efficient industrial microbes, this study integrates cutting-edge systems biology tools with *Halomonas bluephagenesis* TD01, a robust halophilic bacterium. We generated the complete and annotated circular genome sequence for this model organism, constructed and meticulously curated a genome-scale metabolic network, achieving striking 86.32% agreement with Biolog Phenotype Microarray data and visualize the network via an interactive Electron/Thrift server architecture. We then analyzed the genome-scale network using vertex sampling analysis (VSA) and found that productions of biomass, polyhydroxyalkanoates (PHA), citrate, acetate, and pyruvate are mutually competing. Recognizing the dynamic nature of *H. bluephagenesis* TD01, we further developed and implemented the hyper-cube-shrink-analysis (HCSA) framework to predict effects of nutrient availabilities and metabolic reactions in the model on biomass and PHA accumulation. We then, based on the analysis results, proposed and validate multi-step feeding strategies tailored to different fermentation stages. This integrated approach yielded remarkable results, with fermentation culminating in a cell dry weight of 100.4 g/L and 70% PHA content, surpassing previous benchmarks. Our findings exemplify the powerful potential of system-level tools in the design and optimization of industrial microorganisms, paving the way for more efficient and sustainable bio-based processes.

1. Introduction

Microbial production of polyhydroxyalkanoates (PHAs) offers a sustainable alternative to petroleum-based plastics due to their biodegradability and versatile properties (Anderson and Dawes, 1990; Doi, 1990; Lee, 1996; Poirier et al., 1995; Steinbuchel and Fuchtenbusch, 1998). Among PHA-producing microbes, *Halomonas bluephagenesis* TD01, isolated from a salt lake in China, shows exceptional promise (Chen, 2009; Chen et al., 2017; Lan et al., 2016; Tan et al., 2011). This halophile thrives under non-sterile, high-salinity, and continuous

* Corresponding author. Department of Pharmacology and Institute of Systems Biomedicine, School of Basic Medical Sciences, Peking University, Beijing, 100191, China.
** Corresponding author.
*** Corresponding author.
**** Corresponding author.
E-mail addresses: chentao@tju.edu.cn (T. Chen), ma_hw@tib.cas.cn (H. Ma), chengg@mail.com.tsinghua.edu.cn (G. Chen), xiezhengwei@hsc.pku.edu.cn (Z. Xie).
[1] Co-first authors contribute equally in this study.

https://doi.org/10.1016/j.mec.2024.e00251
Received 25 January 2024; Received in revised form 9 November 2024; Accepted 11 November 2024
Available online 19 November 2024

culture conditions, significantly reducing production costs associated with sterilization. Notably, engineered strains of *H. bluephagenesis* TD01 can produce not only the standard P(3HB) but also copolymers like poly (3-hydroxybutyrate-co-3-hydroxyvalerate) (PHBV) and poly (3-hydroxybutyrate-co-4-hydroxybutyrate) [P(3HB-co-4HB)] (Lan et al., 2016) with enhanced mechanical flexibility (Fu et al., 2014).

While recombinant strains of *Escherichia coli* and other species have been employed for P(3HB-co-4HB) production (Tan et al., 2011), their susceptibility to contamination necessitates complex and energy-intensive sterilization. Unlike these organisms, *H. bluephagenesis* TD01's natural tolerance to high pH and salt environments acts as a built-in defense against contaminants, simplifying processing and reducing resource needs (Chen and Jiang, 2018; Yin et al., 2015). This unique characteristic combined with its robust growth under non-sterile conditions establishes *H. bluephagenesis* TD01 as a superior candidate for large-scale, cost-effective PHA production and opens doors for various industrial applications.

However, *Halomonas bluephagenesis* TD01's full biomanufacturing potential remains untapped. A major roadblock lies in the lack of comprehensive metabolic knowledge. To overcome this, we present the first complete and annotated genome sequence of *H. bluephagenesis* TD01, alongside a meticulously curated genome-scale metabolic network (GSMN). The GSMN enabled us to apply advanced constraint-based approaches for strain optimization. Traditional flux balance analysis methods are limited in their ability to predict metabolic behavior and identify optimal strain designs. To overcome these limitations, we previously developed an algorithm called the Hyper-Cube Shrink Algorithm (HCSA) (Xie et al., 2018). HCSA assigns a pair of parameters to each enzyme to represent its activity, enabling the solution space to be reduced based on enzymatic activity rather than a single objective function. This approach allows us to evaluate the flux distribution between multiple outputs, improving the accuracy of our predictions.

In this study, we have constructed a genome-scale metabolic network based on the whole genome sequence. We developed a map and visualization tool for the entire metabolic network. We analyzed metabolic prosperities of *H. bluephagenesis* TD01 with the VSA and HCSA algorithm. Based on the analysis, we proposed and validate multi-step feeding strategies tailored to different fermentation stages, achieving a higher PHA yield.

2. Material and methods

2.1. Genome library preparation and sequencing

We sequenced the genomic DNA using the Illumina Hiseq 4000 and the Pacbio RSII platform, separately. The sequencing library for the Illumina Hiseq 4000 was prepared by BGI Genomics using their in-house reagents. From the resulting libraries, 7.9 million fragments were sequenced in paired end reads with a read length of 90 nt. The sequencing library for the Pacbio RSII was prepared using the SMRTbell template prep kit v1.0, and then 44,274 polymerase reads were sequenced with an average length of 14356 bp. The quality and quantity of libraries were checked by Agilent 2100, qPCR, and Qubit.

2.2. Genome assembly

The long reads from the Pacbio RSII were assembled into contigs using SMRT Analysis v2.3.0 (Pacbio Inc.). Then the hybrid assembly of the Illumina-derived short reads and PacBio-derived contigs was done using Celera Assembler to generate a complete genome (Istrail et al., 2004).

2.3. Genome size estimation using k-mer analysis

A k-mer count analysis was done using Jellyfish (version 2.3.0) on the Illumina data (Marçais and Kingsford, 2011). From the paired end reads, only the first read was used in this analysis, and the second read was omitted from this analysis to avoid counting overlapping k-mers. K-mer sizes were set as 15. After converting the k-mer counts into a histogram format, this file was analyzed using the Genomescope tool (Vurture et al., 2017).

2.4. Full genome comparison for identifying evolutionarily conserved regions

The whole genome comparison between *H. bluephagenesis* TD01 and some other bacteria were conducted using Mauve to identify the evolutionarily conserved regions (ECR) (Darling et al., 2004).

2.5. Genome annotation

For the annotation, we utilized multiple bioinformatics tools, and then we merged and manually curated results from different sources (Supplementary Fig. 1). We first performed genome annotation using RAST (Overbeek et al., 2014) and Prokka (Version 1.7.1) (Seemann, 2014). These two software generated annotation results fully automatically, including annotations for protein coding sequences (CDSs), rRNAs and tRNAs. Besides, we also employed Glimmer (Version: 3.02) (Delcher et al., 2007), RNAmmer (Version: 1.2) (Lagesen et al., 2007), tRNAScan-SE (Version: 1.3.1) (Lowe and Eddy, 1997), Rfam (Version 9.1) (Griffiths-Jones et al., 2005)/Infernal (Nawrocki et al., 2009), CRISPRFinder (Version: 0.4) (Grissa et al., 2007) and Tandem Repeat Finder (Version: 4.04) (Benson, 1999) to complete structure annotations for CDSs, rRNAs, tRNAs, sRNAs, CRISPRs and Tandem Repeats, respectively. After that, we finished function annotations for CDSs using KEGG (Kanehisa et al., 2016), SwissProt (Magrane and UniProt, 2011), InterPro (Jones et al., 2014) and Gene Ontology (Ashburner et al., 2000).

After that, we considered merging results from different sources together. The first thing to merge was the structure annotations for the coding sequences (CDSs). Since we would use the RAST annotation to get the initial draft model in the following steps, in the event of a conflict, results from RAST took precedence over those from Glimmer and Prokka. Merging annotations from RAST and Glimmer, we believed different predictions were for the same CDS, if they met the following conditions: 1) they shared the same start site or end site, and 2) the length difference between predicted structures was less than or equal to 2% of that from RAST, and in the meantime 3) the predicted structures were on the same strand. For such CDSs, we saved function annotations from all different sources and the structure annotations only from RAST. In above situation, if length difference between two predicted structures was more than 2% of that from RAST, only the annotation from RAST would be saved for the CDS. When predictions from RAST and Glimmer overlapped each other and are on the same strand, but neither the start site nor the end site was the same, if the number of overlapping bases was a multiple of 3, only the annotation from RAST would be saved. In other situations, we simply put annotation results together. Then using the same method, we merged the above results with the structure annotations from Prokka.

Structure annotations for rRNAs and tRNAs were also merged in the same way. For rRNAs, when a conflict occurs, results from RNAmmer took precedence, then RAST, then Prokka. For tRNAs, annotations from tRNAScan-SE took precedence, then RAST, then Prokka.

We then merged function annotations from different sources for CDSs. Firstly, we merged annotations from RAST and KEGG. During this process, some databases including KEGG, Uniprot and so forth, were used to find out whether different texts from RAST and KEGG meant the same function. When function annotations from RAST and KEGG for one CDS were different, we would employ NCBI's BLAST tool to decide which annotation was more likely to be right. If annotation results from all three sources were different, we only saved annotations from RAST.

Then annotations from Prokka was the third important, then SwissProt, then InterPro, and then Gene Ontology. After completing the merging process, we might see some different genes have the same name. Therefore, we added a number (1, 2, 3, …) to their names to differentiate them based on the order of their locus.

2.6. Genome-scale metabolic network reconstruction

2.6.1. The initial draft model

The initial model was automatically constructed using the KBase platform (Arkin et al., 2018). Since the KBase could accept the genome annotation from RAST rather than our own curated version, we provided the RAST annotation for the initial construction. Afterword, we curated the model based on our curated genome annotation.

2.6.2. Initial refinement of the model

To construct a high-quality metabolic model and ensure the simulation accuracy, we manually inspected, assessed, and refined the model. The first issues we encountered were writing formats of the metabolite names, some of which cannot be identified by the COBRA (Heirendt et al., 2019), the matlab toolbox for metabolic network simulation. Non-standard writing formats of metabolite names includes spaces, brackets, and apostrophe. All these, therefore, were substituted with underscores. Besides, all the NH3 in the model were changed into NH4+ since it is aquatic inside the bacteria. Also, all the "_c0[c0]" and "_e0[e0]" in the model were changed into "[c0]" and "[e0]", respectively, to make it simplified.

In the preliminary revision process, the COBRA toolbox was used to check the reaction balance. Unbalanced reactions were corrected based on multiple databases, including KEGG, UniProt, BiGG, SEED and PubChem. Also, we checked the reaction direction using the method described below.

2.6.3. Biomass equation

The biomass reaction, which represents cellular growth, is a critical indicator of model quality and accuracy. In our draft model from KBase, a generic biomass reaction formula is included; however, it has not been verified against real-world experimental data. To define the biomass equation for *Halomonas bluephagenesis* TD01, we conducted an extensive literature review. Despite numerous studies on the genetic modifications and production capabilities of this strain, no reports were found on its specific cell composition or component content. We expanded our search to the genus *Halomonas* and identified *Chromohalobacter salexigens* DSM 3043, a γ-proteobacterium closely related to TD01 in phylogenetic studies. Notably, genome-scale metabolic models iOA584 (Ates et al., 2011) and iFP764 (Piubeli et al., 2018) for *C. salexigens* DSM 3043 were published in 2011 and 2018, respectively, with iFP764 demonstrating higher quality. We adopted the biomass composition data and biomass equation from iFP764 for *Halomonas bluephagenesis* TD01. Furthermore, for moderate halophiles like *H. bluephagenesis* TD01, ectoine and its hydroxylated derivative, 5-hydroxyectoine, are the two primary compatible solutes that they preferentially synthesize and accumulate in high osmotic environments to maintain intracellular osmotic balance (Liu et al., 2021), also identified and annotated four ectoine-related genes in *H. bluephagenesis* TD01, underscoring its osmoregulatory adaptations (Cai et al., 2011). Therefore, to better reflect TD01's metabolic characteristics, we incorporated ectoine and 5-hydroxyectoine as precursor metabolites in our biomass equation. This adjustment aims to capture the strain's ability to withstand high salt stress.

