Received: 8 December 2023 | Accepted: 2 February 2024

DOI: 10.1111/1751-7915.14425

MICROBIAL BIOTECHNOLOGY Applied Microbiology International

RESEARCH ARTICLE

Systematic elucidation of independently modulated genes in *Lactiplantibacillus plantarum* reveals a trade-off between secondary and primary metabolism

Sizhe Qiu[1,2] | **Yidi Huang**[3] | **Shishun Liang**[4] | **Hong Zeng**[2] | **Aidong Yang**[1]

[1]Department of Engineering Science, University of Oxford, Oxford, UK

[2]School of Food and Health, Beijing Technology and Business University, Beijing, China

[3]School of Computer Science and Engineering, Beihang University, Beijing, China

[4]Department of Life Science, Imperial College London, London, UK

Correspondence
Hong Zeng, School of Food and Health, Beijing Technology and Business University, Beijing 100048, China.
Email: zenghong@btbu.edu.cn

Aidong Yang, Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK.
Email: aidong.yang@eng.ox.ac.uk

Funding information
National Center of Technology Innovation for Dairy, Grant/Award Number: 2023-QNRC-2; National Natural Science Foundation of China, Grant/Award Number: 32302265

Abstract

Lactiplantibacillus plantarum is a probiotic bacterium widely used in food and health industries, but its gene regulatory information is limited in existing databases, which impedes the research of its physiology and its applications. To obtain a better understanding of the transcriptional regulatory network of *L. plantarum*, independent component analysis of its transcriptomes was used to derive 45 sets of independently modulated genes (iModulons). Those iModulons were annotated for associated transcription factors and functional pathways, and active iModulons in response to different growth conditions were identified and characterized in detail. Eventually, the analysis of iModulon activities reveals a trade-off between regulatory activities of secondary and primary metabolism in *L. plantarum*.

INTRODUCTION

The transcriptional regulatory network (TRN) of a bacterium consists of all regulatory interactions between its transcription factors (TFs) and genes (van Hijum et al., 2009). TFs, also referred to as sequence-specific DNA-binding factors, sense external signals and then bind to promoter regions of operons to regulate gene expression levels (Ishihama, 2012). To identify regulatory interactions between TFs and genes, the most commonly used experimental method is chromatin immunoprecipitation followed by sequencing (CHiP-seq) (Park, 2009). In CHiP-seq, antibodies are used to select TF proteins, and then DNA bound to TF proteins will be purified. DNA sequencing for the DNA-TF protein complex will determine the binding site on the genome. A group of genes with binding sites of the same TF are considered as a regulon. However, the drawbacks of CHiP-seq lie in its high cost, time-intensive nature and challenges in capturing the diverse growth conditions of bacteria (Kidder et al., 2011).

In recent years, many computational methods of in-silico reconstruction of TRN have been developed, such as coexpression network analysis (Lemoine et al., 2021) or supervised learning-based methods (e.g., GENIE3 (Huynh-Thu et al., 2010)). One of the

most popular methods to reconstruct TRN is using independent component analysis (ICA) to decompose the gene expression matrix, which consists of transcriptomic data of different samples, into sets of independently modulated genes, called iModulons (IMs) (Sastry et al., 2019). Apart from derived IMs, ICA can also quantify IM activities in different samples. Unlike CHiP-seq being a 'bottom-up' method, ICA follows a 'top-down' approach. ICA has been extensively applied to study and improve the understanding of many bacteria's TRNs. For example, ICA of *Vibrio natriegens* transcriptomes unveils the genetic basis of its natural competency (Shin et al., 2023). ICA has also been used to discover therapeutic strategies for *Streptococcus pyogenes* by identifying carbon sources that control the expression of haemolytic toxins (Hirose et al., 2023).

Lactiplantibacillus plantarum is a gram-positive lactic acid bacterium that can be found in diverse ecological niches (Seddik et al., 2017). It has been widely used in food and health industries. For instance, it is the major bacterium involved in the fermentation of mozzarella cheese (De Angelis et al., 2008); *L. plantarum*-derived exopolysaccharides (EPSs) have various probiotic effects (Silva et al., 2019) and anticancer properties (Arasu et al., 2016). Due to the importance of *L. plantarum* in different biological processes, such as dairy product fermentation, its gene expression regulation has received interest in several studies. For example, Jung and Lee identified differentially expressed genes when *L. plantarum* was in the acidic condition (Jung & Lee, 2020). Unlike most studies focusing on single regulatory genes, Wels et al. reconstructed the gene regulatory network of *L. plantarum* on the basis of correlations between gene expression levels and conserved regulatory motifs (Wels et al., 2011). Nonetheless, the regulon information of *L. plantarum* in RegPrecise (Novichkov et al., 2013) only recorded 47 regulons and 210 TF binding sites, in contrast to 624 and 943 TF binding sites recorded for *Bacillus subtilis* and *Escherichia coli*, respectively. The lack of gene regulatory information hinders the study of *L. plantarum*'s physiology and rational engineering of its cellular pathways.

Considering the value of *L. plantarum* in industry and research as well as the limited understanding of its TRN, this study managed to infer undiscovered regulatory interactions using ICA decomposition of the gene expression matrix and to further investigate how *L. plantarum* respond to different growth conditions (e.g., acid stress). Moreover, this study, through the analysis of IM activities, explored the growth strategy of *L. plantarum*, in terms of how it balances different biological processes (e.g., energy generation, carbohydrate metabolism, stress responses).

EXPERIMENTAL PROCEDURES

Data acquisition and preprocessing

The transcriptomic data used in the study were obtained from 4 independent studies that included various experimental conditions: response to pH decrease from 6.2 to 5.0 (Jung & Lee, 2020), treatment with N-3-oxododecanoyl homoserine lactone (a quorum sensing molecule) (Spangler et al., 2019), contrasting habitats (e.g., bee extract) (Filannino et al., 2018) and change of carbon sources (Özcan et al., 2021). The metadata of sample conditions can be found in Table S1. In the data from the selected 4 studies, genes were all annotated based on the genome assembly of *L. plantarum* WCFS1 (ASM20385v3) (Siezen et al., 2012). All transcriptomic sequencing reads were normalized as RPKM (Reads Per Kilobase Million). Then, all samples were merged as a compendium of transcriptomic data (100 samples and 3000 genes). Before independent component analysis (ICA) was undertaken, the merged dataset was first log-transformed and then centred by subtracting the expression levels of the reference condition (i.e., wt_pH6.2 in Table S1). The data quality was demonstrated by the higher Pearson correlation coefficients (PCCs) between replicates than PCCs between non-replicates (Rychel et al., 2021) (Figure 1A).

Determination of iModulons

ICA decomposition of the merged dataset (i.e., the expression matrix, 100 samples and 3000 genes) was conducted using scripts in precise-db (https://github.

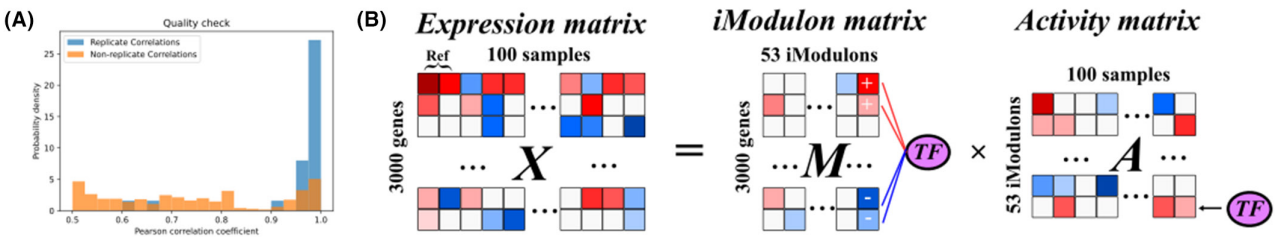

FIGURE 1 ICA decomposes the compendium of transcriptomic data to 45 nonempty iModulons. (A) Quality check of transcriptomic data with PCCs. Blue: replicate correlations; Yellow: non-replicate correlations. (B) Schematic illustration of ICA applied to the gene expression matrix.

com/SBRG/precise-db) (Rychel et al., 2020). The FastICA algorithm in Scikit-Learn (v0.20.3) (Pedregosa et al., 2012) was used to calculate independent components with 100 iterations with a tolerance of 10^{-7}, log(cosh(x)) as the contrast function, and parallel search algorithm. The OptICA method was used to determine the optimal number of independent components (McConn et al., 2021). The outputs of ICA were the iModulon matrix (M matrix, 3000 genes and 53 IMs) and Activity matrix (A matrix, 53 IMs and 100 samples) (Figure 1B). The M and A matrices can be found in https://github.com/SizheQiu/LPiModulons/tree/main/data/IMdata.

