Hybrid *de novo* Genome Assembly of *Erwinia* sp. E602 and Bioinformatic Analysis Characterized a New Plasmid-Borne *lac* Operon Under Positive Selection

Yu Xia[1,2,3*†], Zhi-Yuan Wei[4†], Rui He[1,2,3], Jia-Huan Li[4], Zhi-Xin Wang[4], Jun-Da Huo[4] and Jian-Huan Chen[4*]

[1] State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, China, [2] School of Food Science and Technology, Jiangnan University, Wuxi, China, [3] Collaborative Innovation Center of Food Safety and Quality Control in Jiangsu Province, Jiangnan University, Wuxi, China, [4] Laboratory of Genomic and Precision Medicine, Wuxi School of Medicine, Jiangnan University, Wuxi, China

Our previous study identified a new β-galactosidase in *Erwinia* sp. E602. To further understand the lactose metabolism in this strain, *de novo* genome assembly was conducted by using a strategy combining Illumina and PacBio sequencing technology. The whole genome of *Erwinia* sp. E602 includes a 4.8 Mb chromosome and a 326 kb large plasmid. A total of 4,739 genes, including 4,543 protein-coding genes, 25 rRNAs, 82 tRNAs and 7 other ncRNAs genes were annotated. The plasmid was the largest one characterized in genus *Erwinia* by far, and it contained a number of genes and pathways responsible for lactose metabolism and regulation. Moreover, a new plasmid-borne *lac* operon that lacked a typical β-galactoside transacetylase (*lacA*) gene was identified in the strain. Phylogenetic analysis showed that the genes *lacY* and *lacZ* in the operon were under positive selection, indicating the adaptation of lactose metabolism to the environment in *Erwinia* sp. E602. Our current study demonstrated that the hybrid *de novo* genome assembly using Illumina and PacBio sequencing technologies, as well as the metabolic pathway analysis, provided a useful strategy for better understanding of the evolution of undiscovered microbial species or strains.

Keywords: *Erwinia*, hybrid sequencing, genome assembly, *lac* operon, bioinformatic analysis

INTRODUCTION

Erwinia is a group of the straight rod-shaped, facultative anaerobic, gram-negative bacterium of the *Erwiniaceae* family of *Enterobacteriaceae*. Most of the *Erwinia* species identified by far are pathogens, saprophytes, or epiphytes of plants. It has been reported that some *Erwinia* species ferment lactose as a carbon source. Our previous study characterized a β-galactosidase with relatively high activity at low temperature in the *Erwinia* sp. strain E602 (Xia et al., 2018).

The *lac* operons typically involve three genes encoding the enzymes that enable bacteria to utilize lactose (Diaz-Hernandez and Santillan, 2010). The gene *lacZ* encodes the β-galactosidase, an enzyme that degrades lactose into monosaccharides glucose and galactose. Similarly, *lacY* encodes

a membrane-embedded transporter that helps bring lactose into cells. The gene _lacA_ encodes galactoside O-acetyltransferase that catalyzes the transfer of an acetyl group from acetyl-CoA to the 6-hydroxyl of galactopyranosides, with its exact physiological function remaining unclear. In addition, the gene _lacI_ encodes the _lac_ repressor, which is a protein that represses the transcription of _lac_ operon genes by binding to the promoter and preventing the operon from transcription when lactose is not available. In the presence of lactose, the _lac_ repressor is released from the operon to allow RNA transcription (Marbach and Bettenbrock, 2012). Notably, the _lac_ operon confers a competitive advantage for bacterial survival in the environments (Pinto et al., 2021). Therefore, the study of _lac_ operon in the newly identified strain _Erwinia_ sp. E602 may improve our understanding of the role of lactose metabolism in this genus.

High-throughput sequencing and _de novo_ assembly allow obtaining the total genetic information of microbes. It is widely used to dissect the genome sequence, gene composition, and evolutionary features of novel or unknown species. Notably, long-read sequencing technologies, such as PacBio (Rhoads and Au, 2015; Zhang et al., 2021), have dramatically speeded up the understanding of microorganisms. Moreover, the whole genome sequence of microorganisms provides an important basis for post-genomics research of microorganisms, such as subsequent gene mining and functional validation. In order to further investigate the regulation mechanisms of the β-galactosidase expression in the strain _Erwinia_ sp. E602, we sequenced its genome and conducted _de novo_ assembly, using the hybrid Illumina and PacBio sequencing (De Maio et al., 2019). From the assembly of the _Erwinia_ sp. E602 genome, we identified a large plasmid and a new plasmid-borne _lac_ operon containing a _lacZ_ gene encoding the β-galactosidase with low-temperature activity, as well as evidence supporting positive selection pressure of the _lac_ operon.