2.6.4. Gap filling

Given that no exchange reactions exist in the draft model from KBase, which represent the substances exchange between the media and bacteria, we added exchange reactions for metabolites distributed in the extracellular compartment of the model. 18 of them were set to be reversible (see Supplementary File 5), which provided necessary nutrition for the organism, including the exchange reactions of glucose, oxygen, carbon dioxide, nitrogen, sulfur, phosphorus, trace elements and metal ions. The low and upper bound of the flux of these reactions were set as $-1000\ \mathrm{mmol}/(g_{DW}\cdot h)$) and $1000\ \mathrm{mmol}/(g_{DW}\cdot h)$), separately, to mimic unrestricted conditions, except the glucose exchange reaction, whose lower bound was set to be $-10\ \mathrm{mmol}/(g_{DW}\cdot h)$), which was the same with the *E. coli* models. All other exchange reactions were only allowed to carry fluxes out of the organism.

Besides, with missing reactions in the synthetic pathway(s) of their correlative metabolites, the draft model has poor simulation accuracy or even no prediction results can be attained. Hence iterative manual inspection and curation for the gaps existing in the metabolic pathways were conducted. In practice, gap filling accounts for a significant portion of the model refinement. For this process, the first step is calculating the optimal growth rate using the glucose as the sole carbon source to identify the biomass components that can't be synthesized or consumed. Then, through checking and comparing with the individual metabolic maps obtained from the E coli models, KEGG, and BiGG (King et al., 2016) databases to determine where the gap is and prepare reactions for the gap filling. After mapping these reactions to the corresponding GPR relationships and verifying their directions using the method described below, they were added into the model artificially and the gap(s) were eliminated. For reactions added during gap filling, please see Supplementary File 5.

2.6.5. Eliminate unrealistic reaction fluxes

After gap filling, the biomass flux was not zero anymore, but too large. We, therefore, manually inspect the model again, and found some unreasonable reactions existing inside the model, for example, the reaction rxn09240, "H2O[c0] + ATP[c0] + GTP[c0] + Sulfate[c0] <-> Phosphate[c0] + PPi[c0] + GDP[c0] + H + [c0] + APS[c0]". Enzyme that catalyzes this reaction is the sulfate adenyltransferase, which could transfer the adenyl group from ATP to the sulfate and generate the adenylyl sulfate (APS). If this reaction goes reversely, it could produce ATP and GTP in a large amount unreasonably, which might contribute to the large biomass flux. So, we manually changed the reaction to be irreversible. In total, 17 reactions were changed to be irreversible to avoid unrealistic reaction fluxes (see Supplementary File 5).

2.6.6. OmniLog phenotype MicroArray™ experiment

The Phenotype MicroArray™ (Biolog) experiment (PM experiment) was performed to analyze the capacity of *Halomonas bluephagenesis* TD01 to utilize various carbon sources in the minimal medium and also help refine the in-silico model. On the 96-well plates used in this experiment, the first microwell (A1) served as the negative control, and in other microwells, bacteria were cultured with different carbon sources. In total, the Biolog plates covered 190 carbon sources. For each microwell, we got the value "Max" by following the manufacturer's' protocol, which represents the metabolic rate of bacteria. Then three criteria were set for correctly identifying the possible carbon source for the bacteria growth: 1) If the value of one microwell is lower than the control, and the dyed well is very light, the corresponding carbon source was directly regarded as unusable; 2) If the value is apparently higher than the control and the dyed well is very dark, the represented compound was directly deemed useable; 3) If the value is slightly higher and very close, and the color depth is difficult to differentiate under this circumstance, the represented was also classified as unusable. After that, based on the results which carbon source is the one that the bacteria could utilize and which one is not, we refined our model (see Supplementary File 5 for reactions we modified and added).

2.6.7. Reactions for PHA production

PHA related reactions in the model were checked, and the missing reactions were added.

2.6.8. Thermodynamic consistency analysis

The standard Gibbs free energy change of reactions, $\Delta_r G'^o$, can be estimated using the component contribution method (Noor et al., 2013). The calculation was based upon the standard condition with pH at 7.0, temperature at 298.15 K, zero ionic strength and 1 M concentrations of all species except H+ and water. However, the 1 M reference state for the metabolite concentrations does not reflect the metabolite concentrations found in the cell (approximately 1 mM) (Feist et al., 2007). Thus, $\Delta_r G'^m$, which represented the free energy change of reaction at 1 mM concentrations for all species except H+, water, H_2, O_2 and CO_2, was calculated based on $\Delta_r G'^o$. The reference concentrations for H_2, O_2 and CO_2 are their saturation concentrations in water at 1 atm and 298.15 K, i.e. 0.000034, 0.000055 and 0.0014 M, respectively (Feist et al., 2007). The $\Delta_r G'^m$ of transport reactions also include the energy contribution of the transmembrane electrochemical potential and proton gradient (Fleming et al., 2009; Henry et al., 2006).

Since intracellular metabolite concentrations are typically 0.00001–0.02 M (Albe et al., 1990), the actual free energy change of a reaction, $\Delta_r G'$, can differ significantly from $\Delta_r G'^m$. Herein, the maximum and minimum of $\Delta_r G'^m$ were calculated using the method stated by Feist A.M. et al. (Feist et al., 2007). The ranges of $\Delta_r \boldsymbol{G}'$ were further used in the model construction process to assess reaction reversibility and directionality. Reactions with exclusively negative $\Delta_r \boldsymbol{G}'$ values were identified as thermodynamically irreversible in the forward direction and vice versa. The reactions with both positive and negative $\Delta_r \boldsymbol{G}'$ values were identified as thermodynamically reversible.

2.7. Fermentation for Halomonas bluephagenesis TD01

The *Halomonas bluephagenesis* TD01 was activated on 60LB agar plates from −80 °C glycerol stocks. Single colonies were picked and inoculated into 100 mL shake flasks containing 20 mL of 60LB medium. Once cultures reached an OD600 of 2–3, 1% (v/v) of the primary seed culture was used to inoculate 50 mL of 60LB medium in a 500 mL conical flask. These cultures were incubated on a rotary shaker at 200 rpm and 37 °C for 7–8 h until the OD600 reached 5–6. Subsequently, 10% (v/v) of the secondary seed culture was used to inoculate 3 L of 60 MMG medium (Ye et al., 2018a) in a 7.5 L bioreactor (Bioflo3000, New Brunswick, USA). The pH of the fermentation medium was maintained at 8.5 by adding 5M NaOH.

Cell growth was monitored by measuring the optical density at 600 nm (OD600) using a WPA Biowave DNA spectrophotometer (Biochrom, UK), with appropriate dilution to ensure raw OD600 readings between 0.3 and 0.8. For PHA content analysis, 30 mL of the culture broth was sampled into pre-weighed 50 mL test tubes and the cells were harvested by centrifugation at 10000 rpm for 15 min at room temperature (Hitachi, Japan). The supernatant was discarded, and the cell pellets were washed twice with distilled water. The cells were then frozen at −80 °C for 1.5–2 h and then lyophilized for 18–24 h using a lyophilizator (SCIENTZ-18N, Ningbo, China). The cell dry weight (CDW) was determined using a conventional analytical scale. Approximately 40 mg of lyophilized cell powder was methyl esterified by adding 2 mL of a solution containing 3% concentrated 98% sulfuric acid and 1 g/L benzoic acid in methanol, and 2 mL of chloroform. The mixture was incubated at 100 °C for 3.5 h in sealed 10 mL glass tubes in a hot metal bath. After methanolysis, the PHA content was analyzed by gas chromatography (GC) using a GC-2014 instrument (Shimadzu, Japan), following the method described by Tan et al. (2011).

3. Results

3.1. Genome sequence of Halomonas bluephagenesis TD01

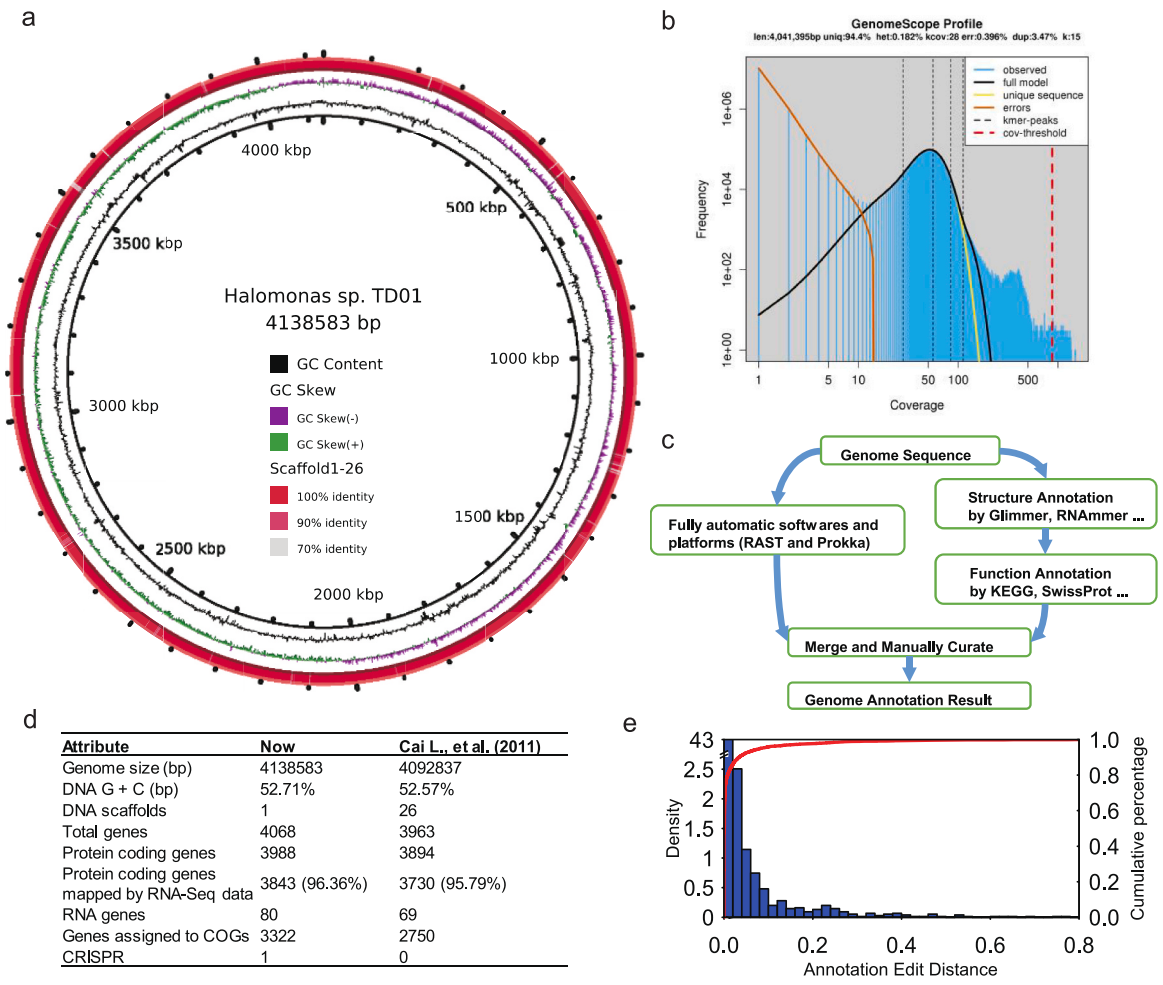
With the combination of the Illumina Hiseq 4000 and Pacbio RSII sequencing technologies, having an average sequencing depth of 133.97 and 104.10, separately (Supplementary Fig. 2), we were able to determine the whole genome sequence of *H. bluephagenesis* TD01 (Supplementary File 1), which consists of 4,138,583 bases pairs (bp) at 99.89% accuracy, rather than 26 scaffolds reported several years ago (Cai et al., 2011). In Fig. 1a, the previous scaffolds were mapped to our genome assembly using NCBI BLAST (highlighted in red). The inner black circle represents our complete genome assembly (Fig. 1a). We then evaluated sequence integrity and completeness of our genome assembly using K-mer analysis (k = 15 here) and the Hiseq data. The whole genome size was estimated to be about 4.04 M (Fig. 1b), which is close to the genome size we actually got.