Gene weights in each column (for the corresponding IM) of the M matrix were used to determine each gene's IM membership. The threshold of gene weight absolute values for each IM was computed based on D'Agostino's K^2 test using the PyModulon package (https://github.com/SBRG/pymodulon) (Sastry et al., 2019). The default K^2-statistic cut-off of 550 was used. The genes with weight absolute values above the threshold were the member genes of the IM. Before annotation, Ims were labelled as IM-1 to 53.

Annotation of iModulons via regulon enrichment analysis

Regulons of *L. plantarum* WCFS1 were obtained from RegPrecise (Novichkov et al., 2013). Ims that overlap with regulons were annotated via regulon enrichment analysis. The set of genes in each IM was compared to each regulon using the two-sided Fisher's exact test (False Discovery Rate (FDR) $< 10^{-5}$) (Sastry et al., 2019). After regulon enrichments were computed for Ims, regulatory annotations were manually determined based on the Venn diagrams of Ims and regulons (see Figure S1). In addition to Ims associated with only one regulon (e.g., PyrR IM (IM-36)), there were two different annotation expressions for combined regulon enrichments: intersection (+) and union (/). If a specific combinatorial regulation (genes controlled by multiple regulators) was observed in the Venn diagram of the IM and enriched regulons, then the IM was annotated with regulators linked by '+' (e.g., MalR+MdxR IM (IM-47)). Otherwise, '/' was used (e.g., ArgR/MleR IM (IM-26)).

Annotation of iModulons via motif comparison

Ims that do not overlap with known regulons were annotated via motif discovery and motif comparison. If a coding gene's 200 bp upstream region does not overlap with another gene (Taboada et al., 2010) and BDGP Neural Network Promoter Prediction (Reese, 2001) predicted this region to be a possible promoter (probability score > 0.8), then this 200 bp upstream region was used to search for sequence motifs using MEME (Bailey, 1994). Motif comparison by TOMTOM (Gupta et al., 2007) then determined the most possible TF based on the similarity of found motifs and TF binding site motifs in databases (e.g., RegTransBase (Cipriano et al., 2013)). The p-value and E-value thresholds set in TOMTOM were 0.05 and 10. To further validate whether genes in the IM are regulated by the found TF, PCCs of the expression levels of the TF gene and IM genes were computed. If the gene had significant correlations (p-value < 0.05) with most genes in the IM, then the TF would be used to annotate the IM.

RESULTS

Regulatory and functional annotations of identified iModulons

The derived 53 Ims account for 85% explained variance of the gene expression matrix. In each IM, genes with absolute values of weights higher than the threshold are determined as IM member genes (see Methods, Section "Determination of iModulons"). The details of IM member genes can be found in https://github.com/SizheQiu/LPiModulons/tree/main/data/IMdata/ as IM_genes.csv. Among 53 Ims, 45 are nonempty and most Ims' sizes are within 20 (Figure 2A). Only 17% IM member genes overlap with genes in known regulons (Figure 2B), and hence, only 13 IMs could be annotated via regulon enrichment (Figure 2C). The details of regulatory annotations can be found in https://github.com/SizheQiu/LPiModulons/blob/main/data/IMdata/IM_annotation.csv.

For the 13 IMs annotated with enriched regulons, most of them have either high recall or high precision (cutoff = 0.6) (Figure 2D). Venn diagrams showing regulon enrichments in IMs are provided in Figure S1. High recall means that the overlap (of IM and regulon) has high coverage of the regulon, while high precision means that the overlap has high coverage of the IM. IMs with low recall and low precision are considered to be incompletely matched with regulons, but that does not necessarily mean the IM's regulatory annotation is inaccurate. For example, the remaining 3 genes in CopR IM that are not included by the current CopR regulon of *L. plantarum* WCFS1 are lp_3055(copA), lp_3057(copper-binding protein) and lp_3058(copper-binding protein), but they are included by the CopR regulon of other closely related lactic acid bacteria (e.g., *Lactococcus lactis subsp. lactis* Il1403) (Magnani et al., 2008). Therefore, the low recall and precision are sometimes resulted by the incompleteness of currently known regulons.

In addition to IMs associated with regulons, there are 11 IMs annotated via motif search and comparison

