

# Whole genome sequencing and bioinformatics analysis of *Photobacterium kishitanii* FJ21

**Shuzheng Yin**

Hunan Agricultural University

**Zilong Yi**

Hunan Agricultural University

**Jia Liu**

Hunan Agricultural University

**Gang Liu**

Hunan Agricultural University

**Jun Fang** (✉ [fangjun1973@hunau.edu.cn](mailto:fangjun1973@hunau.edu.cn))

Hunan Agricultural University

---

## Research Article

**Keywords:** Photobacterium kishitanii, whole-genome sequencing, gene prediction and functional annotation, lux genes

**Posted Date:** January 21st, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1214200/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** *Photobacterium kishitanii* FJ21 is isolated and purified in the laboratory, and it is a kind of gram-negative bacteria that can emit blue-green fluorescence under normal conditions. In order to understand the *lux* operon that has luminescence activity, the whole-genome sequencing and bioinformatics analysis of *Photobacterium kishitanii* were carried out by using Nanopore sequencing technology. The sequencing data were predicted by the secondary structure and tertiary structure of the protein encoded by *lux* genes, genome assembly, GC content, gene prediction and functional annotation, phylogenetic tree, genome collinearity analysis and secondary metabolite synthesis gene cluster prediction.

**Result:** Sequencing results showed that there were *luxC*, *luxD*, *luxA*, *luxB*, *luxF*, *luxE* and *luxG* genes in *Photobacterium kishitanii* FJ21, and the protein encoded by *lux* genes had certain hydrophilicity. The genome of this strain contains a chromosome with a total length of 4853277bp and a GC content of 39.23%. There are 3141, 1769, 2472, 4070, 3514 and 1413 genes annotated in COG, KEGG, GO, Refseq, Pfam and TIGRFAMs databases, respectively. At the same time, three types of secondary metabolite gene clusters are predicted, namely RiPP-like, betalactone and arylpolyene. 870 genes were annotated in PHI database. The antibiotic resistance genes of the strain were annotated with CARD database.

**Conclusion:** This study reported the whole genome sequence *Photobacterium kishitanii* FJ21, and the related research results will provide a basis for further study of *lux* genes, photogenic activities and mining of other functional genes of this strain.

## Background

Luminescent bacteria are a type of Gram-negative bacteria that can emit blue-green fluorescence under normal conditions[1], and live in the marine environment mainly with the form of free organisms or parasites. The bioluminescence of luminescent bacteria is regulated by the enzyme-catalyzed reaction encoded by the *lux* genes. The *luxA* and *luxB* genes encode the  $\alpha$  and  $\beta$  subunits of luciferase, respectively. *luxC*, *luxD* and *luxE* constitute the fatty acid reductase complex, responsible for the synthesis of long Aldehyde substrate, *luxG* encodes flavin reductase[2]. Conserved genes *luxC*, *luxD*, *luxA*, *luxB*, *luxE*, and *luxG* exist in all luminescent bacteria that have been discovered[3], in addition, there are other genes such as *luxI*, *luxR* and *luxF*[4]. Although luminescent bacteria have the same *lux* genes, these bacteria show great differences in characteristics such as growth behavior, luminescence intensity, or bioluminescence regulation[5]. The luminescence process of luminescent bacteria is oxidize FMNH<sub>2</sub> and RCHO to FMN and RCOOH under the catalysis of intracellular specific luciferase and the participation of molecular oxygen, and at the same time release blue-green light[6]. The luminescence reaction is as follows:



Luminescent bacteria can not only act as biosensors[7], but also produce antibacterial compounds[8], lipase[9], asparagine[10] and esterase[11]. Due to the luminescent bacteria method has the advantages of high sensitivity, simple processing, rapid response, and real-time monitoring, it has been widely used in the monitoring of water toxicity and environmental pollutants[12], and the acute and chronic toxicity tests of

heavy metal mixtures[13]. At present, the luminescent bacteria commonly used in water quality and environmental monitoring are *Photobacterium phosphoreum*, *Vibrio fischeri* and *Vibrio Qinghaiensis*[14]. The bright luminescence is usually used in the national standard GB/T15441-1995 for the determination of acute toxicity of water quality. The basic principle of luminescent bacteria for acute toxicity detection is that the luminescence process is easily affected. As long as the respiration or physiological process of bacteria is disturbed, the luminescence intensity of the bacteria will change[15].