Our genome assembly of *Halomonas bluephagenesis* TD01 consists of 4,138,583 base pairs, representing a 1.12% increase in total size compared to the previous draft genome. The GC content is 52.71%, slightly higher than the previously reported 52.57%. We identified 170 repetitive sequences in the genome with an average length of 630.49 bp. The total length of repetitive sequences is 107,184 bp, covering 2.59% of the complete genome.

3.2. Genome annotation of Halomonas bluephagenesis TD01

The genome of *Halomonas bluephagenesis* TD01 was annotated using a combination of multiple bioinformatics tools, and the annotations from different sources were mutually validated and manually merged to produce a final curated annotation (Fig. 1c and Supplementary File 2). This strategy allowed us to discover more information about the genome than we could from any single tool. In total, we identified 3988 protein-coding genes, which is more than any single tool could detect. The final curated annotation comprised 3814 protein-coding genes from RAST, 3268 protein-coding genes from Glimmer, and 3500 protein-coding genes from Prokka. Additionally, we identified 61 tRNA genes, 18 rRNA genes, and one snoRNA gene in the genome. We also predicted the presence of 1 CRISPR sequence and 223 tandem repeats in the genome. Then, to evaluate the completeness of the genome annotation, we identified evolutionarily conserved regions (ECRs) by alignment of the *H. bluephagenesis TD01*'s genome sequence to the genome sequences of *Halomonas elongata DSM 2581*, *Halomonas* sp. *R57-5*, *Halomonas* sp. *Y2R2* and *Halomonas titanicae GPM3* (Darling et al., 2004; Gregory et al., 2006). We found that 339 of the 354 ECRs conserved in all 5 genomes overlap with annotated genes, suggesting that gene annotation is about 95.8% complete (Supplementary File 3). The remaining 15 ECRs may represent additional exons without supporting evidence, or highly conserved regulatory and structural elements.

In comparison to the annotation result reported by Cai L. et al.(Cai et al., 2011), our annotation identified 4068 genes, representing a 2.65% increase in the total number of genes (Fig. 1d). Notably, the number of genes assigned to Clusters of Orthologous Groups (COGs) increased by 20.8%, indicating that our annotation provides more useful genome information (Fig. 1d). The GC content of protein-coding genes was found to be 53.35%, which is slightly higher than the GC content of the entire sequence (52.71%). The protein coding gene distributed across the coordinates and protein length distribution was shown in Supplementary Fig. 3. To identify the exact genes between the two genome annotations, we performed reciprocal best hits analysis, which revealed 3753 identical genes, 141 deleted genes, and 235 newly added genes (Supplementary File 4). The annotation edit distance (AED), ranging between 0 and 1, is able to quantify the amount of structural changes to individual genes between different annotation releases (Eilbeck et al., 2009). Among the genes of *H. bluephagenesis TD01*, 1775 of them have 0 AED value, and 3054 of them have AED values less than 0.01 (Fig. 1e–Supplementary File 4). However, 150 genes had AED values larger than 0.1. For genes with large AEDs, many "N"s were found in the previous version of the sequencing result, indicating that the exact nucleic acids were not assigned for this locus (Supplementary File 4). This highlights the superiority of the third-generation sequencing technology used in our sequencing project compared with the



Attribute	Now	Cai L., et al. (2011)
Genome size (bp)	4138583	4092837
DNA G + C (bp)	52.71%	52.57%
DNA scaffolds	1	26
Total genes	4068	3963
Protein coding genes	3988	3894
Protein coding genes mapped by RNA-Seq data	3843 (96.36%)	3730 (95.79%)
RNA genes	80	69
Genes assigned to COGs	3322	2750
CRISPR	1	0

Fig. 1. Whole genome sequence and annotation of H. Bluephagenesis TD01 and comparison with previous release. (a) A panoramic view of the entire genome, visualized using BRIG. GC content and GC skew are visualized. The bold outer circle delineates the previously reported fragmented sequence. (b) K-mer analysis for sequencing data using the GenomeScope software. (c) A concise depiction of the annotation pipeline. When annotating, fully automatic tools were used, including RAST (http://rast.nmpdr.org/rast.cgi) and Prokka (Version 1.7.1). Software used for structure annotation included Glimmer (Version: 3.02), RNAmmer (Version: 1.2), tRNAScan-SE (Version: 1.3.1), Rfam (Version 9.1)/Infernal, CRISPRFinder (Version: 0.4) and Tandem Repeat Finder (Version: 4.04). Databases employed for function annotation included KEGG, SwissProt, InterPro and Gene Ontology. Results from different sources were merged manually in the end. (d) Comparison of genome statistics. The average length of all protein-coding genes is 963.4, and 86.3% of protein-coding genes have a size within the range (100, 1500] (see Supplementary Fig. 3 for detailed distribution). (e) Density histogram and cumulative distribution curve underscore the accuracy of annotations, with minimal edit distances observed. Note the left vertical axis is noncontinuous between 2.5 and 43.

traditional next-generation sequencing technology used by Cai L. et al.

The obtained genome of *H. bluephagenesis* TD01 was 4,138,583 bp in size with a 52.7% G + C content. Of the 4068 total genes, 3988 are encoding proteins and 80 are ncRNAs. Compared with TD01's closest neighbors and its closely related stains, showed in The SEED Viewer and Genome Neighbor report gained from NCBI (https://www.ncbi.nlm.nih.gov/genome/neighbors), respectively, the genome size and the total gene number of *H. bluephagenesis* TD01 were at moderate level, while the GC content seemed to be slightly lower and on the contrary, the coding regions seemed to be slightly larger (Supplementary Table 1).

We used WEGO 2.0 (Ye et al., 2018b) to classify the genes in the whole genome of *Halomonas bluephagenesis* TD01 based on level 2 terms in three ontologies of the Gene Ontology (cellular component, molecular function, and biological process). The majority of genes were found to be involved in catalytic activity and metabolic processes, as opposed to other functions (Supplementary Fig. 4). Additionally, the KEGG pathway classification for protein coding genes was shown in Supplementary Fig. 5 it was obvious that the amount of coding sequences was 3988, of which assigned to "Carbohydrate metabolism" and "Amino acid metabolism" categories occupied the first two in the subsystem distribution, with 344 and 340 coding sequences, respectively. In addition, "Membrane Transport" category with 176 coding sequences accounted for a large proportion, this maybe have an intimate association with its ability to resist high salinity environments.

3.3. Genome-scale metabolic network of Halomonas bluephagenesis TD01

Based on the genome information, we constructed the metabolic model for *Halomonas bluephagenesis* TD01 as described in the method part, and the whole process was summarized in Supplementary Fig. 6 and Supplementary File 5.

The initial draft model from KBase was comprised of 1473 reactions, 1554 metabolites and 1049 genes. During the initial refinement process, the reaction formats were standardized and to be more simplified than original KBase formats, and directions of 103 reactions were changed according to the Gibbs free energy we calculated (Supplementary File 5). Biomass reaction was then curated as mentioned in the method part.

Categories of biomass precursors were listed in Supplementary File 6. When doing the in-silico simulation, however, biomass precursors could not be synthesized. Gap filling, therefore, was an essential step. 130 exchange reactions were first added for metabolites distributed in the extracellular compartment of the model, and then 81 reactions (including 15 transport reactions and 1 demand reaction) were added for filling the gaps in the metabolite pathways of biomass precursors (Supplementary File 5). Two reactions were changed from irreversible to reversible and one reaction was removed due to duplication with another one in the model (Supplementary File 5). One reaction was simplified in format (see Supplementary File 5). After these modifications, the biomass was able to be produced in the model, but the biomass flux was, however, too big. We, thus, re-checked the model carefully, and found some unreasonable reaction fluxes when doing the simulation (Supplementary File 5). As mentioned in the method part, we changed the directions for 17 reactions to fix this problem (Supplementary File

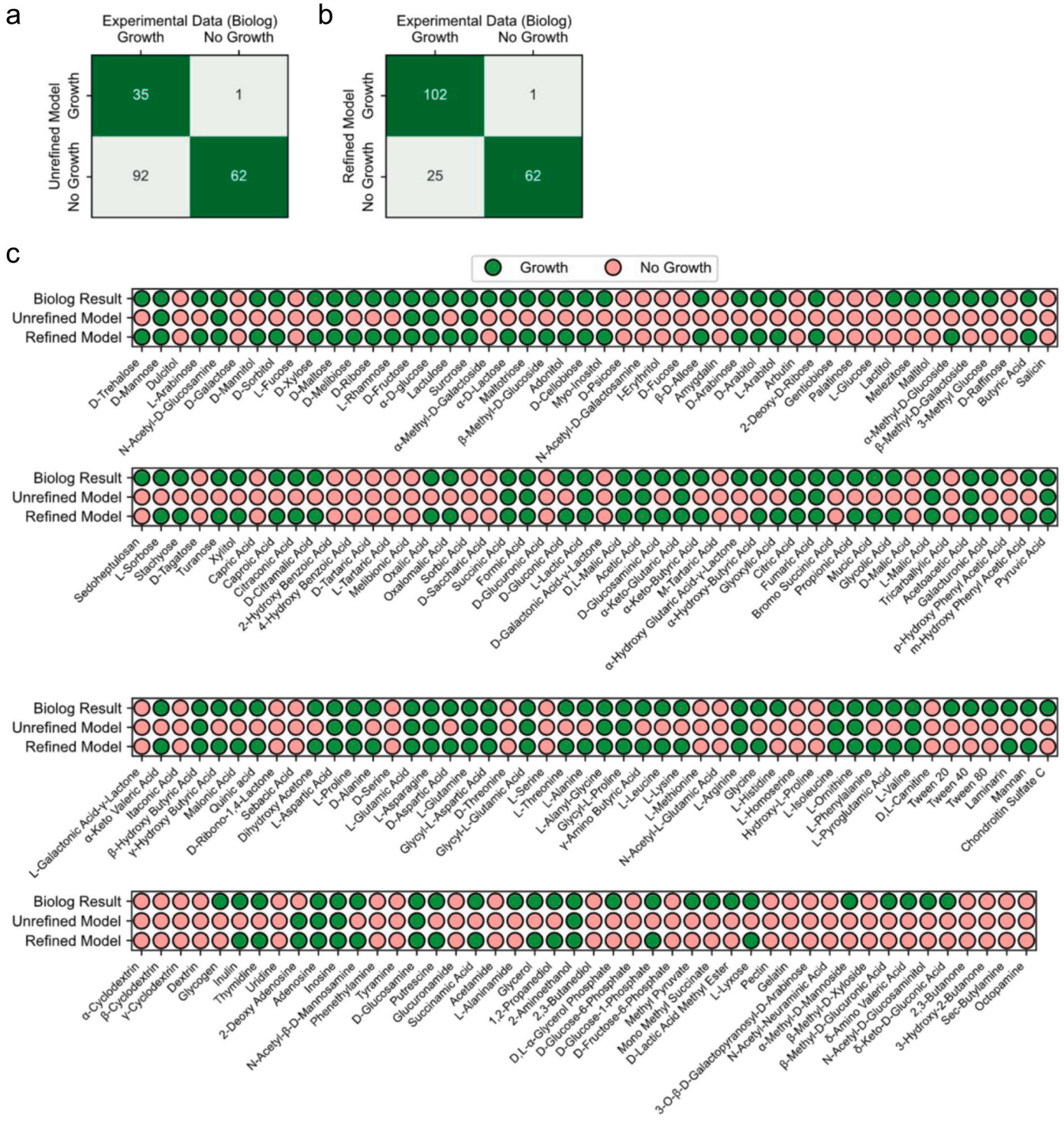


Fig. 2. Unveiling Metabolic Versatility of Halomonas bluephagenesis TD01 through Biolog Phenotype Microarrays (PM) and Model Refinement. (a and b) Confusion matrices comparing Biolog PM results with model predictions before and after model refinement based on the Biolog PM results. (c) Biolog PM results and model predictions before and after refinement on the 190 carbon sources.

5).

From genome annotation until fixing unreasonable reaction fluxes, all these steps were done computationally, whereas phenotypic data from known publications is lacking and insufficient to further refine and verify the model. Biolog Phenotype Microarray (PM) experiment in minimal medium was, therefore, employed to comprehensively determine growth capabilities of *H. bluephagenesis* TD01 on 190 different carbon sources (see Methods, Supplementary Files 5 and 7), which could be divided into the following eleven groups: Polymers (n = 12), Sugars and Sugar Derivatives (n = 66), Methyl Esters (n = 3), Carboxylic Acids (n = 40), Amides (n = 4), Amino Acids/Peptides and Related Chemicals (n = 32), Nucleosides and nucleotides (n = 5), Amines (n = 5), Alcohols (n = 4), Sugar Phosphates (n = 4) and Others (n = 15).