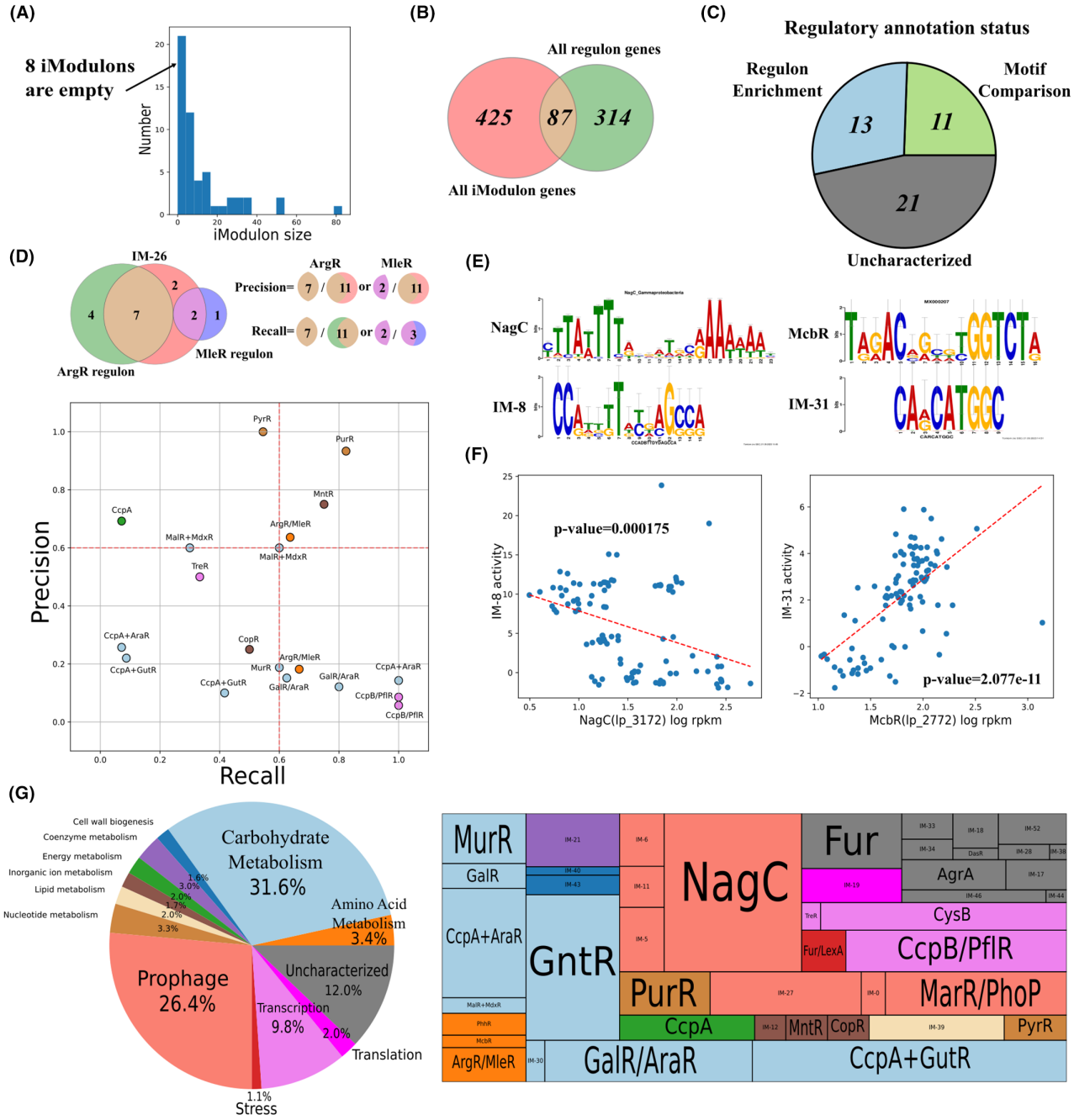

FIGURE 2 Regulatory and functional pathway annotations of IMs. (A) The histogram of IM sizes, 45 out of 53 IMs are nonempty. (B) The Venn diagram of all IM genes and regulon genes. 87 genes in IMs are contained in known regulons. (C) The pie chart of regulatory annotation status. Blue: regulon enrichment; Green: motif comparison; Grey: uncharacterized. (D) Recall and precision of IMs with matched regulons. (E) Motif comparison of IM-8 and IM-31. (F) The significant correlations between IM activities and gene expression levels of associated TFs identified via motif comparison for IM-8 and IM-31 (p-value < 0.05). (G) The pie chart and treemap of functional annotations of IMs, the size of each fraction is scaled with the IM size.

(Figures 2C and S2). Two representative examples are NagC IM (IM-8) and McbR IM (IM-31) (Figure 2E). Their regulatory annotations are validated by significant correlations between expression levels of TF genes and IM activities (Figure 2F). The remaining 21 IMs (Figure 2C) cannot be annotated via motif search and comparison either because the IM does not contain multiple possible promoter sequences for motif search (e.g., IM-19) or TOMTOM (Methods, Section "Annotation of iModulons via Motif Comparison") fails to find a TF binding site motif with a high similarity to the found motif (e.g., IM-6).

IMs were also annotated with enriched functional pathways (see SI, 3.1), and the details of functional annotations can be found in https://github.com/SizheQiu/LPiModulons/blob/main/data/IMdata/IM_annot

ation.csv. Apart from the uncharacterized group, 3 dominant functions of derived IMs are carbohydrate metabolism, prophage proteins and transcription (Figure 2G). Fur/LexA IM (IM-1) was functionally annotated as 'Stress', as LexA has already been found as a TF for stress response (Ravcheev et al., 2013). IM-19 was annotated as 'Translation', because genes in IM-19 were all ribosomal genes (e.g., rplV (lp_1039), large ribosomal subunit protein uL22). 12% (scaled with IM sizes) of IMs are uncharacterized in functional annotation due to the lack of enriched functional pathways.

Comparison between iModulons and regulons

The difference between IMs and regulons can provide undiscovered regulatory information. Regulon enrichments of some IMs show combinatorial regulations of multiple TFs, such as MalR+MdxR IM. Based on the genomic organization, 6 genes in the region between 151,222 and 158,185 bp belong to the same operon (Figure 3A). While mdxE (lp_0175), mdxG (lp_0177) and lp_0178 are already included by both MalR and MdxR regulons, MalR+MdxR IM also captures the combinatorial regulatory signals for malS (lp_0179) and msmX (lp_0180), which share the same promoter with genes in the overlap of MalR and MdxR regulons. All genes in MalR+MdxR IM are involved in maltose/maltodextrin metabolism, which is the biological process regulated by MalR and MdxR (Muscariello et al., 2011; Ravcheev et al., 2013).

IMs also have the ability to identify genes with strong regulatory interactions with TFs from known regulons. For example, the Pearson correlation coefficients (PCCs) between TF genes and genes in the overlap (of the IM and regulon) exhibit higher distributions compared to those of genes in the regulon for ArgR and CcpA (Figure 3B,C). Nevertheless, the overlap does not always show stronger regulatory interactions. For example, genes in PyrR IM do not have significantly higher PCCs with the PyrR gene than with the genes in PyrR regulon (Figure 3D).

Active iModulons in response to different growth conditions

In addition to the M matrix, the A matrix is another output of ICA decomposition, which reveals IM activities of *L. plantarum* under different growth conditions. In response to acid stress (in terms of pH decrease), 4 active IMs are observed: Fur/LexA IM, CopR IM, McbR IM and PyrR IM (Figure 4A). IM activities of all 4 active IMs identified increase with the decrease of pH (Figure 4B-E). The gene expression levels of Fur (lp_3247) and LexA (lp_2063) both decrease with the decrease of pH, though the trends over three pH values are not consistently decreasing (Figure 4F).

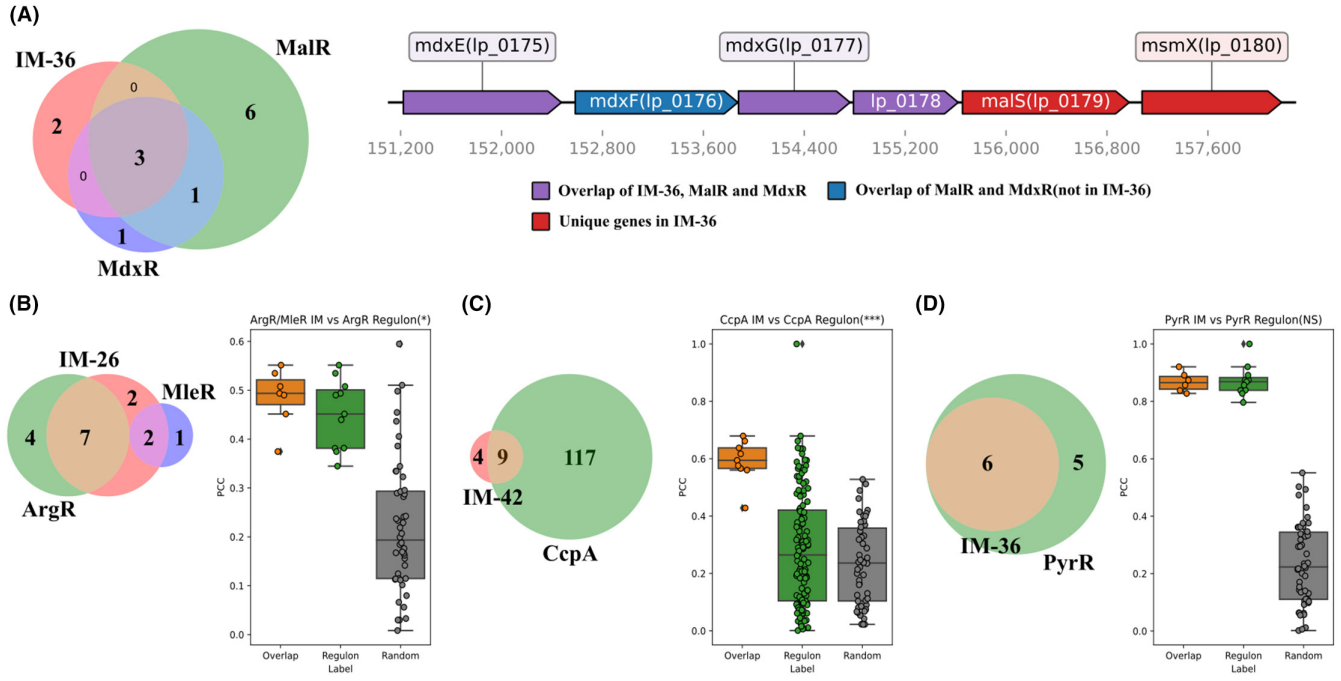

FIGURE 3 Comparison between IMs and regulons of *Lactiplantibacillus plantarum*. (A) Left: The Venn diagram of MalR+MdxR IM (IM-47) and Mdx, MalR regulons; Right: Genomic organization of genes in MalR+MdxR IM. (B–D) Comparison of PCCs of the gene expression levels of TF gene and genes in the overlap of IM and regulon (orange), regulon (green) and randomly sampled genes (grey) for (B) ArgR/MelR, (C) CcpA and (D) PyrR IMs. *p-value < 0.05; ***p-value < 0.001; NS, not significant.