Recently, with the development of high-throughput sequencing technology, many microorganisms have completed genome sequencing. Whole-genome sequencing is an important foundation of microbial molecular mechanism research and development. To further understand the structure of the protein encoded by the *lux* genes of *Photobacterium kishitanii* and some functional genes, the whole genome sequencing and bioinformatics analysis of *Photobacterium kishitanii* FJ21 were carried out by Nanopore sequencing technology in this study. The related research results will provide a basis for the in-depth study of the *lux* genes, luminescent activities and other functional genes of *Photobacterium kishitanii* FJ21.

## Results

### Genome Sequencing and Assembly

The sequencing data are shown in Table 1 and the assembly results are shown in Table 2. The genome size was 4853277 bp, the number of coding genes was 4131, and the N50 was 3252201 bp. ATGC content accounted for 30.49%, 30.29%, 19.72%, and 19.50% of the total base, respectively, and the GC content was 39.23%. The sequencing depth distribution is shown in Figure 1, and the genome circle is shown in Figure 2. The genome sequence of *Photobacterium kishitanii* FJ21 has been submitted to the GenBank database with accession number SRX10356131.

Table 1  
Sequencing data statistics.

Name	Base(bp)	Longest Reads(bp)	mean length(bp)	mean quality	Number Reads	Reads N50(bp)
Raw Reads	2200768252	172754	22420.21	8.74	98160	32173
FilteredReads	1823049947	172754	23391.03	9.58	77938	32731

Table 2  
Assembly result statistics.

Total bases	Contig number	Contig N50	Longest Contig	Shortest Contig
4853277	2	3252201	3252201	1601076

### Genome Structure and Function Annotation

The genome contains 4131 CDSs, 4027962 bp in length, and a CRISPR sequence (a cluster of regularly spaced short palindromic repeats, often found in many bacteria and archaea). Gene islands are not predicted on the genome. The results of genome structure prediction are shown in Table 3. There are 1,769, 3,141, 2472,

4070, 3514, and 1413 genes were annotated respectively related to KEGG pathway, COG category, GO, Refseq, Pfam and TIGRFAMs databases.

Table 3  
Genome structure prediction results.

Type	Number	Length (bp)	% genome
tRNA	226	18001	0.37
rRNA(16s)	22	33990	0.70
rRNA(23s)	22	63668	1.31
rRNA(5s)	24	2784	0.06
CDS	4131	4027962	82.99
CRISPR	1	807	0.02

## COG function classification

In the COG category(Figure 3), there are Energy production and conversion (232 genes), Amino acid transport and metabolism (328 genes), Carbohydrate Transport and metabolism (210 genes), Translation, ribosomal structure and biogenesis (265 genes), Transcription (266 genes), Cell wall/membrane/envelope biogenesis (267 genes) and Inorganic ion transport and metabolism (194 genes).

The corresponding protein sequence was compared with the COG database to complete the annotation classification of homologous genes, and the coding genes including information storage and processing, cell biology process and signal transduction, basic metabolism, and unknown functions were obtained[34]. As shown in Figure 3, a total of 3141 proteins obtained COG functional annotations, accounting for 76.03% of the total number of predicted genes, including Energy production and conversion, Amino acid transport and metabolism, Carbohydrate Transport and metabolism, Translation/ribosomal structure and biogenesis, Transcription, Cell wall/membrane/envelope biogenesis, Inorganic ion transport and metabolism, and the number of genes was 232, 328, 210, 265, 266, 267 and 194, respectively.