RESULTS

De novo Assembly and Annotation of _Erwinia_ sp. E602 Genome

As summarized in **Supplementary Tables 1–3**, the Illumina and PacBio reads were used to conduct _de novo_ genome assembly. After hybrid _de novo_ assembly using SPAdes, two circular scaffolds were obtained, including a chromosome sequence of 4,842,717 bp with a GC content of 58.91% (GenBank: CP046582.1) and a large plasmid sequence of 325,969 bp (GenBank: CP046581.1). A total of 4,739 genes, including 4,543 protein-coding, 25 rRNAs, 82 tRNAs, and 7 other ncRNAs genes were annotated as shown in **Table 1** and **Figures 1A,B**. Among these genes, 4,472 genes including 4,296 protein-coding, 25 rRNAs, 82 tRNAs, and 7 other ncRNAs genes were located on the chromosome, while 277 protein-coding genes were located on the large plasmid.

EggNOGv5.0 was then used to predict the functions of these genes. The KEGG pathway enrichment analysis was also performed (**Supplementary Material 1**). Pathways with a number of genes greater than 5 were shown, and finally, 122

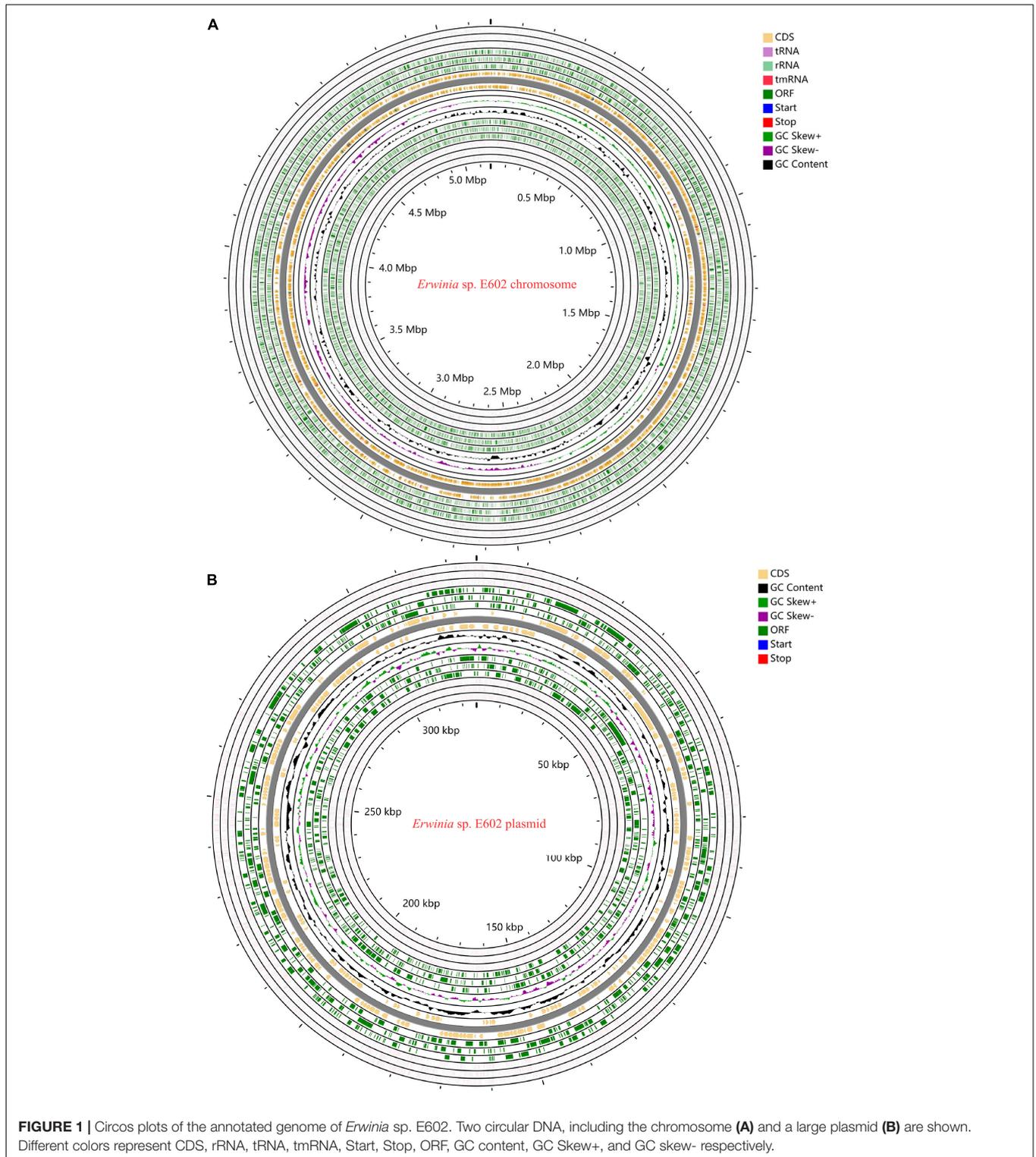**TABLE 1 |** Summary of _Erwinia_ sp. E602 genome annotation.