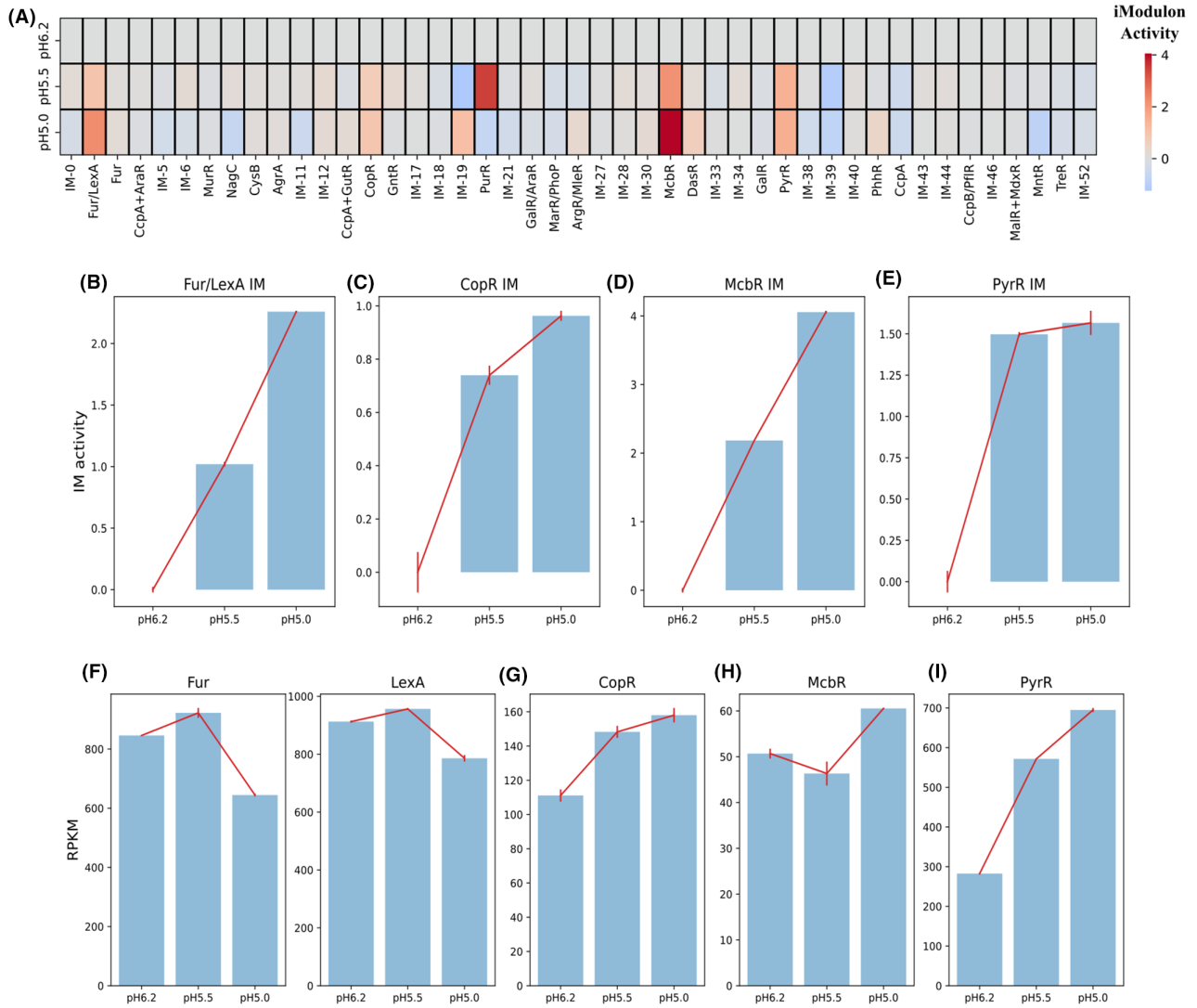

FIGURE 4 Identification of active IMs under the acidic condition. (A) The heatmap of IM activities at pH 6.2, 5.5 and 5.0. (B) IM activities of Fur/LexA IM at different pH values. (C) IM activities of CopR IM at different pH values. (D) IM activities of McbR IM at different pH values. (E) IM activities of PyrR IM at different pH values. (F) The expression levels of Fur (lp_3247) and LexA (lp_2063) at different pH values. (G) The expression levels of CopR (lp_3365) at different pH values. (H) The expression levels of McbR (lp_2772) at different pH values. (I) The expression levels of PyrR (lp_2704) at different pH values.

Genes in Fur/LexA IM are related to the biosynthesis of exopolysaccharide (EPS), an important secondary metabolite (Welman & Maddox, 2003), including lp_0302 (extracellular transglycosylase), lp_0304 (extracellular transglycosylase), lp_2809 (extracellular protein of unknown function), lp_2810 (glycosyl hydrolase, family 25), lp_2845 (extracellular transglycosylase, with LysM peptidoglycan binding domain), lp_3014 (extracellular transglycosylase, with LysM peptidoglycan binding domain) and lp_3050 (extracellular transglycosylase, membrane-bound). Oppositely, the gene expression levels of CopR (lp_3365), McbR (lp_2772) and PyrR (lp_2704) increase with the decrease of pH (Figure 4G–I). CopR, McbR and PyrR regulate copper homeostasis, amino acid metabolism and pyrimidine metabolism, respectively.

To further characterize acid-active IMs, regulatory networks are reconstructed as weighted correlation networks, and genomic organizations of genes in those IMs are further investigated. Fur/LexA IM, based on gene locations and the weighted correlation network, appear to contain two operons regulated by Fur and LexA separately: lp_0302 and lp_0304 regulated by Fur; lp_2809 and lp_2810 regulated by LexA (Figure 5A,B). The correlations between Fur and lp_0302, lp_0304 and lp_3014 are all negative, consistent with the previous finding that Fur is a repressor (Bagg & Neilands, 1987) (Figure 5A). The correlations between LexA and its regulated genes (i.e., lp_2809, lp_2810 and lp_3050) are positive, indicating that LexA functions as an activator to those genes (Figure 5A). For CopR, McbR and PyrR IMs, the correlations between TFs and regulated genes are all positive, suggesting that associated TFs

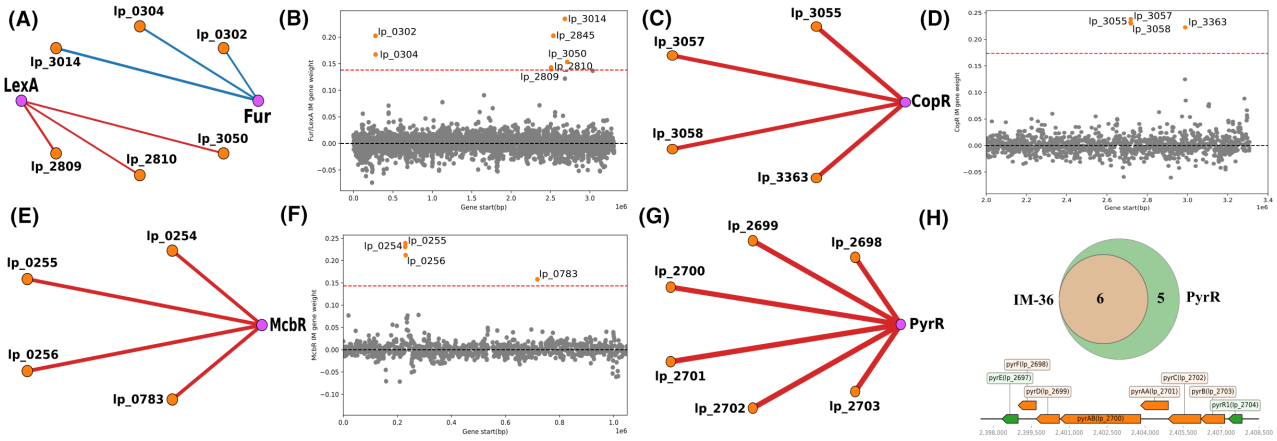

FIGURE 5 Characterization of genes in acidity-active IMs. (A) The weighted correlation network of Fur, LexA and genes in Fur/LexA IM (IM-1). (B) Gene weights and gene locations of Fur/LexA IM. (C) The weighted correlation network of CopR and genes in CopR IM (IM-15). (D) Gene weights and gene locations of CopR IM. (E) The weighted correlation network of McbR and genes in McbR IM (IM-31). (F) Gene weights and gene locations of McbR IM. (G) The weighted correlation network of PyrR and genes in PyrR IM (IM-36). (H) Genomic organization of genes in PyrR IM (IM-36). Orange: overlap of IM and regulon; Green: genes only in the regulon. Edge weights in weighted correlation networks are scaled to PCCs. Red: positive correlation; Blue: negative correlation; Orange node: the gene in the IM; Purple node: the TF gene.

all function as activators (Figure 5C,E,G). Unlike Fur/LexA IM, member genes of those three IMs are mainly in single operons (Figure 5D,F,H).