PM results showed that 127 compounds could be oxidized to support growth, while 63 couldn't (Fig. 2). Meanwhile, during the model simulation, we found 98 compounds are consistent with the PM experiment, which means the preliminary coincidence rate was 51.58%. Generally, the discrepancies can be sub-divided into two categories. The first is that some compounds couldn't be utilized for the in-silico growth, while PM results turned out they could. 92 compounds were included in this category, including turanose, D-fructose, D-mannitol, thymidine, putrescine and so forth, which probably due to metabolic gaps, incomplete gene annotations, or the lack of transport and/or exchange reactions. Such issues were solved by searching the metabolism information in online databases or literatures and adding related reactions to the model. The other category, which includes a compound called 2-deoxyadenosine, was shown by model simulations to support the growth of *Halomonas bluephagenesis* TD01, but not by PM experiments. This inconsistency might be due to wrong metabolic reactions/pathways in the model. This issue might be solved by improved genome annotations in the future help identifying incorrect reactions in the current model. However, currently, it is hard to identify the exact reactions that should not be in the model due to the lack of the experimental data. We, therefore, just leave the model unchanged, but will avoid using such compounds as the carbon source when doing simulations.

Despite the utmost efforts, 25 of the 93 inconsistent carbon sources remain unable to support in-silico growth, but this is within acceptable limits. Ultimately, the overall prediction consistency increases to 86.32% (164/190, Supplementary File 7, Fig. 2, and Table 1), which is similar to 85% for iFP764 (C. salexigens) (Piubeli et al., 2018), and higher than 76% for iAF1260 (*E. coli* K12 MG1655) (Feist et al., 2007). After that, we examined the reactions added previously for gap filling to check whether they are still needed or not for biomass production. If 1) the reaction was not needed for biomass anymore, 2) no genes were related to this reaction, and 3) its removal will not affect the consistency between model and PM experiment, we will remove the reaction since it was added purely for gap filling previously (Supplementary File 5).

In total, this update ended up with adding 194 new reactions (including 58 transport and 57 exchange reactions) and modifying 4 reactions for the 93 initially inconsistent carbon sources. 17 reactions (including 2 exchange reactions) previously added for gap filling were removed. Besides, for simulating PHA production using the model, PHA-related reactions were checked, and 7 new reactions were added (Supplementary File 5).

In summary, the final metabolic model including 1674 reactions (except 185 exchange reactions, 4 demand reactions, 3 sink reactions and the biomass reaction) and 1699 metabolites, accounts for two cellular compartments: the cytoplasm and extracellular space (Supplementary File 8 and Table 2). In total, 1072 genes were included in the metabolic network, which accounted for 26.89% of the genes in the whole genome of *H. bluephagenesis* TD01.87.4% reactions (1454 in 1674) are related to one or more genes, while the remaining 12.6% reactions are also included in the model in the form of reactions for filling gaps or transport reactions. From the perspective of gene coverage, our model resembles other published models (Supplementary Table 2), some of which belong to the microorganisms have been extensively studied or halophiles, such as *iJO1366* for *Escherichia coli* K-12 MG1655 (32% coverage (Orth et al., 2011b)), *iEM439* for *Zymomonas mobilis* ZM1 (22.7% coverage (Motamedian et al., 2016)), *iWX1009* for *Bacillus licheniformis* WX-02 (22% coverage (Guo et al., 2016)), *iFP764* for *Chromohalobacter salexigens* (22.4% coverage (Piubeli et al., 2018)), *iCKL708* for *Clostridium kluyveri* (17% coverage (Zou et al., 2018)), *iKYA1142* for *Halomonas smyrnensis* AAD6T (35.5% coverage (Diken et al., 2015)), the model for *Halobacterium salinarum* R-1 (17.09% coverage (Gonzalez et al., 2008)).

We categorized ORFs, reactions, and metabolites in the metabolic network based on clusters of orthologous groups (COGs). The number of ORFs from each COG functional class that were included in the metabolic network are shown in Fig. 3. Amino acid transport and metabolism has the highest number of ORFs, while lipid transport and metabolism has the highest percent coverage in the metabolic network. The classification of reactions in the metabolic network is shown in Fig. 3b, where lipid transport and metabolism is the most significant COG functional class. The classification of metabolites in the metabolic network is shown in Fig. 3c, where the lipid transport and metabolism class contains the most metabolites. Moreover, lipid transport and metabolism has the highest reaction to ORF ratio (5.1), followed by secondary metabolites biosynthesis, transport, and catabolism (3.7), indicating that these two classes have many multiple functional enzymes (Feist et al.,

Table 1
Model refinement based on the phenotype microarray (PM) experiment.

Carbon source category	Compound number	Initial consistency rate	Revised consistency rate
Polymers	12	4/12	7/12
Sugars and Sugar Derivatives	66	28/66	59/66
Methyl Esters	3	0/3	0/3
Carboxylic Acids	40	23/40	38/40
Amides	4	2/4	3/4
Amino Acids/Peptides and Related Chemicals	32	21/32	31/32
Nucleosides and nucleotides	5	3/5	4/5
Amines	5	3/5	4/5
Alcohols	4	2/4	4/4
Sugar Phosphates	4	2/4	3/4
Others	15	9/15	10/15
Total	190	97/190	164/190

Table 2
Features of the in-silico genome-scale metabolic model of Halomonas bluephagenesis TD01.

Features	Number
Genome	
Genome size (base pairs, bp)	4138583
No. of open reading frames (ORFs)	3988
No. of ORFs assigned in metabolic network	1067
ORF coverage (%)	26.76%
Reactions (except exchange/demand/sink/biomass reactions)	1674
Metabolic reactions	1467
Transport reactions	207
Gene-protein-reaction associations	
Gene associated (metabolic/transport)	1320/134 (87.4%)
No gene association (metabolic/transport)	147/73
Exchange reactions	185
Demand reactions	4
Sink reactions	3
Metabolites	1699
Unique metabolites	1522
Cytoplasmic	1514
Extracellular	185

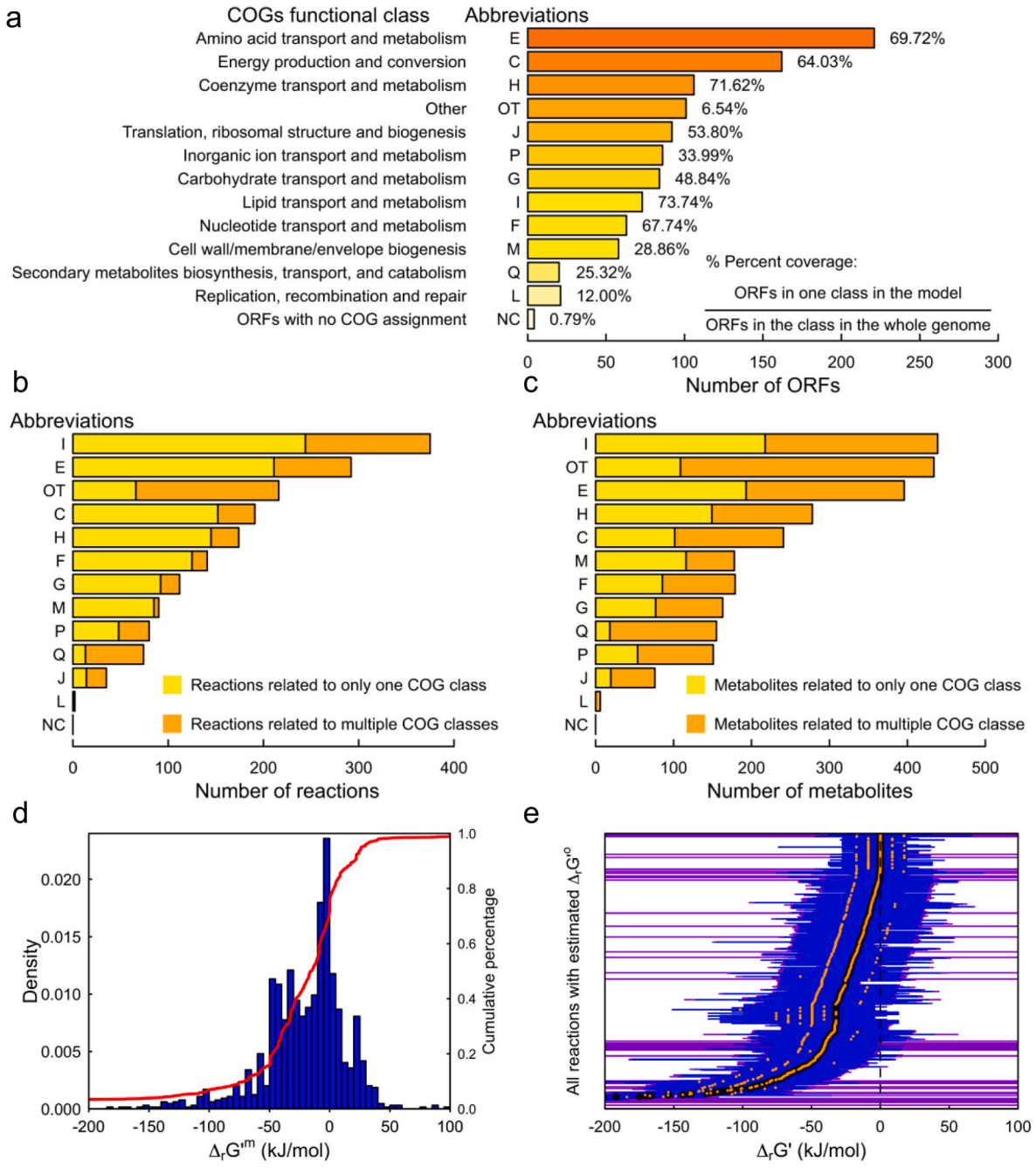


Fig. 3. Classification of the ORFs, reactions and metabolites; thermodynamic properties of reactions in the metabolic network of Halomonas bluephagenesis TD01. (a) Mapping Functional Domains in the Genome: This panel shows the coverage of characterized ORFs from each COG functional class included in the reconstructed metabolic network. Each percentage value represents the proportion of ORFs within a specific class relative to the entire genome. (b and c) Unveiling Reaction and Metabolite Landscape: These panels illustrate the distribution of reactions and metabolites within the network, categorized by their associated COG functional classes. Yellow bars represent unique reactions/metabolites within a class, while orange bars show those involved in multiple classes. Other and unassigned classes are grouped separately. (d) The distribution of estimated $\Delta rG'm$ values for the reactions in the metabolic network has been calculated for 1290 reactions (77.1%) in the reconstruction. 851 reactions have a negative $\Delta rG'm$, and 109 reactions have a zero $\Delta rG'm$. (e) The range of possible $\Delta rG'$ values for the reactions in the metabolic network. $\Delta rG'$ differs from $\Delta rG'o$ (black points) and $\Delta rG'm$ (orange diamonds) due to variations in metabolite concentrations from the 1 M and 1 mM reference states, respectively. Metabolite concentrations typically range from 0.00001 to 0.02 M, resulting in a wide range of values for $\Delta rG'$ (blue error bars). Taking uncertainty into account, the range of possible values for $\Delta rG'$ can be extended (purple error bars). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2007). Additionally, the highest number of unique metabolites involved in one class is cell wall/membrane/envelope biogenesis (53.37%), followed by coenzyme transport and metabolism (43.88%), indicating the specialized nature of the proteins in these pathways in *Halomonas bluephagenesis* TD01.

Following Heirendt, L. et al. (Heirendt et al., 2019), we identified 157 essential genes in the metabolic network, which account for 14.65% of the genes in the metabolic network and 3.94% of the genes in the whole genome. These essential genes were then categorized by eggNOG-Mapper (Huerta-Cepas et al., 2017), which assigned them to different COGs. Our findings revealed that the COG class "amino acid transport and metabolism" has the highest number of essential genes,

whereas the COG class "cell wall/membrane/envelope biogenesis" has the highest percentage of essential genes (Supplementary Fig. 7), indicating genes related to these pathways are crucial for the growth of *Halomonas bluephagenesis* TD01. In addition, we performed double gene deletion analysis as shown in Supplementary Fig. 7 and Supplementary Table 3.