Type	Chromosome	Plasmid
RefSeq	NZ_CP046582.1	NZ_CP046581.1
INSDC	CP046582.1	CP046581.1
Size (Mb)	4.84	0.33
GC%	58.9	59.0
Protein	4,296	277
rRNA	25	–
tRNA	82	–
Other RNA	7	–
Gene	4,472	82
Pseudogene	62	5

enriched pathways were obtained. The most-enriched pathways (gene counts) included biosynthesis of secondary metabolites (357), microbial metabolism in diverse environments (260), ABC transporters (204), biosynthesis of cofactors (156), biosynthesis of amino acids (132), two-component systems (115), carbon metabolism (98), purine metabolism (71), quorum sensing (63), pyrimidine metabolism (51), ribosome (51), cysteine and methionine metabolism (50), bacterial secretion system (48), and amino sugar and nucleotide sugar metabolism (47). By dbCAN2 database annotation (Zhang et al., 2018), 593 carbohydrases were found on the chromosome and 51 on the plasmid (**Supplementary Material 1**). Through Resfams database annotation (Gibson et al., 2015), we found 211 resistance genes on the chromosome and 21 resistance genes on the plasmid, respectively (**Supplementary Material 1**).

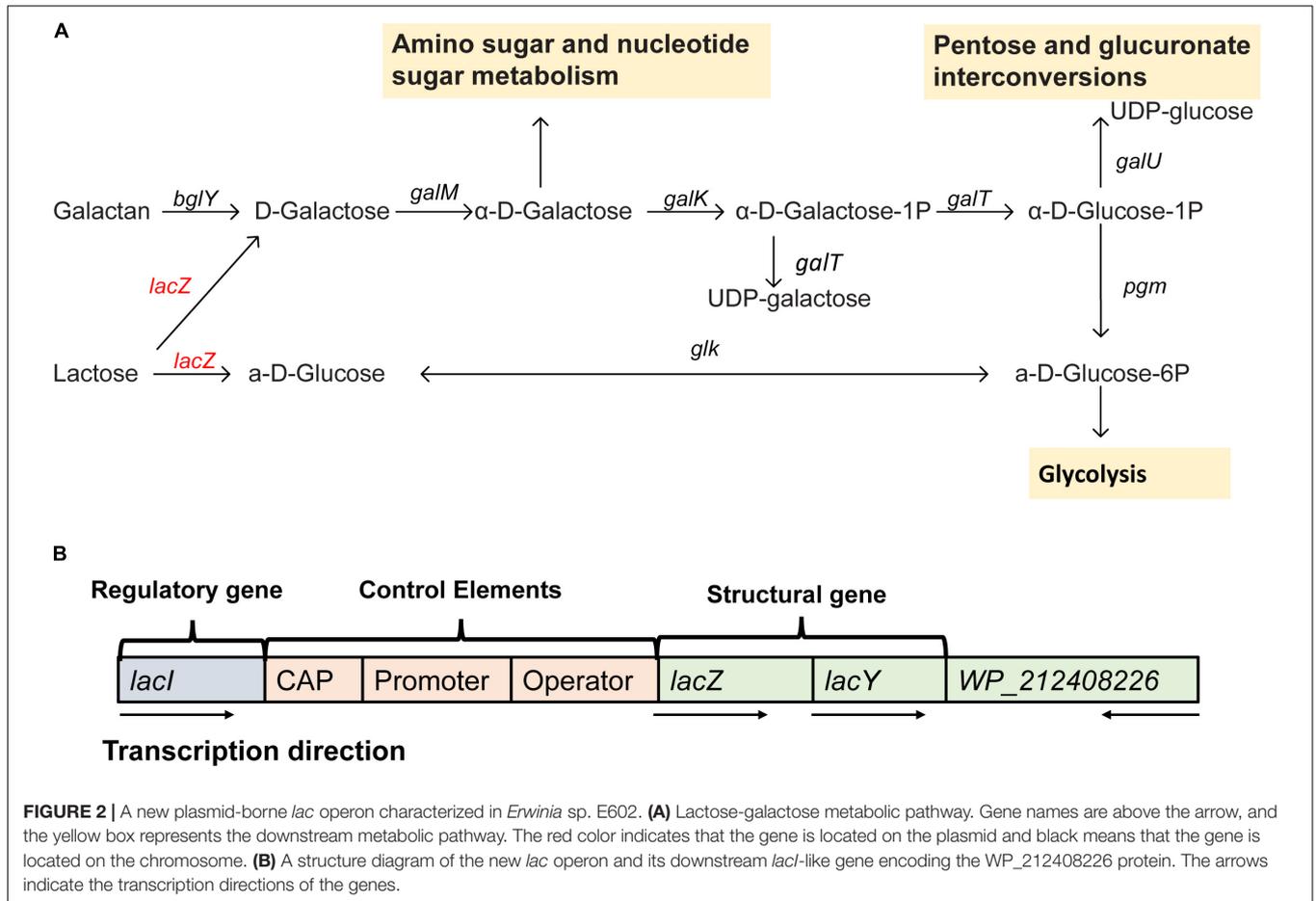### New Plasmid-Borne _lac_ Operon in _Erwinia_ sp. E602

From the gene annotation results of _de novo_ assembly, _Erwinia_ sp. E602 was found to contain a number of genes and pathways for lactose metabolism and regulation (**Figure 2A**). A list of lactose/galactose-related functional genes were shown in **Table 2**.