On the other hand, the change of carbon sources can result in transcriptional regulations of carbohydrate metabolism (Deutscher, 2008), where GntR IM (IM-16) was found to be the most active IM in this study (Figure 6A). Genes in GntR IM mainly encode for the utilization of different carbon sources (e.g., pts9C (lp_0576), uptake of mannose; panD (lp_0579), aspartate 1-decarboxylase) and the biosynthesis of capsular polysaccharide (CPS) in the cell wall (e.g., cps1F (lp_1182), CPS biosynthesis protein CpsC). The biosynthesis of CPS is a part of primary metabolism (cellular biomass formation), different from that of EPS, belonging to secondary metabolism (Whitfield et al., 2020). GntR IM is annotated via motif comparison (Figure S2) due to the lack of regulon information, and hence, it is hard to determine which TF in the GntR family regulate genes in this IM. Top 4 GntR TF genes with highest PCCs with activities of GntR IM are lp_2615, lp_2651, lp_3633 and lp_0563 (Figure 6B). The PCCs between TF genes and genes in GntR IM show that lp_2615 and lp_0563 have significant negative correlations with genes in GntR IM, while lp_2651 and lp_3633 have significant positive correlations with genes in GntR IM (Figure 6C), which are consistent with the PCCs (Figure 6B). Possibly, genes in GntR IM are regulated by multiple GntR family TFs.

The trade-off between primary and secondary metabolism revealed by iModulon activities

Member genes of IMs derived in this study encode connected reactions in one or several metabolic pathways, and those reactions were visualized as networks (see SI, 3.2) to investigate the links between IMs and cellular metabolism (Figure 7). For acid-active IMs identified in Section "Active iModulons in Response to Different Growth Conditions", genes in McbR IM and PyrR IM encode for the biosynthesis of L-cysteine and uridine monophosphate, respectively (Figure 7A,B). EPS biosynthetic reactions encoded by genes in Fur/LexA IM and copper homeostasis encoded by genes in CopR IM are currently not included by model iBT721.

Next, 4 representative IMs, namely ArgR/MleR IM, CcpA IM, GntR IM and GalR/AraR IM, functionally annotated for amino acid metabolism, energy metabolism and carbohydrate metabolism, are selected to reconstruct metabolic pathways encoded by their member genes (Figure 2G). ArgR/MleR IM member genes encode for the biosynthesis of N-Acetyl-L-glutamate 5-semialdehyde from L-glutamine (Figure 7C). CcpA IM, as an IM for energy metabolism, contains a part of glycolysis, the conversion of glycerol to dihydroxyacetone phosphate and phosphorylation of nucleosides (Figure 7D). GntR IM member genes mainly encode for CPS biosynthesis, from the activation of monosaccharides to the polymerization as explained in Section "Active iModulons in Response to Different Growth Conditions" (Figure 7E). Two important carbohydrate metabolic pathways, namely galactose metabolism and pentose phosphate pathway are contained by GalR/AraR IM (Figure 7F).

In contrast to Fur/LexA IM controlling secondary metabolism (EPS biosynthesis induced by acid stress) as shown in Section "Active iModulons in Response to Different Growth Conditions", ArgR/MleR, CcpA, GntR and GalR/AraR IMs (metabolic pathways visualized in Figure 7C–F) regulate primary metabolism.

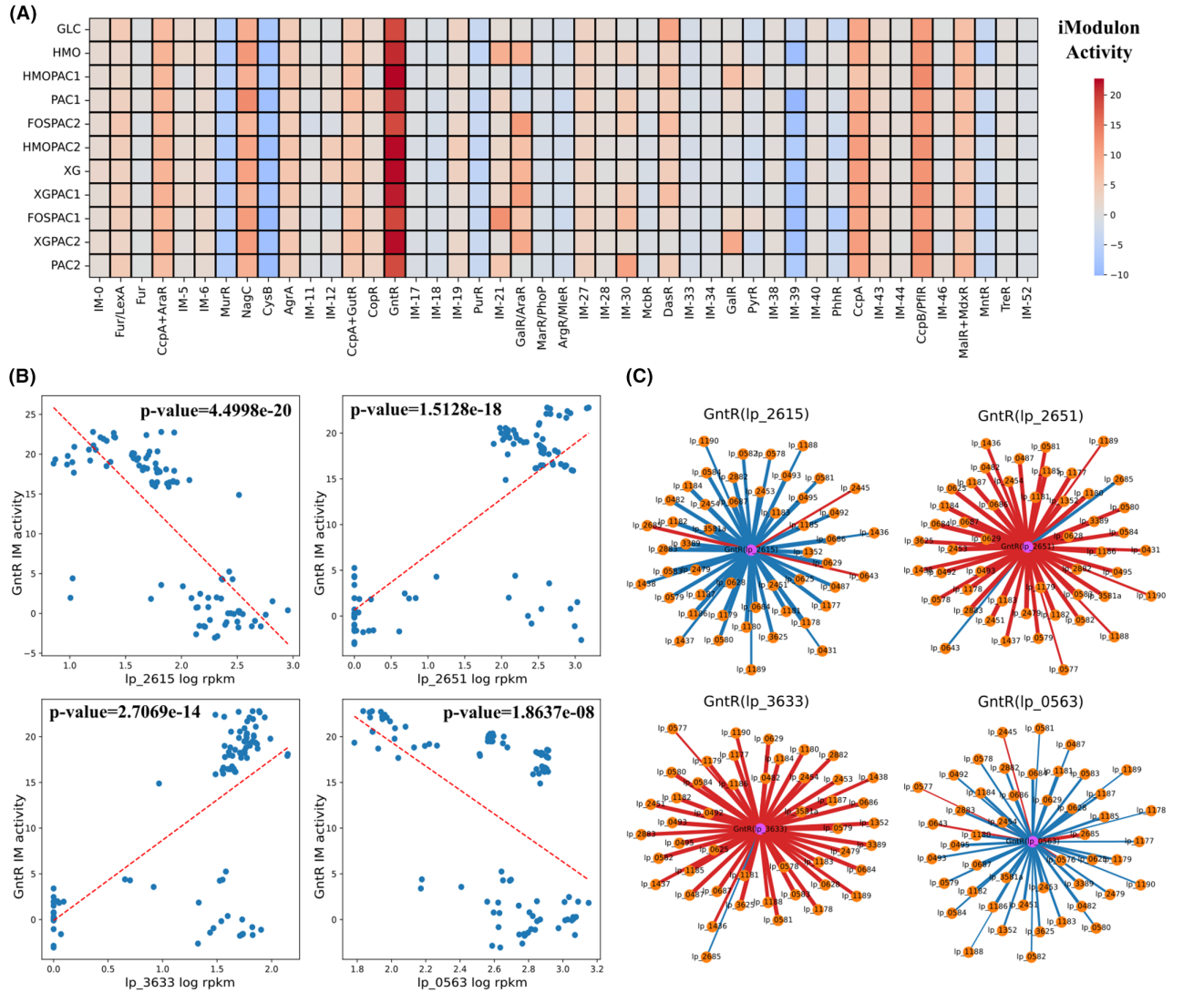

FIGURE 6 Identification of the most active IM in response to different carbon sources: GntR IM (IM-16). (A) The heatmap of IM activities with different carbon sources. FOS, fructooligosaccharides; GLC, glucose; HMO, human milk oligosaccharides; PAC1, proanthocyanidin fraction 1; PAC2, proanthocyanidin fraction 2; XG, xyloglucans. Detailed information can be found in Özcan et al. (2021). (B) The correlations between expression levels of 4 GntR family TF genes and GntR IM activities (p-value < 0.05). Red dashed line: linear fit. (C) The weighted correlation networks of 4 GntR family TF genes and genes in GntR IM (p-value < 0.05). Edge weights are scaled to PCCs. Red: positive correlation; Blue: negative correlation; Orange node: the gene in the IM; Purple node: the TF gene.