3.4. Thermodynamic consistency analysis

The standard Gibbs free energy change of reactions, $\Delta_r G'^o$, was estimated using the component contribution method (Noor et al., 2013). The calculation was based on standard conditions with a pH of 7.0, temperature of 298.15 K, zero ionic strength, and 1 M concentrations of all species except H+ and water. However, the use of 1 M as the reference state for metabolite concentrations does not reflect the concentrations found in cells, which are typically around 1 mM (Feist et al., 2007). Therefore, $\Delta_r G'^o$, which represents the free energy change of the reaction at 1 mM concentrations for all species except H+, water, H2, O2, and CO2, was calculated based on $\Delta_r G'^o$. The reference concentrations for H2, O2, and CO2 are their saturation concentrations in water at 1 atm and 298.15 K, i.e., 0.000034, 0.000055, and 0.0014 M, respectively (Feist et al., 2007). The $\Delta_r G'^m$ values of transport reactions also include the energy contribution of the transmembrane electrochemical potential and proton gradient (Fleming et al., 2009; Henry et al., 2006). We found that 74.2% of the estimated $\Delta_r G'^m$ values are less than or equal to zero, indicating that the majority of the represented reactions are thermodynamically feasible at 1 mM metabolite concentrations (Fig. 3d).

The actual free energy change of a reaction, denoted as ΔG', can deviate significantly from the estimated ΔG'0 value when the intracellular metabolite concentrations differ from the standard 1 mM condition used for the calculation (Albe et al., 1990). To evaluate the range of ΔG' values that could be expected in vivo, we employed the method proposed by Feist et al. (2007) to estimate the maximum and minimum values of ΔG' for each reaction.

We then used the calculated ranges of ΔG' values to determine the thermodynamic reversibility of reactions in the metabolic network (Fig. 3e and Supplementary Table 4). Reactions with only negative ΔG' values were deemed thermodynamically irreversible in the forward direction, and those with only positive values were deemed irreversible in the reverse direction. Reactions with both positive and negative ΔG' values were considered reversible.

By comparing the predicted and estimated ΔG' values, we identified 108 reactions in the draft model that were initially specified as reversible but were found to be thermodynamically irreversible., and 1 reaction in the draft model having wrong direction (Supplementary Table 4). We corrected these reactions to match our thermodynamic estimates, resulting in 852 reversible and 822 irreversible reactions (the numbers do not include exchange/demand/sink/biomass reactions) in the final metabolic network.

3.5. A visualization application to access the metabolic network

To better visualize the metabolic network of *Halomonas*

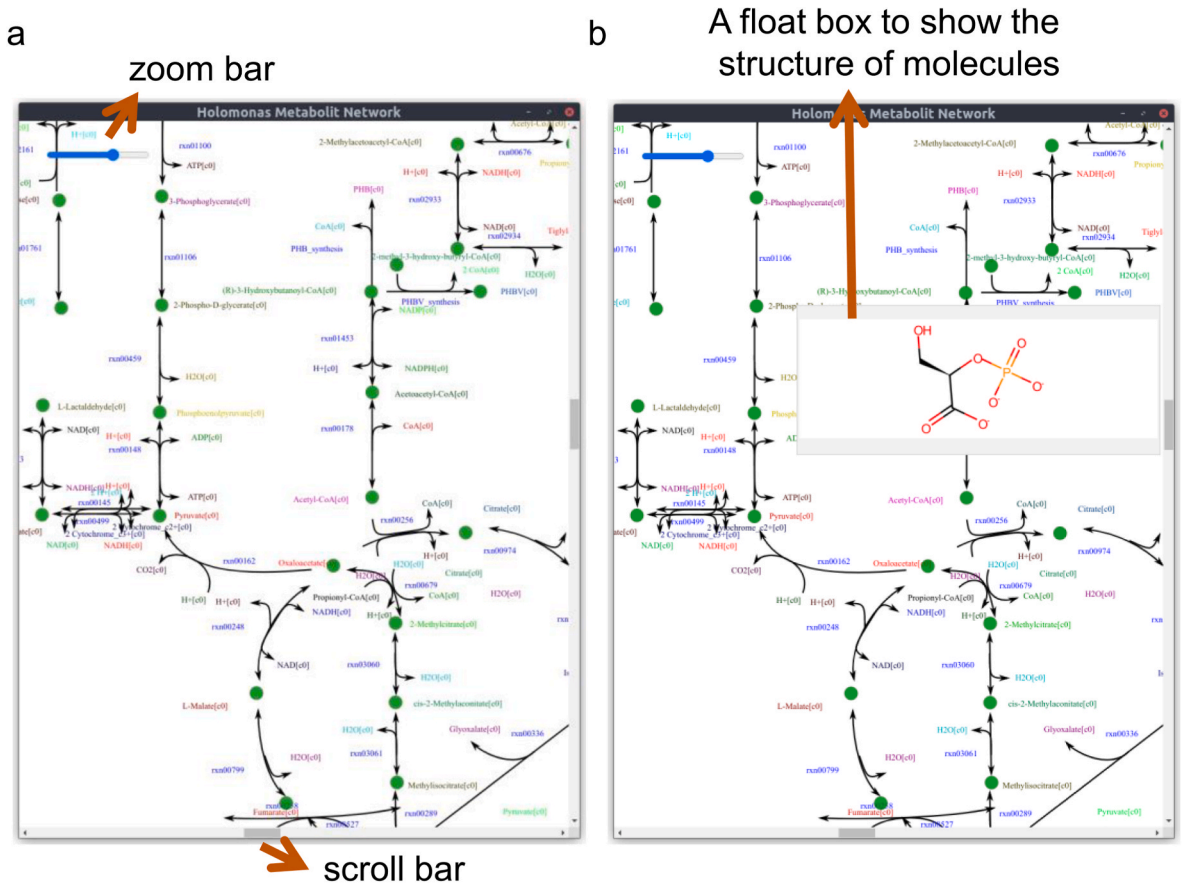


Fig. 4. Interactive Visualization of Metabolic Networks. This figure showcases the interface of our interactive visualization application for exploring metabolic networks. (a) Network Overview: This panel displays a portion of the metabolic network. Users can navigate the map by dragging and adjust the zoom level using the provided slider. Hovering over metabolite labels displays enlarged versions for better readability. (b) Molecule Details: Clicking on a metabolite triggers a pop-up window displaying the chemical structure of the corresponding molecule. This provides users with instant access to detailed information about each metabolite within the network.

bluephagenesis TD01, we constructed a metabolic network diagram in SVG format (Fig. 4). We first generated vector diagrams for each individual reaction using a Perl script, and then manually connected these reactions. This work was very labor-intensive. Then, we developed an application based on a client-server framework. The front end is built on Electron, and the back end uses a Thrift server based on Python. This allows us to browse the metabolic network in a scalable way on the front end, and to transmit data to the front end through the back end, for example, when we click on a metabolite, it can display its molecular image, which is generated by RDKit on the back end. We believe that this tool can help researchers better understand halophilic bacteria. The source code and application are deposited at github (https://github.com/kekegg/HalomonasBluephagenesisTD01).

3.6. Vertex sampling analysis for reducing byproducts

We have previously reported the real-time monitoring of the fermentation process of *Halomonas bluephagenesis* TD01, including the cell growth (OD600), PHA, and byproduct accumulation (Ye et al., 2018a). During the fermentation process, both OD600 and PHA accumulation increased simultaneously, while several byproducts, such as citrate, acetate, and pyruvate, were also produced.

We utilized the convex basis method (Bordel et al., 2010) to conduct vertex sampling analysis, which helped us uncover the impact of different metabolites (e.g., the byproduct citrate and acetate) on PHA output. This algorithm enables the exploration of the solution space in flux balance analysis by generating a two-dimensional projection of the steady-state flux space and analyzing the relationships between different products. Here, we performed vertex sampling analysis for some metabolites in our model. The mathematical expression of this algorithm is as follows:

$$\max \text{ or } \min f = v_i \cos\theta + v_j \sin\theta \; i, j \in N, \theta \in [0, \pi]$$

s.t.

$$S_{m\times n} v_n = 0 \; m, n \in N$$

$$l_z < v_z < u_z \; z \in N$$

Where $S_{m\times n}$ is a stoichiometric matrix of the network, and l_z and u_z is the lower and upper bounds of reaction z. θ was sampled from 0 to π when maximizing and minimizing f respectively.

Although we performed multiple samplings in different directions, the solution collapsed to a limited vertex. By plotting the solution vectors in Cartesian coordinates, we were able to obtain a two-dimensional projection of the steady-state flux space, which allowed us to analyze the relationships between two metabolites.

Using vertex sampling analysis, we investigated the relationships between biomass, PHA, and several byproducts identified during the fermentation process (Fig. 5a). The geometries in these six images in Fig. 5a all form lower-left triangles or similar shapes, indicating that these two metabolites, as products of the model, compete with each other. Therefore, to increase the yield of biomass and PHA, it is necessary to reduce the production of these byproducts and make more carbon available for biomass and PHA production. We also performed vertex sampling analysis between PHA and other metabolites, and a similar conclusion was reached (Supplementary Fig. 8). Furthermore, we conducted a three-dimensional vertex sampling analysis, as shown in Supplementary Fig. 9.

To explore ways to reduce the output of byproducts, we utilized traditional FBA for double nutrient analysis (Fig. 5b). The objective function was a linear combination of biomass and PHA. During the early stages of the fermentation process, when the proportion of PHA in the objective function is low, a large amount of oxygen is required to maximize the metabolic flux compared to the later stages as shown in Fig. 5b and Supplementary Movie 1. Therefore, a high ventilation rate for the culture system is necessary in the early stages, while the ventilation rate should be reduced later. Specifically, the amounts of glucose and oxygen consumed by *H. bluephagenesis* TD01 must meet the optimal values shown in Fig. 5b during different fermentation stages. This will result in the most efficient accumulation of biomass and PHA, with no other byproducts being produced, such as acetate and pyruvate. Based on the above analysis results, the fermentation medium could be optimized to increase PHA yield. Additionally, the analysis process was performed among some other nutrients (Supplementary Fig. 10), providing more information for optimization of the culture medium.

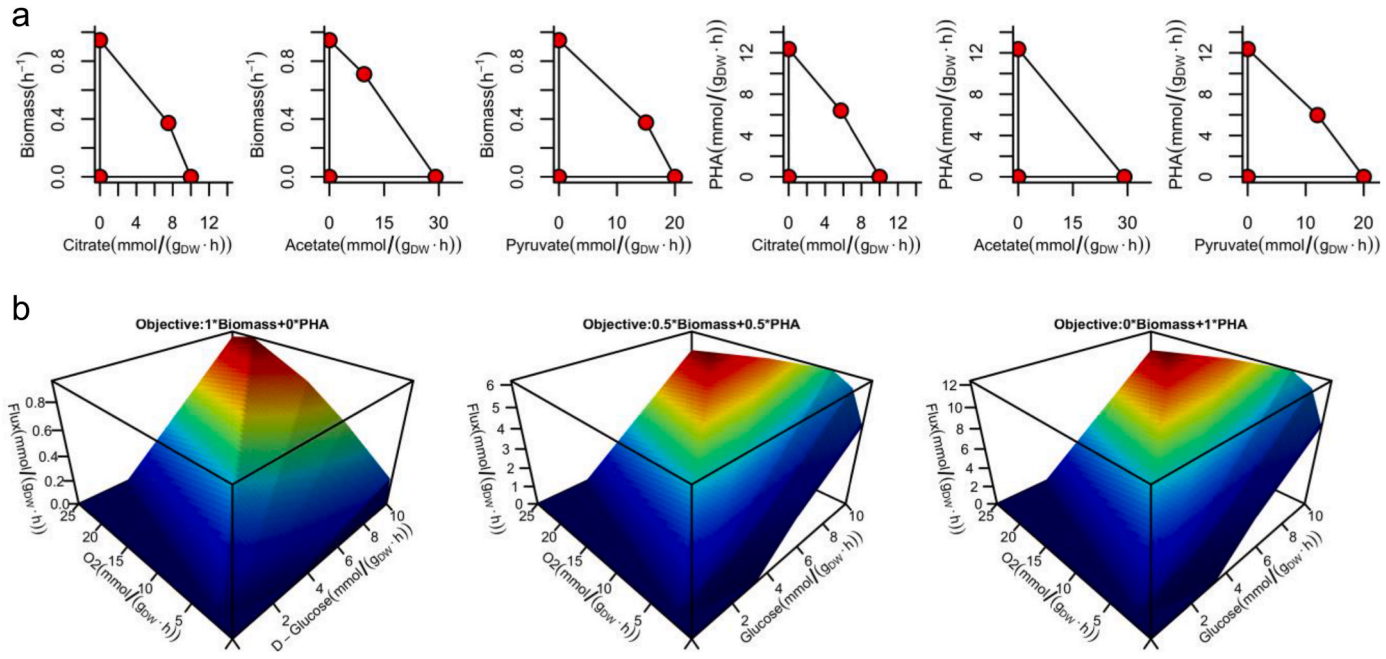


Fig. 5. Trade-offs between biomass/PHA production and other byproducts. (a) Vertex sampling analysis for biomass, PHA and byproducts. (d) Phase plane analysis using flux balance analysis (FBA).