Our previous study characterized a new β-galactosidase that exhibited activity at low temperatures, indicating its potential unique biological function (Xia et al., 2018). Intriguingly, the _de novo_ assembly results showed that the exact _lacZ_ gene encoding the new β-galactosidase was not located on the chromosome of _Erwinia_ sp. E602. Instead, further analysis of the sequence context surrounding the _lacZ_ gene found that the gene was a component of a new _lac_ operon on the large plasmid. The _lac_ operons are known to be composed of regulatory genes, manipulated regions, and structural genes (Diaz-Hernandez and Santillan, 2010). The newly identified _lac_ operon in the current study contained _lacI, lacZ, lacY,_ as well as a regulatory element-CAP and a promoter, between the location of _lacI_ and _lacZ_ (**Figure 2B**). BLAST results of the plasmid-borne _lac_ operon showed that it shares low sequence similarity with other sequences in other _Erwinia_ strains. These results are included in **Supplementary Materials**. Moreover, this newly discovered _lac_ operon, compared with the classical ones, lacked a _lacA_ gene but had a reversed _lacI_-like gene encoding protein WP_212408226. Further comparison of the related genes revealed that although other _lacZ, lacI_ and

FIGURE 1 | Circos plots of the annotated genome of *Erwinia* sp. E602. Two circular DNA, including the chromosome **(A)** and a large plasmid **(B)** are shown. Different colors represent CDS, rRNA, tRNA, tmRNA, Start, Stop, ORF, GC content, GC Skew+, and GC skew- respectively.

lacY paralogs were also observed in the genome of *Erwinia* sp. E602, they did not form a canonical *lac* operon due to the lack of essential structure. Moreover, the *lacI* and *lacY* genes of the plasmid-borne *lac* operon share low similarity with their homologs in the genomes of *Erwinia* species (**Table 3**), suggesting

that this *lac* operon was distinct from canonical ones. In addition, the NCBI BLAST search using the nucleic acid database was performed to identify potentially similar operon-related genes, and the results showed that the genes in this operon were more similar to those in genus *Citrobacter* (**Figure 3**). Likewise, the

FIGURE 2 | A new plasmid-borne *lac* operon characterized in *Erwinia* sp. E602. **(A)** Lactose-galactose metabolic pathway. Gene names are above the arrow, and the yellow box represents the downstream metabolic pathway. The red color indicates that the gene is located on the plasmid and black means that the gene is located on the chromosome. **(B)** A structure diagram of the new *lac* operon and its downstream *lacI*-like gene encoding the WP_212408226 protein. The arrows indicate the transcription directions of the genes.

lacZ gene of the plasmid-borne *lac* operon shared a relatively low identity with its paralogs in the other *Erwinia* species (**Table 4**). Instead, its *lacZ* and *lacY* had high similarities with their homologs in *Citrobacter* species. The genes of *lacI* and *lacI*-like *protein* (WP_212408226) were compared using the NCBI nucleic acid database, and no sequences with similarities greater than 80% were found. In addition, no *lacA* was found either on the chromosome of Erwinia sp. E602, or on its large plasmid.

Selective Pressure in Genes in the Newly Identified *lac* Operon

To better understand whether genes in the plasmid-borne *lac* operon are subject to natural selection during adaptation to the environment, positive selection sites were identified using branch-site models (model A vs. model A null) and site models (M1a vs. M2a; M7 vs. M8) implemented in Phylogenetic Analysis by Maximum Likelihood (PAML) (20). Due to fewer sequences similar to genes *lacI* and the *lacI*-like protein (WP_212408226) in other bacteria strains, we chose the top 100 sequences ranked by identity with *lacY* or *lacZ* to be included in the two datasets separately for selection pressure analysis. The results of the branch-site model comparison showed that the gene *lacY* had a site at codon 187 under positive selection (**Table 5**). Surprisingly,

TABLE 2 | Lactose/galactose related functional genes annotated in *Erwinia* sp. E602 genome.

Function	Counts	Gene name
lac operon repressor	11	*lacI, AscG, galS, scrR, cytR, aglR, gntR, aglR, etc.*
Beta-galactosidase	3	*lacZ, lacZ3, bglY*
Lactose/galactose transport	9	*lacY, lacY-2, mglA-1, mglA, celC, ulaB, ptxB, bglE*
Galactose operon repressor	1	*galR*

there are multiple sites under strong positive selection in the gene *lacZ*, indicating the strong positive selection pressure in it (**Table 6**).

DISCUSSION

By using a strategy combining Illumina and PacBio sequencing technology, the *de novo* assembly of the whole genome of *Erwinia* sp. E602 was performed. Moreover, our study characterized the largest plasmid in the genus *Erwinia* by far, and reported a new plasmid-borne *lac* operon.

TABLE 3 | Sequence similarity of *lac* operon related genes.