To investigate the relationship between regulatory activities of two branches of cellular metabolism, PCCs were computed for the activities of Fur/LexA IM and 4 IMs for primary metabolism (Figure 8A–D). Significant inverse correlations between the activity of Fur/LexA IM and activities of ArgR/MleR IM, CcpA IM, GntR IM and GalR/AraR IM can be observed, suggesting a trade-off between the regulatory activities of secondary and primary metabolisms. *Lactiplantibacillus plantarum* in acidic media (e.g., bee extract (pH 4.7), tomato juice (pH 3.5), see Table S1) have higher Fur/LexA IM activities and lower IM activities of the 4 IMs for primary metabolism than those in relatively neutral media (e.g., faecal extract (pH 5.9), see Table S1). Therefore, the balance between regulations of EPS biosynthesis and primary metabolism in *L. plantarum* appears to significantly depend on the acidity of extracellular environments.

To assess whether a trade-off relationship also exists between gene expression levels (in addition to regulatory activities) of secondary and primary metabolism, PCCs were computed between the total expression levels of genes in Fur/LexA IM (EPS biosynthetic genes) and (i) all glycolytic genes (central carbon catabolism) (Figure 8E) and (ii) genes in Translational IM (IM-19, ribosomal genes) (Figure 8F). An inverse correlation between gene expression levels of EPS biosynthetic genes and glycolytic genes is also observed (Figure 8E), though the correlation is not statistically significant. For EPS biosynthetic genes versus ribosomal genes, there is no inverse correlation between them (Figure 8F).

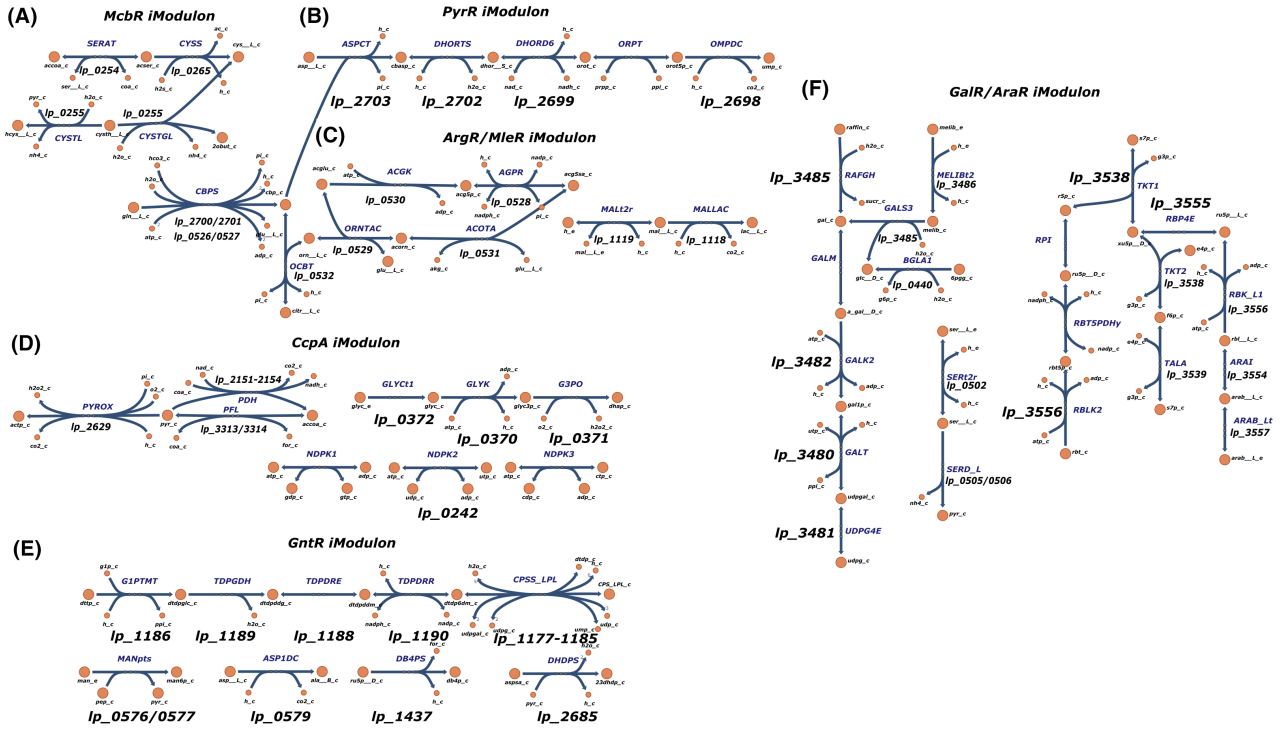

FIGURE 7 Metabolic pathways encoded by IM member genes. Reaction information (names, associated genes and IMs) can be found in Table S2. Reaction abbreviations are adopted from the BIGG database (http://bigg.ucsd.edu/) (King et al., 2016). (A) McbR IM. (B) PyrR IM. (C) ArgR/MleR IM. (D) CcpA IM. (E) GntR IM. (F) GalR/AraR IM.

DISCUSSION

ICA decomposition of *L. plantarum* transcriptomes allowed us to identify 45 nonempty IMs, 53.3% of which were annotated with associated TFs via either regulon enrichment analysis (13 IMs) or motif comparison (11 IMs). Annotated IMs revealed several regulatory interactions that have not been reported by known regulons of *L. plantarum*, for example, malS (lp_0179) and msmX (lp_0180) captured by MalR+MdxR IM (Section "Comparison between iModulons and Regulons"), which contributed to the reconstruction of a more complete TRN. Furthermore, the Activity matrix (A matrix) output by ICA decomposition showed the change of regulatory activities of TFs in response to different growth conditions (e.g., acid stress, carbon source switch), leading to the identification and characterization of relevant active IMs (Section "Active iModulons in Response to Different Growth Conditions"). Lastly, the analysis of relationships between IM activities unveiled a trade-off between secondary metabolism (acid stress-induced EPS biosynthesis) and primary metabolism in *L. plantarum* (Section "The Trade-off between Primary and Secondary Metabolism Revealed by iModulon Activities"), which might shed light on evolutionarily beneficial growth strategies.

Though IMs derived in this study provided regulatory information for the reconstruction of the TRN of *L. plantarum*, the performance of ICA decomposition was limited by the size of the expression matrix, compared to other ICA-based studies of bacterial transcriptomes (e.g., ICA of *Corynebacterium glutamicum* collected 263 samples from 29 independent projects (Zhao et al., 2023)). Compared to well-studied organisms such as *E. coli*, the amount of existing transcriptomic data of *L. plantarum* on NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) (Edgar et al., 2002) is much smaller. Also, due to the lack of operon annotation in *L. plantarum*'s genome, motif search for TF binding sites in this study used estimated promoter regions, which lowered the accuracy and might explain why some IMs were uncharacterized. It is also worth noting that the novel regulatory interactions shown by ICA are just 'predicted' instead of 'confirmed'. To obtain a more valid conclusion, CHiP-seq experiments are needed to confirm those findings in future studies.