3.7. Hyper cube shrink analysis yield a strate segmented feeding

During the fermentation process, the availability of dissolved oxygen in the water poses a significant limitation for cell growth, and improving oxygen supply can be challenging. In the early stages of fermentation, the bacteria have sufficient oxygen, resulting in a high biomass accumulation rate. However, as more biomass accumulates, the amount of oxygen available to each bacterium decreases, leading to a decrease in the biomass accumulation rate and an increase in PHA accumulation rate since carbon is still available for PHA synthesis. When the oxygen available to each bacterium becomes very limited, biomass accumulation stops, and only PHA production continues. Toward the end of fermentation, the PHA production rate decreases due to bacterial aging and decreased enzyme activity.

To optimize production, we suggested adopting different fermentation strategies in different periods. The early fermentation stage requires a strategy that favors biomass, while in the later stages, strategies that promote PHA accumulation should be employed to achieve a high PHA production rate. In the middle stage of fermentation, transitional strategies should be applied to promote the ideal conditions for bacterial growth and improve PHA yield.

We initially conducted simulations for single nutrient analysis to identify the optimal substrate composition strategy (Fig. 6a–c). Since FBA can only maximize biomass or PHA production at one time and cannot predict the causal relationships between inputs and outputs, we employed the previously developed HCSA in this analysis, which allows us to explore the flux distribution between multiple outputs, ensuring the accuracy of our predictions (Xie et al., 2018). In HCSA, we defined a pseudo flux V which defines a "shrink cube" when it increases:

$$l_i + C_i V \le v_i \le u_i - D_i(1 - C_i)V$$

v_i represents the flux of the reaction i that we pose the "shrink cube" on. When performing the nutrient analysis in this study, we will apply this "shrink cube" on the corresponding nutrient import reactions in the model. $C_i v_i$ represents the enzyme regulatory activities for reaction i. l_i (u_i) is the lower (upper) bound of the flux for reaction i, which are calculated by flux variability analysis (FVA). $D_i = u_i - l_i$. An inner point of the metabolic flux solution space can be found by solving:

maximize V s.t.

$$[S_{m\times n} \quad 0_{m\times 1}]\begin{bmatrix} v_{n\times 1} \\ V_{1\times 1} \end{bmatrix} = 0_{(n+1)\times 1}$$

$$\begin{bmatrix} -I_{z\times z} & C_{z\times 1} \\ I_{z\times z} & G_{z\times 1} \end{bmatrix}\begin{bmatrix} v_{z\times 1} \\ V_{1\times 1} \end{bmatrix} \le \begin{bmatrix} l_{z\times 1} \\ u_{z\times 1} \end{bmatrix}$$

$$v_i \in [\, l_i \quad u_i \,]$$

where S is the stoichiometric matrix for the metabolic model, I is an identity matrix with its diagonal elements all being 1 and the rest being 0, and $G_i = D_i(1 - C_i)$. m and n represent number of metabolites and reactions in the metabolic network respectively, and z represents number of reactions we apply the "shrink cube" on.

HCSA optimizes the consistency between enzymatic activity and flux and generate flux outputs for multiple targets at the same time. The nutrients we analyzed here include oxygen, nitrogen, and phosphorus. As shown in Fig. 6a–c, when intake rates of these nutrients, except oxygen, are within the range shown on the line charts, they all have inhibitory effects on PHA output but are beneficial to biomass

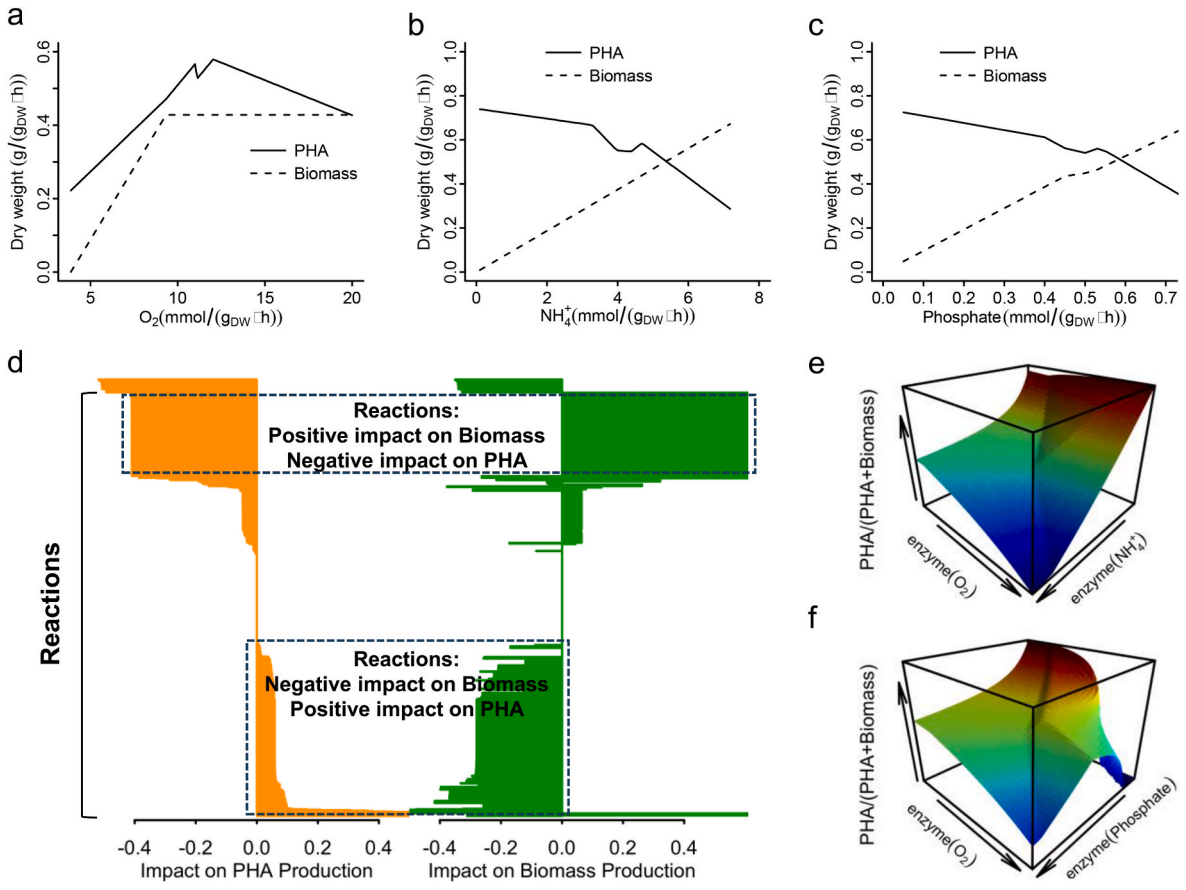


Fig. 6. Effects of nutrient uptake and other metabolic reactions on biomass and PHA production. (a–c) Single nutrient analysis using HCSA. The intake rate of glucose was fixed at 10 mmol/(g_{DW}.

accumulation.

The above results provide important guidance for optimizing fermentation strategies to improve PHA yield. Specifically, in the early stage of fermentation when biomass accumulation is the objective, it is essential to provide sufficient oxygen, nitrogen, and phosphorus to facilitate biomass synthesis. In contrast, in the late stage of fermentation, when PHA production is the primary objective, the addition of these nutrients should be limited to avoid inhibiting PHA production. This strategy can also prevent the early synthesis of PHA, which can occupy the cytosol and inhibit bacterial growth and metabolism. By inhibiting PHA production in the early stage of fermentation, enough biomass can be obtained for PHA production in the later stage, ultimately resulting in a larger PHA yield.

To gain a comprehensive understanding of the causal relationships between other reactions and biomass growth rate or PHA production, we performed HCSA on all metabolic reactions in the model. Fig. 6d shows that a large portion (305 of 941 reaction fluxes being evaluated, 32.4%) of reaction have positive impact on biomass but negative impact on PHA, that is, flux increases in these reactions will promote biomass accumulation but inhibit PHA production. Besides, there is another large portion (343 of 941 reaction fluxes being evaluated, 36.5%) of reactions in the model shown in Fig. 6d have negative impact on biomass but positive impact on PHA. This might indicate that PHA and biomass productions mostly compete with each other. We are also able to see that there are small groups of reactions having negative impacts on both biomass and PHA production, and there are also small group of reactions having very positive impacts on both PHA and biomass production. The detailed HCSA analysis results are shown in Supplementary Table 5.

The double-nutrients analysis was also conducted using HCSA (Fig. 6e and f). In this analysis, we focused on the effect of nutrient intake on biomass and PHA productions, and therefore, we set the fluxes of all other exchange reactions to zero, except for the biomass function, PHA exchange reaction, and exchange reactions for necessary nutrients. In this HCSA simulation, by changing the activities of hypothetical enzymes that control nutrient intake reactions in the metabolic model, we could simulate changes in nutrient intake rate and observe its effect on biomass and PHA yield. The results showed that decreasing the inputs of oxygen, nitrogen, and phosphorus can increase the PHA production (Fig. 6e and f).

The metabolic network predictions of *Halomonas bluephagenesis* TD01 were consistent with results from wet experiments. We previously reported wet experiment results of single nutrient analysis for nitrogen (Ye et al., 2018a), using urea or corn steep powder as the nitrogen source in the culture medium, where high nitrogen medium inhibited PHA synthesis but promoted biomass accumulation, but low nitrogen medium would increase PHA production, which agrees with our model predictions. The results are reproduced in Fig. 7a and b with different metrics from the original paper. Here, we performed wet experiments of single nutrient analysis for phosphorus, with dihydrogen phosphate being the phosphorus source in the culture medium (Fig. 7c). Our wet experimental results on phosphorus were also consistent with the model predictions, where high phosphorus concentrations in the culture medium inhibited PHA synthesis but promoted biomass accumulation. Conversely, low phosphorus concentrations were beneficial for PHA production.