Title	Query coverage	*E*-value	Per. ident
plasmid_238-lacI vs. genome_2601-lacI	41%	2.00 E-38	72.04%
plasmid_238-lacI vs. plasmid_241_wp	4%	7.00 E-06	89.29%
genome_2601–lacI vs. plasmid_241_wp	3%	0.004	100.00%
plasmid_240-lacY vs. genome_2604-bglY		No sig	
plasmid_240-lacY vs. genome_3788-lacY	94%	5.00 E-161	70.88%
genome_2604-bglY vs. genome_3788-lacY		No sig	
plasmid_239-lacZ vs. plasmid_250-lacZ3		No sig	
plasmid_239-lacZ vs. plasmid_240-lacY	0	0.039	100.00%

From the assembly, a large plasmid with a length of 325,969 bp and a total of 297 genes was characterized. Before our current study, 6 plasmids had been reported in the genus *Erwinia*, among which the largest plasmid was CP037949 derived from *Erwinia* sp. QL-Z3. That plasmid was 149,889 bp in length and encoded a total of 124 genes.[1] Therefore, the plasmid characterized in our current *de novo* genome assembly of *Erwinia* sp. E602 is the largest plasmid found in the genus *Erwinia* by far. In general, the genes in plasmids might provide microorganisms with potential genetic advantages. Yet the specific role of the large plasmid in *Erwinia* sp. E602 still remains unclear.

The *lac* operon is a hallmark gene of the regulatory circuit for bacteria to regulate metabolism according to nutrient conditions in the environment (Leonard et al., 2015; Malakar, 2015; Karkare et al., 2021; Pinto et al., 2021). Through genome function annotation, a large number of genes related to lactose metabolism were found in *Erwinia* sp. E602, a cold-adapted strain. In particular, a new plasmid-borne *lac* operon was characterized in this study. The *lac* operon shared low similarity with homologous genes in *Erwinia* species. It lacked a typical *lacA*. Instead, the position of β-galactoside transacetylase was replaced by a *lacI*-like gene encoding uncharacterized function. The gene *lacZ* in the newly identified *lac* operon encodes a β-galactosidase with the characteristics of low-temperature adaptation (Xia et al., 2018), which might help the organism survive in extreme environments. Similarly, identification and isolation of a 127-kb large plasmid in a rat *E. coli* isolate (EC93), containing the *cdiI* gene, would show higher toxic potency, thus helping it to exert a competitive advantage (Waneskog et al., 2021). Plasmids are mobile parts of bacterial genomes that carry genes that benefit the survival of the organism and confer selective advantages such as antibiotic resistance, thus helping bacterial communities to quickly adapt to the environment (Bogomazova et al., 2020). Similarly, the large plasmid we found in *Erwinia* sp. E602 also contains a group of I toxin-antitoxin modules, mainly including *vagC, vapC, relE, stbD, yefM,* and other related genes.

The plasmid-borne *lac* operon might enable *Erwinia* sp. E602 to utilize substrates from environmental resources for lactose metabolism, thereby gaining an adaptive advantage over competitors. Similar functions have been found in *E. coli* (Harwani, 2014). Nevertheless, our analysis showed that the *lac*

---

[1] https://www.ncbi.nlm.nih.gov/genome/browse/#!/plasmids/13524/

operon in *Erwinia* sp. E602 had low sequence similarity to that of *E. coli* and was evolutionarily distant from its *E. coli* counterparts. Likewise, other studies reported the *lac* operons in specific strains (Vaughan et al., 1998; Bidart et al., 2018), which also suggested that the *lac* operon could vary in different habitats. Furthermore, the *lacZ* and *lacY* genes on the *lac* operon in *Erwinia* sp. E602 were under positive selection, which might benefit the bacterium in terms of survival or competition (Bundalovic-Torma et al., 2020). Meanwhile, the strain *Erwinia* sp. E602 was isolated from the high latitudes of northeast China and the enzyme encoded by the plasmid-borne *lacZ* gene exhibited good activity at low temperatures. These findings might be in line with the possible adaptation of this *Erwinia* strain to the cold environment.

The plasmid-borne *lac* operon lacked the *lacA* gene in *Erwinia* sp. E602. It has been reported that the transacetylase encoded by the *lacA* gene is not an essential element for lactose catabolism (Lagesen et al., 2007). In contrast, a *lacI*-like gene was found in the opposite strand downstream of the *lac* operon in *Erwinia* sp. E602. It remains to be elucidated whether the *lacI*-like gene contributes to the regulation of the *lac* operon.