With regard to the relationship between secondary and primary metabolism, theoretical models such as Grime's competitor-stress-ruderal triangle (Bruggeman et al., 2023; Grime, 1977), Synthetic Chemostat Model (Panikov, 2021) and regulatory proteome allocation model (Qiu et al., 2023) all adopted a resource allocation framework to capture the balance between two branches of cellular metabolism. Through the correlations between the activities of identified IMs, this study provided evidence to the theoretical models for secondary metabolism proposed in previous studies by showing the growth strategy of *L. plantarum* that adjusts regulatory

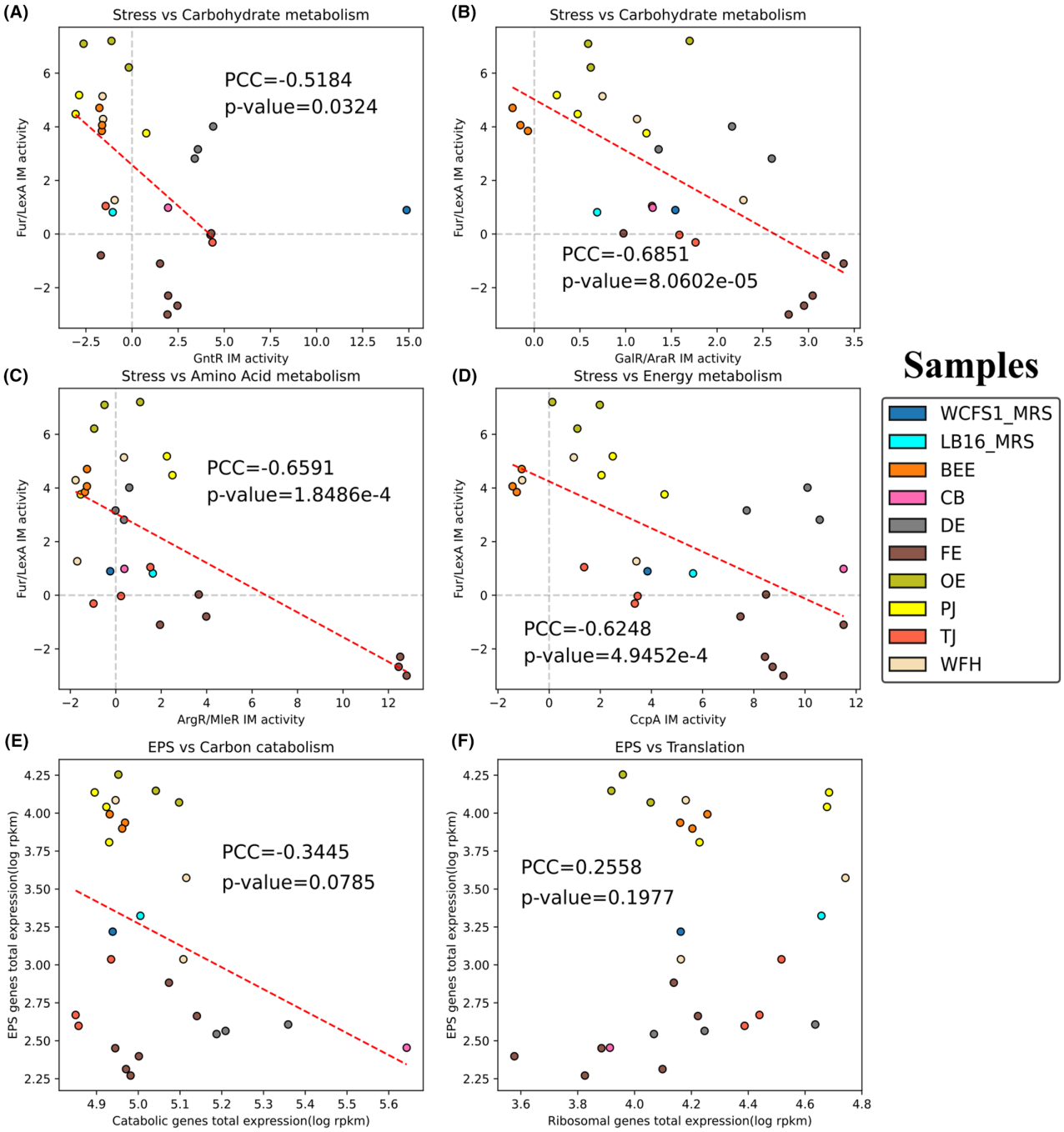

FIGURE 8 The relationships between secondary and primary metabolisms for *Lactiplantibacillus plantarum* cultivated in different growth conditions. (A) Fur/LexA IM activity versus GntR IM activity. (B) Fur/LexA IM activity versus GalR/AraR IM activity. (C) Fur/LexA IM activity versus ArgR/MleR IM activity. (D) Fur/LexA IM activity versus CcpA IM activity. (E) The total expression levels (log RPKM) of central catabolic genes and EPS biosynthetic genes (genes in Fur/LexA IM). (F) The total expression levels (log RPKM) of ribosomal genes (genes in Translation IM (IM-19)) and EPS biosynthetic genes. BEE, bee extract; CB, cheese broth; DE, *Drosophila* sp. extract; FE, faecal extract; LB16_MRS: *L. plantarum* LB16 in MRS broth; OE, olive extract; PJ, pineapple juice; TJ, tomato juice; WCFS1_MRS: *L. plantarum* WCFS1 in MRS broth; WFH, wheat flour hydrolysate.

activities for different metabolic pathways to react to external stress signals (Section "The Trade-off between Primary and Secondary Metabolism Revealed by iModulon Activities"). However, the curated data in this study could not support a significant trade-off relationship between gene expression levels of primary and secondary metabolism. More transcriptomic and proteomic profiling for *L. plantarum* under different growth conditions is needed to quantitatively study the balance between stress and cellular growth.

To conclude, this study provided the in-silico TRN reconstruction for *L. plantarum* in a top-down manner and unveiled its growth strategy to balance primary and secondary metabolism with IM activities, in spite

of the limitations discussed above. With the growing amount of gene expression data of *L. plantarum* as expected, the quality of IMs derived by ICA will be improved, thus enabling researchers to acquire a better understanding of the underlying rationale of its cellular activities.

AUTHOR CONTRIBUTIONS

Sizhe Qiu: Conceptualization (lead); data curation (lead); formal analysis (lead); methodology (lead); visualization (lead); writing – original draft (lead). **Yidi Huang:** Formal analysis (supporting); writing – original draft (supporting). **Shishun Liang:** Formal analysis (supporting); writing – original draft (supporting). **Hong Zeng:** Funding acquisition (lead); supervision (equal); writing – review and editing (equal). **Aidong Yang:** Supervision (equal); writing – review and editing (equal).

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support provided by National Natural Science Foundation of China (project No. 32302265; Beijing) and the National Center of Technology Innovation for Dairy (project No. 2023-QNRC-2).

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interests.

DATA AVAILABILITY STATEMENT

The code and data are openly available at https://github.com/SizheQiu/LPiModulons.

ORCID

Sizhe Qiu https://orcid.org/0000-0002-1936-1223

REFERENCES

Arasu, M.V., Al-Dhabi, N.A., Ilavenil, S., Choi, K.C. & Srigopalram, S. (2016) In vitro importance of probiotic *Lactobacillus plantarum* related to medical field. *Saudi Journal of Biological Sciences*, 23(1), S6–S10.

Bagg, A. & Neilands, J.B. (1987) Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. *Biochemistry*, 26(17), 5471–5477.

Bailey, T.L. (1994) *Fitting a mixture model by expectation maximization to discover motifs in bipolymers*. San Diego: Department of Computer Science and Engineering, University of California.

Bruggeman, F.J., Teusink, B. & Steuer, R. (2023) Trade-offs between the instantaneous growth rate and long-term fitness: consequences for microbial physiology and predictive computational models. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 45(10), e2300015.

Cipriano, M.J., Novichkov, P.N., Kazakov, A.E., Rodionov, D.A., Arkin, A.P., Gelfand, M.S. et al. (2013) RegTransBase—a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes. *BMC Genomics*, 14, 213.

De Angelis, M., de Candia, S., Calasso, M.P., Faccia, M., Guinee, T.P., Simonetti, M.C. et al. (2008) Selection and use of autochthonous multiple strain cultures for the manufacture of high-moisture traditional Mozzarella cheese. *International Journal of Food Microbiology*, 125(2), 123–132.