We have previously reported three-stage feeding fermentation and

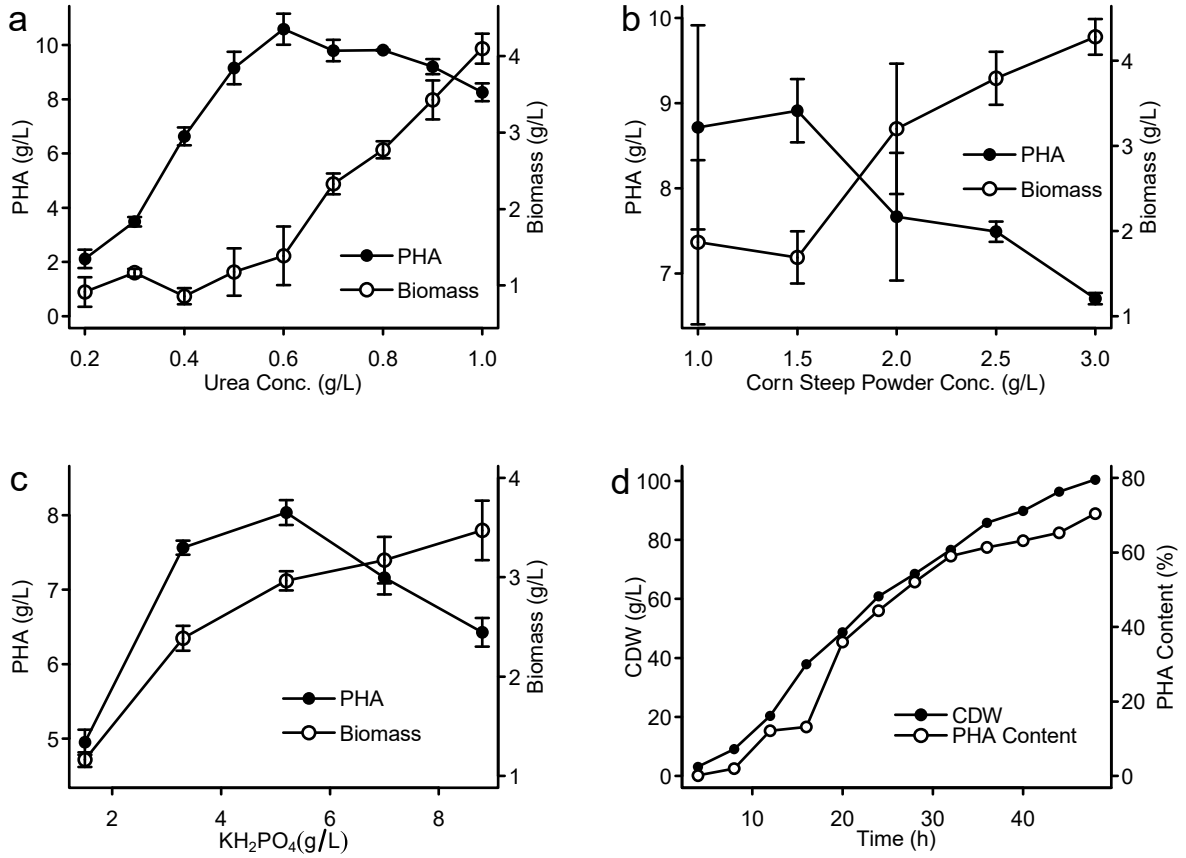


Fig. 7. Experimental results of single nutrient analysis and multiple-stage feeding fermentation. (a–c) Wet experiment results for single-nutrient analysis using urea, corn steep powder, and dihydrogen phosphate. Both urea and corn steep powder can provide nitrogen for the bacteria. Dihydrogen phosphate provides the phosphorus. (a) and (b) are reproduced with different metrics from Ye et al. (2018). Specifically, the data for the cell dry weight (CDW) and PHA content from the previous study was obtained to calculate the PHA yield (CDW x PHA content) and biomass yield (CDW – PHA yield) in (a) and (b). (d) Cell dry weight (CDW) and PHA content in four-stage feeding fermentation.

achieved a high PHA yield, 82.6 g/L CDW containing 79.86% PHA in the lab-scale 7.5 L bioreactor with *H. bluephagenesis* TD01, 90 g/L CDW containing 74.1% PHA in the scale-up 5000 L fermentor with *H. bluephagenesis* TD40 (a *H. bluephagenesis* TD01 derivative) (Ye et al., 2018a).

In light of our model simulations, we decided to conduct a four-stage feeding fermentation, modifying and optimizing our previous three-stage feeding fermentation process. The first stage of the previous three-stage feeding fermentation is between the 7th and 18th hour of the fermentation in the 7.5 L bioreactor, which is in the logarithmic phase of the cell growth (Ye et al., 2018a). Here, we split the first stage of the original three-stage feeding into two sections, splitting at the 15th hour of the fermentation, with the urea being distributed between first and second sections in a ratio of 7:3. Besides, corn steep powder was also added into these two sections and the concentration were 5 g/L and 3 g/L, respectively. This allowed more nitrogen in the very early stage of the fermentation to accelerate biomass accumulation and inhibit the accumulation of PHA. The four-stage feeding fermentation was performed in the lab-scale 7.5 L bioreactor, and results are shown in Fig. 7d.

Compared to previous three-stage feeding fermentation performed in the 7.5 L bioreactor, we obtained a lower PHA content at the early stage of fermentation. The PHA content was 13.2% at the 16th hour of the fermentation, whereas in the previous three-stage feeding fermentation, the PHA content was more than 30% at the same time point. However, the final total cell dry weight obtained by the four-stage feeding fermentation was 100.4 g/L, which was 25% higher than that obtained by the three-stage one, and we got a higher PHA yield (100.4 g/L CDW with 70.4% PHA content).

4. Discussion

By combining third-generation and NGS technologies, we obtained a complete and richly annotated genome sequence for *Halomonas bluephagenesis* TD01, significantly advancing our understanding of this bacterium. Comprehensive information on gene functions, regulatory elements, and potential metabolic pathways now paves the way for in-depth studies and novel applications.

Leveraging this detailed genomic data, we built the first genome-scale metabolic network for *H. bluephagenesis* TD01, providing a holistic view of its metabolic capabilities. While uncertainties remain due to the presence of hypothetical genes (17% potential impact), the network captures the core metabolic machinery and reveals promising new targets for metabolic engineering. Its comprehensive nature, as a "bird's-eye view" of TD01's metabolism, has already informed our strategies for improving the production rate of PHA and holds significant potential for future bio-based production processes.

With our metabolic model, we conducted vertex sampling analysis using the convex basis method to identify competitive metabolites within the network. This analysis revealed that the productions of biomass, PHA, citrate, acetate, and pyruvate are mutually compensable, guiding our efforts to enhance PHA yield through strategic inhibition or activation of specific pathways.

Recognizing the dynamic nature of *H. bluephagenesis* TD01, we employed our previously proposed HCSA algorithm to identify optimal feeding strategies tailored to different fermentation stages. Notably, HCSA's strength lies in its ability to introduce a flexible-regulatory constraint to multiple reactions, unlike traditional flux balance analysis (FBA) which focuses on a single objective function. Previously, we employed HCSA on a genome-scale level, achieving a narrow range of flux variations. In this work, we apply the HCSA constraint to specific reactions within a full-scale network, thereby circumventing its limitations and yielding a more significant redistributed flux distribution across reactions. This comprehensive approach led to the design of effective batch fermentation strategies, culminating in a remarkable cell dry weight of 100.4 g/L with a 70.4% PHA content, surpassing previously reported values (82.6 g/L CDW containing 79.86% PHA in the 7.5 L bioreactor, and 90 g/L CDW containing 74% PHA in a 5000 L fermentor) (Ye et al., 2018a).

Recently, the genome-scale metabolic model (HaloGEM) for *Halomonas campaniensis* was reconstructed (Deantas-Jahn et al., 2024). *Halomonas campaniensis*, similar to *Halomonas bluephagenesis*, is another *Halomonas* spp. that performed well in PHA accumulation. *Halomonas campaniensis* was reported to produce 55–70 g/L CDW containing over 70 % PHA (Yue et al., 2014), whereas *H. bluephagenesis* is able to accumulate 100.4 g/L biomass containing 70.4 % of PHA in our study. Specifically for the metabolic model, the *Halomonas campaniensis* model includes 888 genes, 1528 reactions, and 1274 metabolites (Deantas-Jahn et al., 2024), which is smaller than our model for *Halomonas bluephagenesis* TD01, which contains 1067 genes, 1860 reactions, and 1701 metabolites. The *Halomonas campaniensis* model shows that nitrogen source could limit aerobic growth of *Halomonas campaniensis* in minimal media and improving nitrogen source by using a combination of glutamate and arginine in the culture media could improve the biomass production, promoting a higher PHA yield at the end, which was predicted by their model and verified by wet experiments (Deantas-Jahn et al., 2024). We also discovered a similar feature in *H. bluephagenesis* TD01, where in the early stage of fermentation for *H. bluephagenesis* TD01, when the cell density in the culture media is low and oxygen is not a limiting factor for cell growth, a higher nitrogen source supply would promote rapid biomass accumulation, which in the end will promote a higher PHA yield, as shown by our four-stage feeding strategy. Taken together, genome-scale metabolic models in both strains will be beneficial for improvements in future PHA production.

H. bluephagenesis TD01's robust growth under non-sterile, high-salt conditions positions it as a prime candidate for industrial PHA production. This study presents the first genome-scale metabolic model for *H. bluephagenesis* TD01, unlocking its metabolic potential and paving the way for future optimization. Through network analysis and fermentation strategies guided by the model, we achieved significantly higher PHA yields compared to previous reports. As research on *H. bluephagenesis* TD01 progresses, refinement and expansion of this model will undoubtedly lead to further breakthroughs in microbial systems biology and bio-based PHA production.

CRediT authorship contribution statement

Luhui Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xinpei Sun:** Writing – review & editing, Writing – original draft, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jianwen Ye:** Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **QianQian Yuan:** Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xin Zhang:** Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Fei Sun:** Visualization, Validation, Methodology, Investigation, Conceptualization. **Yongpan An:** Methodology, Investigation, Formal analysis, Data curation. **Yutong Chen:** Methodology, Investigation, Formal analysis, Data curation. **Yuehui Qian:** Software, Methodology, Investigation, Formal analysis. **Daqian Yang:** Validation, Methodology, Investigation, Formal analysis, Data curation. **Qian Wang:** Methodology, Investigation, Formal analysis, Data curation. **Miaomiao Gao:** Visualization, Validation, Methodology, Investigation. **Tao Chen:** Project administration, Investigation, Funding acquisition, Conceptualization. **Hongwu Ma:** Supervision, Software, Funding acquisition, Conceptualization. **Guoqiang Chen:** Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Zhengwei Xie:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Funding acquisition, Data curation,

Conceptualization.

Data availability

Source data are provided with this paper in Supplementary Files. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request. The datasets analyzed during the current study are available in the NCBI repository, accession numbers: CP133631 for *Halomonas bluephagenesis* TD01, complete genome.

Declaration of competing interest

The authors declare that there is no conflict of interest.

Acknowledgements

This work was supported by the National key R&D program of China 2018YFA0900200, the National Natural Science Foundation of China 32170756.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mec.2024.e00251.

Data availability

Data will be made available on request.