In the current study, by combining Illumina and PacBio sequencing technologies, a hybrid *de novo* assembly was performed to obtain a more accurate and complete map of the genome of *Erwinia* sp. E602, providing a reference basis for subsequent transcriptome studies. Furthermore, a new plasmid-borne *lac* operon was characterized with evidence supporting evolution and adaptation to the environment of the genus *Erwinia*. The specific role of the large plasmid on *Erwinia* sp. E602 needs further studies.
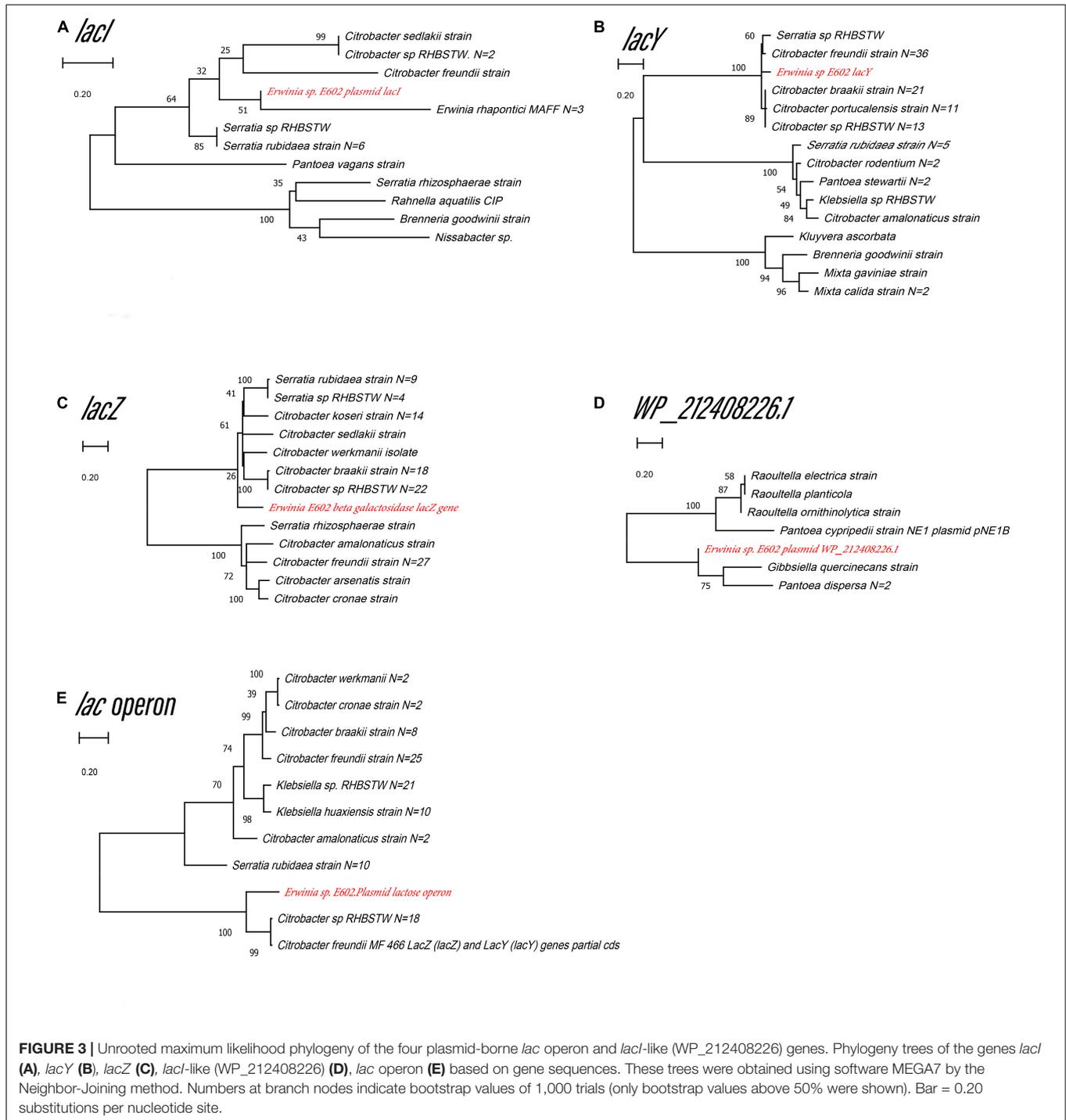
MATERIALS AND METHODS

Bacteria Culture and DNA Extraction

Bacterial culture and DNA extraction of *Erwinia* sp. E602 were performed as described in our previous report (Xia et al., 2018). The genomic DNA was used for subsequent Illumina and PacBio sequencing.

Illumina Sequencing and Data Processing

The library for short read sequencing was prepared from 1 μg *Erwinia* sp. E602 genomic DNA using NEBNext Ultra DNA Library Prep Kit (NEB, Ipswich, Massachusetts) according to the manufacturer's instructions, assessed using a Qubit® 2.0 Fluorometer (Invitrogen, Waltham, Massachusetts) and an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, California) for quality control (QC), and sequenced on the Hi-Seq 2500 platform (Illumina, San Diego, California) using a 100 bp paired-end configure. All adapter sequences and low-quality bases contained in the reads were filtered and removed using Trimmomatic (version 0.31), to keep only reads with sequencing read length more than 90 nucleotides (nt), and ensure the proportion of base quality greater than 20 (Q20) was not less than 90%.

FIGURE 3 | Unrooted maximum likelihood phylogeny of the four plasmid-borne *lac* operon and *lacI*-like (WP_212408226) genes. Phylogeny trees of the genes *lacI* **(A)**, *lacY* **(B)**, *lacZ* **(C)**, *lacI*-like (WP_212408226) **(D)**, *lac* operon **(E)** based on gene sequences. These trees were obtained using software MEGA7 by the Neighbor-Joining method. Numbers at branch nodes indicate bootstrap values of 1,000 trials (only bootstrap values above 50% were shown). Bar = 0.20 substitutions per nucleotide site.