Deutscher, J. (2008) The mechanisms of carbon catabolite repression in bacteria. *Current Opinion in Microbiology*, 11(2), 87–93.

Edgar, R., Domrachev, M. & Lash, A.E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1), 207–210.

Filannino, P., De Angelis, M., Di Cagno, R., Gozzi, G., Riciputi, Y. & Gobbetti, M. (2018) How *Lactobacillus plantarum* shapes its transcriptome in response to contrasting habitats. *Environmental Microbiology*, 20(10), 3700–3716.

Grime, J.P. (1977) Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *The American Naturalist*, 111(982), 1169–1194.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. (2007) Quantifying similarity between motifs. *Genome Biology*, 8(2), R24.

Hirose, Y., Poudel, S., Sastry, A.V., Rychel, K., Lamoureux, C.R., Szubin, R. et al. (2023) Elucidation of independently modulated genes in *Streptococcus pyogenes* reveals carbon sources that control its expression of hemolytic toxins. *mSystems*, 8(3), e0024723.

Huynh-Thu, V.A., Irrthum, A., Wehenkel, L. & Geurts, P. (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE*, 5(9), e12776. Available from: https://doi.org/10.1371/journal.pone.0012776

Ishihama, A. (2012) Prokaryotic genome regulation: a revolutionary paradigm. *Proceedings of the Japan Academy, Series B, Physical and Biological Sciences*, 88(9), 485–508.

Jung, S. & Lee, J.-H. (2020) Characterization of transcriptional response of *Lactobacillus plantarum* under acidic conditions provides insight into bacterial adaptation in fermentative environments. *Scientific Reports*, 10(1), 19203.

Kidder, B.L., Hu, G. & Zhao, K. (2011) ChIP-Seq: technical considerations for obtaining high-quality data. *Nature Immunology*, 12(10), 918–922.

King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A. et al. (2016) BiGG models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44(D1), D515–D522.

Lemoine, G.G., Scott-Boyer, M.-P., Ambroise, B., Périn, O. & Droit, A. (2021) GWENA: gene co-expression networks analysis and extended modules characterization in a single bioconductor package. *BMC Bioinformatics*, 22(1), 267.

Magnani, D., Barré, O., Gerber, S.D. & Solioz, M. (2008) Characterization of the CopR regulon of *Lactococcus lactis* IL1403. *Journal of Bacteriology*, 190(2), 536–545.

McConn, J.L., Lamoureux, C.R., Poudel, S., Palsson, B.O. & Sastry, A.V. (2021) Optimal dimensionality selection for independent component analysis of transcriptomic data. *BMC Bioinformatics*, 22(1), 584.

Muscariello, L., Vastano, V., Siciliano, R.A., Sacco, M. & Marasco, R. (2011) Expression of the *Lactobacillus plantarum* malE gene is regulated by CcpA and a MalR-like protein. *Journal of Microbiology*, 49(6), 950–955.

Novichkov, P.S., Kazakov, A.E., Ravcheev, D.A., Leyn, S.A., Kovaleva, G.Y., Sutormin, R.A. et al. (2013) RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics*, 14(1), 1–12.

Özcan, E., Rozycki, M.R. & Sela, D.A. (2021) Cranberry proanthocyanidins and dietary oligosaccharides synergistically modulate *Lactobacillus plantarum* physiology. *Microorganisms*, 9(3), 656. Available from: https://doi.org/10.3390/microorganisms9030656

Panikov, N.S. (2021) Genome-scale reconstruction of microbial dynamic phenotype: successes and challenges. *Microorganisms*,

9(11), 2352. Available from: https://doi.org/10.3390/microorganisms9112352
Park, P.J. (2009) ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics*, 10(10), 669–680.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2012) Scikit-learn: machine learning in python. *arXiv [cs.LG]*, 85, 2825–2830. Available from: https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html
Qiu, S., Yang, A. & Zeng, H. (2023) Flux balance analysis-based metabolic modeling of microbial secondary metabolism: current status and outlook. *PLoS Computational Biology*, 19(8), e1011391.
Ravcheev, D.A., Best, A.A., Sernova, N.V., Kazanov, M.D., Novichkov, P.S. & Rodionov, D.A. (2013) Genomic reconstruction of transcriptional regulatory networks in lactic acid bacteria. *BMC Genomics*, 14, 94.
Reese, M.G. (2001) Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Computers & Chemistry*, 26(1), 51–56.
Rychel, K., Decker, K., Sastry, A.V., Phaneuf, P.V., Poudel, S. & Palsson, B.O. (2021) iModulonDB: a knowledgebase of microbial transcriptional regulation derived from machine learning. *Nucleic Acids Research*, 49(D1), D112–D120.
Rychel, K., Sastry, A.V. & Palsson, B.O. (2020) Machine learning uncovers independently regulated modules in the *Bacillus subtilis* transcriptome. *Nature Communications*, 11(1), 6338.
Sastry, A.V., Gao, Y., Szubin, R., Hefner, Y., Xu, S., Kim, D. et al. (2019) The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nature Communications*, 10(1), 5536.
Seddik, H.A., Bendali, F., Gancel, F., Fliss, I., Spano, G. & Drider, D. (2017) *Lactobacillus plantarum* and its probiotic and food potentialities. *Probiotics and Antimicrobial Proteins*, 9(2), 111–122.
Shin, J., Rychel, K. & Palsson, B.O. (2023) Systems biology of competency in *Vibrio natriegens* is revealed by applying novel data analytics to the transcriptome. *Cell Reports*, 42(6), 112619.
Siezen, R.J., Francke, C., Renckens, B., Boekhorst, J., Wels, M., Kleerebezem, M. et al. (2012) Complete resequencing and reannotation of the *Lactobacillus plantarum* WCFS1 genome. *Journal of Bacteriology*, 194(1), 195–196.
Silva, L.A., Lopes Neto, J.H.P. & Cardarelli, H.R. (2019) Exopolysaccharides produced by *Lactobacillus plantarum*: technological properties, biological activity, and potential application in the food industry. *Annals of Microbiology*, 69(4), 321–328.
Spangler, J.R., Dean, S.N., Leary, D.H. & Walper, S.A. (2019) Response of *Lactobacillus plantarum* WCFS1 to the gram-negative pathogen-associated quorum sensing molecule N-3-Oxododecanoyl homoserine lactone. *Frontiers in Microbiology*, 10, 715.
Taboada, B., Verde, C. & Merino, E. (2010) High accuracy operon prediction method based on STRING database scores. *Nucleic Acids Research*, 38(12), e130.
van Hijum, S.A.F.T., Medema, M.H. & Kuipers, O.P. (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiology and Molecular Biology Reviews*, 73(3), 481–509.
Welman, A.D. & Maddox, I.S. (2003) Exopolysaccharides from lactic acid bacteria: perspectives and challenges. *Trends in Biotechnology*, 21(6), 269–274.
Wels, M., Overmars, L., Francke, C., Kleerebezem, M. & Siezen, R.J. (2011) Reconstruction of the regulatory network of *Lactobacillus plantarum* WCFS1 on basis of correlated gene expression and conserved regulatory motifs. *Microbial Biotechnology*, 4(3), 333–344.
Whitfield, C., Wear, S.S. & Sande, C. (2020) Assembly of bacterial capsular polysaccharides and exopolysaccharides. *Annual Review of Microbiology*, 74, 521–543.
Zhao, J., Sun, X., Mao, Z., Zheng, Y., Geng, Z., Zhang, Y. et al. (2023) Independent component analysis of *Corynebacterium glutamicum* transcriptomes reveals its transcriptional regulatory network. *Microbiological Research*, 276, 127485.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Qiu, S., Huang, Y., Liang, S., Zeng, H. & Yang, A. (2024) Systematic elucidation of independently modulated genes in *Lactiplantibacillus plantarum* reveals a trade-off between secondary and primary metabolism. *Microbial Biotechnology*, 17, e14425. Available from: https://doi.org/10.1111/1751-7915.14425