References

Albe, K.R., Butler, M.H., Wright, B.E., 1990. Cellular concentrations of enzymes and their substrates. J. Theor. Biol. 143, 163–195.
Anderson, A.J., Dawes, E.A., 1990. Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. Microbiol. Rev. 54, 450–472.
Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M.W., Henderson, M.L., Riehl, W. J., Murphy-Olson, D., Chan, S.Y., Kamimura, R.T., Kumari, S., Drake, M.M., Brettin, T.S., Glass, E.M., Chivian, D., Gunter, D., Weston, D.J., Allen, B.H., Baumohl, J., Best, A.A., Bowen, B., Brenner, S.E., Bun, C.C., Chandonia, J.M., Chia, J. M., Colasanti, R., Conrad, N., Davis, J.J., Davison, B.H., DeJongh, M., Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J.N., Fang, G., Faria, J.P., Frybarger, P.M., Gerlach, W., Gerstein, M., Greiner, A., Gurtowski, J., Haun, H.L., He, F., Jain, R., Joachimiak, M.P., Keegan, K.P., Kondo, S., Kumar, V., Land, M.L., Meyer, F., Mills, M., Novichkov, P.S., Oh, T., Olsen, G.J., Olson, R., Parrello, B., Pasternak, S., Pearson, E., Poon, S.S., Price, G.A., Ramakrishnan, S., Ranjan, P., Ronald, P.C., Schatz, M.C., Seaver, S.M.D., Shukla, M., Sutormin, R.A., Syed, M.H., Thomason, J., Tintle, N.L., Wang, D., Xia, F., Yoo, H., Yoo, S., Yu, D., 2018. KBase: the United States department of energy systems biology knowledgebase. Nat. Biotechnol. 36, 566–569.
Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.
Ates, O., Oner, E.T., Arga, K.Y., 2011. Genome-scale reconstruction of metabolic network for a halophilic extremophile, Chromohalobacter salexigens DSM 3043. BMC Syst. Biol. 5, 12.
Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580.
Bordel, S., Agren, R., Nielsen, J., 2010. Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. PLoS Comput. Biol. 6, e1000859.
Cai, L., Tan, D., Aibaidula, G., Dong, X.R., Chen, J.C., Tian, W.D., Chen, G.Q., 2011. Comparative genomics study of polyhydroxyalkanoates (PHA) and ectoine relevant genes from Halomonas sp. TD01 revealed extensive horizontal gene transfer events and co-evolutionary relationships. Microb. Cell Factories 10, 88.
Chen, G.Q., 2009. A microbial polyhydroxyalkanoates (PHA) based bio- and materials industry. Chem. Soc. Rev. 38, 2434–2446.
Chen, G.Q., Jiang, X.R., 2018. Next generation industrial biotechnology based on extremophilic bacteria. Curr. Opin. Biotechnol. 50, 94–100.
Chen, X., Yin, J., Ye, J., Zhang, H., Che, X., Ma, Y., Li, M., Wu, L.P., Chen, G.Q., 2017. Engineering Halomonas bluephagenesis TD01 for non-sterile production of poly(3-hydroxybutyrate-co-4-hydroxybutyrate). Bioresour. Technol. 244, 534–541.
Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403.
Deantas-Jahn, C., Mendoza, S.N., Licona-Cassani, C., Orellana, C., Saa, P.A., 2024. Metabolic modeling of Halomonas campaniensis improves polyhydroxybutyrate production under nitrogen limitation. Appl. Microbiol. Biotechnol. 108, 310.
Delcher, A.L., Bratke, K.A., Powers, E.C., Salzberg, S.L., 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23, 673–679.
Diken, E., Ozer, T., Arikan, M., Emrence, Z., Oner, E.T., Ustek, D., Arga, K.Y., 2015. Genomic analysis reveals the biotechnological and industrial potential of levan producing halophilic extremophile, Halomonas smyrnensis AAD6T. SpringerPlus 4, 393. Doi, Y., 1990. Microbial polyesters. Vch.
Eilbeck, K., Moore, B., Holt, C., Yandell, M., 2009. Quantitative measures for the management and comparison of annotated genomes. BMC Bioinf. 10, 67.
Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.O., 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3, 121.
Fleming, R.M., Thiele, I., Nasheuer, H.P., 2009. Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to Escherichia coli. Biophys. Chem. 145, 47–56.
Fu, X.Z., Tan, D., Aibaidula, G., Wu, Q., Chen, J.C., Chen, G.Q., 2014. Development of Halomonas TD01 as a host for open production of chemicals. Metab. Eng. 23, 78–91.
Gonzalez, O., Gronau, S., Falb, M., Pfeiffer, F., Mendoza, E., Zimmer, R., Oesterhelt, D., 2008. Reconstruction, modeling & analysis of Halobacterium salinarum R-1 metabolism. Mol. Biosyst. 4, 148–159.
Gregory, S.G., Barlow, K.F., McLay, K.E., Kaul, R., Swarbreck, D., Dunham, A., Scott, C. E., Howe, K.L., Woodfine, K., Spencer, C.C., Jones, M.C., Gillson, C., Searle, S., Zhou, Y., Kokocinski, F., McDonald, L., Evans, R., Phillips, K., Atkinson, A., Cooper, R., Jones, C., Hall, R.E., Andrews, T.D., Lloyd, C., Ainscough, R., Almeida, J. P., Ambrose, K.D., Anderson, F., Andrew, R.W., Ashwell, R.I., Aubin, K., Babbage, A. K., Bagguley, C.L., Bailey, J., Beasley, H., Bethel, G., Bird, C.P., Bray-Allen, S., Brown, J.Y., Brown, A.J., Buckley, D., Burton, J., Bye, J., Carder, C., Chapman, J.C., Clark, S.Y., Clarke, G., Clee, C., Cobley, V., Collier, R.E., Corby, N., Coville, G.J., Davies, J., Deadman, R., Dunn, M., Earthrowl, M., Ellington, A.G., Errington, H., Frankish, A., Frankland, J., French, L., Garner, P., Garnett, J., Gay, L., Ghori, M.R., Gibson, R., Gilby, L.M., Gillett, W., Glithero, R.J., Grafham, D.V., Griffiths, C., Griffiths-Jones, S., Grocock, R., Hammond, S., Harrison, E.S., Hart, E., Haugen, E., Heath, P.D., Holmes, S., Holt, K., Howden, P.J., Hunt, A.R., Hunt, S.E., Hunter, G., Isherwood, J., James, R., Johnson, C., Johnson, D., Joy, A., Kay, M., Kershaw, J.K., Kibukawa, M., Kimberley, A.M., King, A., Knights, A.J., Lad, H., Laird, G., Lawlor, S., Leongamornlert, D.A., Lloyd, D.M., Loveland, J., Lovell, J., Lush, M.J., Lyne, R., Martin, S., Mashreghi-Mohammadi, M., Matthews, L., Matthews, N.S., McLaren, S., Milne, S., Mistry, S., Moore, M.J., Nickerson, T., O'Dell, C.N., Oliver, K., Palmeiri, A., Palmer, S.A., Parker, A., Patel, D., Pearce, A.V., Peck, A.I., Pelan, S., Phelps, K., Phillimore, B.J., Plumb, R., Rajan, J., Raymond, C., Rouse, G., Saenphimmachak, C., Sehra, H.K., Sheridan, E., Shownkeen, R., Sims, S., Skuce, C.D., Smith, M., Steward, C., Subramanian, S., Sycamore, N., Tracey, A., Tromans, A., Van Helmond, Z., Wall, M., Wallis, J.M., White, S., Whitehead, S.L., Wilkinson, J.E., Willey, D.L., Williams, H., Wilming, L., Wray, P.W., Wu, Z., Coulson, A., Vaudin, M., Sulston, J.E., Durbin, R., Hubbard, T., Wooster, R., Dunham, I., Carter, N.P., McVean, G., Ross, M.T., Harrow, J., Olson, M.V., Beck, S., Rogers, J., Bentley, D.R., Banerjee, R., Bryant, S.P., Burford, D.C., Burrill, W.D., Clegg, S.M., Dhami, P., Dovey, O., Faulkner, L.M., Gribble, S.M., Langford, C.F., Pandian, R.D., Porter, K.M., Prigmore, E., 2006. The DNA sequence and biological annotation of human chromosome 1. Nature 441, 315–321.
Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., Bateman, A., 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 33, D121–D124.
Grissa, I., Vergnaud, G., Pourcel, C., 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 35, W52–W57.
Guo, J., Zhang, H., Wang, C., Chang, J.W., Chen, L.L., 2016. Construction and analysis of a genome-scale metabolic network for Bacillus licheniformis WX-02. Res. Microbiol. 167, 282–289.
Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdottir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., Magnusdottir, S., Ng, C.Y., Preciat, G., Zagare, A., Chan, S.H.J., Aurich, M.K., Clancy, C.M., Modamio, J., Sauls, J.T., Noronha, A., Bordbar, A., Cousins, B., El Assal, D.C., Valcarcel, L.V., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Ben Guebila, M., Kostromins, A., Sompairac, N., Le, H.M., Ma, D., Sun, Y., Wang, L., Yurkovich, J.T., Oliveira, M.A.P., Vuong, P.T., El Assal, L.P., Kuperstein, I., Zinovyev, A., Hinton, H. S., Bryant, W.A., Aragon Artacho, F.J., Planes, F.J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M.A., Maranas, C.D., Lewis, N.E., Sauter, T., Palsson, B.O., Thiele, I., Fleming, R.M.T., 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat. Protoc. 14, 639–702.
Henry, C.S., Jankowski, M.D., Broadbelt, L.J., Hatzimanikatis, V., 2006. Genome-scale thermodynamic analysis of Escherichia coli metabolism. Biophys. J. 90, 1453–1461.
Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., Bork, P., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34, 2115–2122.
Istrail, S., Sutton, G.G., Florea, L., Halpern, A.L., Mobarry, C.M., Lippert, R., Walenz, B., Shatkay, H., Dew, I., Miller, J.R., Flanigan, M.J., Edwards, N.J., Bolanos, R., Fasulo, D., Halldorsson, B.V., Hannenhalli, S., Turner, R., Yooseph, S., Lu, F., Nusskern, D.R., Shue, B.C., Zheng, X.H., Zhong, F., Delcher, A.L., Huson, D.H., Kravitz, S.A., Mouchard, L., Reinert, K., Remington, K.A., Clark, A.G., Waterman, M. S., Eichler, E.E., Adams, M.D., Hunkapiller, M.W., Myers, E.W., Venter, J.C., 2004.

Whole-genome shotgun assembly and comparison of human genome assemblies. Proc. Natl. Acad. Sci. U.S.A. 101, 1916–1921.
Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240.
Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726–731.
King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., Lewis, N.E., 2016. BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 44, D515–D522.
Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T., Ussery, D.W., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100–3108.
Lan, L.H., Zhao, H., Chen, J.C., Chen, G.Q., 2016. Engineering Halomonas spp. as A low-cost production host for production of bio-surfactant protein PhaP. Biotechnol. J. 11, 1595–1604.
Lee, S.Y., 1996. Bacterial polyhydroxyalkanoates. Biotechnol. Bioeng. 49, 1–14.
Liu, M., Liu, H., Shi, M., Jiang, M., Li, L., Zheng, Y., 2021. Microbial production of ectoine and hydroxyectoine as high-value chemicals. Microb. Cell Factories 20, 76.
Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.
Magrane, M., UniProt, C., 2011. UniProt Knowledgebase: a hub of integrated protein data. Database 2011, bar009.
Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770.
Motamedian, E., Saeidi, M., Shojaosadati, S.A., 2016. Reconstruction of a charge balanced genome-scale metabolic model to study the energy-uncoupled growth of Zymomonas mobilis ZM1. Mol. Biosyst. 12, 1241–1249.
Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., 2009. Infernal 1.0: inference of RNA alignments. Bioinformatics 25, 1335–1337.
Noor, E., Haraldsdottir, H.S., Milo, R., Fleming, R.M., 2013. Consistent estimation of Gibbs energy using component contributions. PLoS Comput. Biol. 9, e1003098.
Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.O., 2011. A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. Mol. Syst. Biol. 7, 535.
Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R., 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. 42, D206–D214.
Piubeli, F., Salvador, M., Argandoña, M., Nieto, J.J., Bernal, V., Pastor, J.M., Cánovas, M., Vargas, C., 2018. Insights into metabolic osmoadaptation of the ectoines-producer bacterium Chromohalobacter salexigens through a high-quality genome scale metabolic model. Microb. Cell Factories 17, 2.
Poirier, Y., Nawrath, C., Somerville, C., 1995. Production of polyhydroxyalkanoates, a family of biodegradable plastics and elastomers, in bacteria and plants. Biotechnology 13, 142–150.
Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.
Steinbuchel, A., Fuchtenbusch, B., 1998. Bacterial and other biological systems for polyester production. Trends Biotechnol. 16, 419–427.
Tan, D., Xue, Y.S., Aibaidula, G., Chen, G.Q., 2011. Unsterile and continuous production of polyhydroxybutyrate by Halomonas TD01. Bioresour. Technol. 102, 8130–8136.
Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M.C., 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33, 2202–2204.
Xie, Z., Zhang, T., Ouyang, Q., 2018. Genome-scale fluxes predicted under the guidance of enzyme abundance using a novel hyper-cube shrink algorithm. Bioinformatics 34, 502–510.
Ye, J., Huang, W., Wang, D., Chen, F., Yin, J., Li, T., Zhang, H., Chen, G.Q., 2018a. Pilot scale-up of poly(3-hydroxybutyrate-co-4-hydroxybutyrate) production by Halomonas bluephagenesis via cell growth adapted optimization process. Biotechnol. J. 13, e1800074.
Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., Xu, H., Huang, X., Li, S., Zhou, A., Zhang, X., Bolund, L., Chen, Q., Wang, J., Yang, H., Fang, L., Shi, C., 2018b. Wego 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. Nucleic Acids Res. 46, W71–W75.
Yin, J., Chen, J.C., Wu, Q., Chen, G.Q., 2015. Halophiles, coming stars for industrial biotechnology. Biotechnol. Adv. 33, 1433–1442.
Yue, H., Ling, C., Yang, T., Chen, X., Chen, Y., Deng, H., Wu, Q., Chen, J., Chen, G.-Q., 2014. A seawater-based open and continuous process for polyhydroxyalkanoates production by recombinant Halomonas campaniensis LS21 grown in mixed substrates. Biotechnol. Biofuels 7, 108.
Zou, W., Ye, G., Zhang, J., Zhao, C., Zhao, X., Zhang, K., 2018. Genome-scale metabolic reconstruction and analysis for Clostridium kluyveri. Genome 61, 605–613.