PacBio Sequencing and Data Processing

The library for single-molecule long read sequencing was also prepared from 5 μg genomic DNA using SMRTbell DNA Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, California) according to the manufacturer's instructions, assessed using a Qubit® 2.0 Fluorometer and an Agilent 2100 Bioanalyzer for quality control (QC), and sequenced on the PacBio RS System. Raw data was filtered to obtain clean data with a cutoff of minimum polymerase read quality 0.75 and minimum polymerase read length 3,500 bp.

Hybrid Assembly of Illumina and PacBio Sequencing Data

The filtered Illumina reads and PacBio subreads were then used to conduct the hybrid assembly using SPAdes software (v3.5.0) (Bankevich et al., 2012). PacBio subreads were provided

TABLE 4 | Identity of plasmid-borne *LacZ* in *Erwinia* sp. E602 with its paralogs in other *Erwinia* species.

Erwinia strains	Chromosome/ Plasmid	Counts	GI (Identity%)
Erwinia amylovora CFBP1430	Chromosome	3	GI:490258078 (13%); GI:490258079 (13%); GI:490258942 (61%)
Erwinia billingiae Eb661	Chromosome	2	GI:502965951 (16%); GI:502966002 (65%)
Erwinia gerundensis E_g_EM595	Chromosome	2	GI:1055842378 (64%); GI:1055871078 (16%)
Erwinia gerundensis E_g_EM595	Plasmid pEM01	1	GI:1055874842 (59%); GI:1055874917 (11%)
Erwinia sp. J780	Chromosome	2	GI:1783198844 (15%); GI:1783198856 (12%); GI:1783199144 (64%)
Erwinia sp. QL-Z3	Chromosome	2	GI:1606668623 (16%); GI:1828892330 (65%)
Erwinia tasmaniensis Et1/99	Chromosome	1	GI:501410713 (62%)

TABLE 5 | Bayes Empirical Bayes analysis of positively selected sites identified in *lacY* of *Erwinia* sp. E602 with the branch-site model A.

Model	Codon	Amino acid	Posterior probability
Branch-site model	75	T	0.663
	147	G	0.564
	187	G	0.956*
	191	T	0.748
	207	S	0.801

*Posterior probability from BEB analysis > 95%; ratio for foreground branch (ω > 1).

to SPAdes using the –PacBio option for gap closure and repeat resolution.

Gene Prediction

The software Prokka (Seemann, 2014) (v 1.14.6) was used to predict genes in the assembly results, Carbohydrate enzymes were annotated using dbCAN2 (Zhang et al., 2018). Resistance genes were annotated using the Resfams database (Gibson et al., 2015). The prediction of rRNA was performed using rnammer1.2 (Lagesen et al., 2007). The prediction of tRNA was performed by the software tRNAscan-SE-1.23 (Chan and Lowe, 2019). The sRNA annotation was obtained by the software infernal-1.1rc4 (Nawrocki and Eddy, 2013). The CGView server (Stothard and Wishart, 2005) was used as an interactive comparison genomics tool to draw the circos plots.

Functional Annotation and Pathway Enrichment Analysis

EggNOG v5.0 (Huerta-Cepas et al., 2019) were used for the functionally annotation of genes. Pathway enrichment analysis was performed using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and KEGG Orthology (KO) databases.[2]

---

[2]https://www.genome.jp/kegg/mapper/search.html

TABLE 6 | Positively selected sites in *lacZ* of *Erwinia* sp. E602 identified with site model M2a using Bayes Empirical Bayes analysis.

Codon	Amino acid	Probability	Post mean
29	A	0.91	1.767
34	R	0.991**	1.873
37	I	0.901	1.755
38	T	0.960*	1.833
39	L	0.959*	1.83
114	T	0.936	1.801
115	G	1.000**	1.884
140	S	0.987*	1.867
146	V	0.9	1.755
151	A	0.974*	1.851
159	S	0.974*	1.85
386	S	0.995**	1.877
412	N	0.901	1.756
421	R	0.979*	1.857
425	P	0.998**	1.882
426	A	0.926	1.788
427	T	0.955*	1.825
429	R	0.995**	1.878

*Posterior probability of BEB analysis > 95%; **, posterior probability of BEB analysis > 99%. BEB, Bayes Empirical Bayes.

Prediction of *lac* Operon

The prediction of *lac* operon was conducted using Operon-mapper (Taboada et al., 2018).

Phylogenetic Analysis

Phylogenetic analysis was performed using software MEGA7 (Kumar et al., 2016) using the Neighbor-Joining method. The 1,000 replicates of bootstrap test were used to evaluate the clustering of taxa at branch nodes.

Select Pressure Assessment

In order to analyze the selection pressure of the related genes, PAML (v4.1) was used to identify sites under positive selection in the genes with site and branch-site models (Yang, 2007). For the branch-site models, *Erwinia*. sp. E602 was selected as the foreground branch of the branch-site model, and other bacteria were used as the background branch.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

YX and J-HC conceived the project, planned the experiments, and participated in the results discussions. Z-YW performed the bioinformatic analysis and improved the data analysis pipelines. RH participated in the DNA extraction from bacteria. J-HL,

Z-XW, and J-DH participated in the data analysis. All authors contributed to the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.783195/full#supplementary-material

REFERENCES

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bidart, G. N., Rodriguez-Diaz, J., Perez-Martinez, G., and Yebra, M. J. (2018). The lactose operon from *Lactobacillus casei* is involved in the transport and metabolism of the human milk oligosaccharide core-2 N-acetyllactosamine. *Sci. Rep.* 8:7152. doi: 10.1038/s41598-018-25660-w

Bogomazova, A. N., Gordeeva, V. D., Krylova, E. V., Soltynskaya, I. V., Davydova, E. E., Ivanova, O. E., et al. (2020). Mega-plasmid found worldwide confers multiple antimicrobial resistance in *Salmonella* Infantis of broiler origin in Russia. *Int. J. Food. Microbiol.* 319:108497.

Bundalovic-Torma, C., Whitfield, G. B., Marmont, L. S., Howell, P. L., and Parkinson, J. (2020). A systematic pipeline for classifying bacterial operons reveals the evolutionary landscape of biofilm machineries. *PLoS Comput. Biol.* 16:e1007721. doi: 10.1371/journal.pcbi.1007721

Chan, P. P., and Lowe, T. M. (2019). tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* 1962, 1–14.

De Maio, N., Shaw, L. P., Hubbard, A., George, S., Sanderson, N. D., Swann, J., et al. (2019). Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb. Genom.* 5:294. doi: 10.1099/mgen.0.000294

Diaz-Hernandez, O., and Santillan, M. (2010). Bistable behavior of the *lac* operon in *E. coli* when induced with a mixture of lactose and TMG. *Front. Physiol.* 1:22.

Gibson, M. K., Forsberg, K. J., and Dantas, G. (2015). Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* 9, 207–216. doi: 10.1038/ismej.2014.106

Harwani, D. (2014). Regulation of gene expression: Cryptic-glucoside (*bgl*) operon of *Escherichia coli* as a paradigm. *Braz. J. Microbiol.* 45, 1139–1144.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Karkare, K., Lai, H. Y., Azevedo, R. B. R., and Cooper, T. F. (2021). Historical contingency causes divergence in adaptive expression of the *lac* operon. *Mol. Biol. Evol.* 38, 2869–2879. doi: 10.1093/molbev/msab077

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874.

Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Leonard, S. R., Lacher, D. W., and Lampel, K. A. (2015). Acquisition of the *lac* operon by *Salmonella enterica. BMC Microbiol.* 15:173. doi: 10.1186/s12866-015-0511-8

Malakar, P. (2015). Pre-induced Lac operon effect on non specific sugars: pre-culture effect is dependent on strength of induction, exponential phase and substrate concentration. *Open Microbiol. J.* 9, 8–13. doi: 10.2174/1874285801509010008

Marbach, A., and Bettenbrock, K. (2012). *lac* operon induction in *Escherichia coli*: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J. Biotechnol.* 157, 82–88.

Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509

Pinto, C., Melo-Miranda, R., Gordo, I., and Sousa, A. (2021). The selective advantage of the *lac* operon for *Escherichia coli* is conditional on diet and microbiota composition. *Front. Microbiol.* 12:709259. doi: 10.3389/fmicb.2021.709259

Rhoads, A., and Au, K. F. (2015). PacBio sequencing and its applications. *Genom. Proteom. Bioinf.* 13, 278–289. doi: 10.1016/j.gpb.2015.08.002

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Stothard, P., and Wishart, D. (2005). Circular genome visualization and exploration using CGView. *Bioinformatics* 21, 537–539. doi: 10.1093/bioinformatics/bti054

Taboada, B., Estrada, K., Ciria, R., and Merino, E. (2018). Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics* 34, 4118–4120. doi: 10.1093/bioinformatics/bty496

Vaughan, E. E., Pridmore, R. D., and Mollet, B. (1998). Transcriptional regulation and evolution of lactose genes in the galactose-lactose operon of *Lactococcus lactis* NCDO2054. *J. Bacteriol.* 180, 4893–4902.

Waneskog, M., Halvorsen, T., Filek, K., Xu, F., Hammarlof, D. L., Hayes, C. S., et al. (2021). *Escherichia coli* EC93 deploys two plasmid-encoded class I contact-dependent growth inhibition systems for antagonistic bacterial interactions. *Microb. Genom.* 7:534. doi: 10.1099/mgen.0.000534

Xia, Y., He, L., Mao, J., Fang, P., Ma, X., and Wang, Z. (2018). Purification, characterization, and gene cloning of a new cold-adapted β-galactosidase from *Erwi*nia sp. E602 isolated in northeast China. *J. Dairy Sci.* 101, 6946–6954. doi: 10.3168/jds.2018-14605

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi: 10.1093/nar/gky418

Zhang, P., Jiang, D., Wang, Y., Yao, X., Luo, Y., and Yang, Z. (2021). Comparison of *de novo* assembly strategies for bacterial genomes. *Int. J. Mol. Sci.* 22:22147668. doi: 10.3390/ijms22147668