## GO function classification

As shown in Figure 4, a total of 2472 genes were annotated for GO function, accounting for 59.84% of the total number of predicted genes. GO function mainly divides them into molecular function, biological process and cellular component[16]. In molecular function, there are many genes annotated by molecular transducer activity, antioxidant activity and transporter activity. In the biological process, there are many genes annotated by metabolic process, positive regulation of the biological process, negative regulation of the biological process. In cell components, there are many genes annotated by cell, cell part and membrane part. Therefore, GO functional annotation is more convenient for us to understand the biological significance behind genes.

## KEGG pathway analysis

The 2945 genes in the KEGG pathway were enriched in 208 metabolic pathways (Figure 5), and the number of effectively annotated genes was 1769, accounting for 42.82% of the total predicted genes. There are five categories, namely Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular processes and Organismal Systems. The most annotated genes in metabolism are carbohydrate metabolism, energy metabolism, and amino acid metabolism. The main pathways are oxidative phosphorylation pathway (ko00190) (40 genes), arginine and proline pathway (ko00330) (16 genes), glycolysis/ gluconeogenesis pathway (ko00010) (31 genes), citric acid cycle (TCA cycle) (ko00020) (19 genes). The least genes were annotated in organismal systems.

## Analysis of interaction genes between pathogen and host

870 genes were annotated in the PHI database (Figure 6), of which 326 genes (55.72%) resulted in reduced virulence. There were 43 increased virulence genes, 217 unaffected pathogenicity genes, 82 loss of pathogenicity genes, 103 effector genes, and resistance to chemical and sensitivity to chemical genes were the least. In this annotation, most of the genes belonged to the reduced virulence genes and unaffected pathogenicity genes. Effector genes are associated with pathogenicity, but increased virulence genes are the key genes.

## Annotation of Resistance Genes in the CARD Database

The information was shown in Table 4. Including the classification of ARO, the Identities, the classification of antibiotics, the resistance mechanism, and the classification of the AMR gene family. The highest identities can reach 100%. The resistance mechanisms are antibiotic efflux and antibiotic target Alteration.

Table 4  
Antibiotic resistance gene annotation.

ARO	%Identities	Drug Class	Resistance Mechanism	AMR Gene Family
rsmA	89.29	fluoroquinolone antibiotic, diaminopyrimidine antibiotic, phenicol antibiotic	antibiotic efflux	resistance-nodulation-cell division (RND)
vanWB	100	glycopeptide antibiotic	antibiotic target Alteration	antibiotic efflux pump vanW, glycopeptide
CRP	94.29	macrolide antibiotic, fluoroquinolone antibiotic; penam	antibiotic efflux	resistance gene cluster resistance-nodulation-cell division (RND) antibiotic efflux pump
Haemophilus influenzae PBP3 conferring resistance to beta-lactam antibiotics	46.85	cephalosporin; cephamycin; penam	antibiotic target alteration	Penicillin-binding protein mutations conferring resistance to beta-lactam antibiotics

# Analysis of carbohydrate-related enzymes (CAZy)

Carbohydrates are the main source of energy needed to maintain life activities and are the most widely distributed organic compounds in nature. The carbohydrate-related enzymes (CAZy) database collected six categories of enzymes, namely glycoside hydrolases (GHs), glycosyltransferases (GTs), carbohydrate esterases (CEs), carbohydrate-binding modules (CBMs), auxiliary module enzymes ( AAs ) and polysaccharide lyases (PLs)[17].

In this database, *Photobacterium kishitanii* FJ21 contains 51 carbohydrate-related enzymes. Among them, glycoside hydrolases (GHs) gene annotation is the most. There are 19 types of 43 genes, accounting for 34.4%. Glycosides produced by enzymatic hydrolysis of glycosidic bonds can be used in biological metabolic pathways. Glycosyltransferases (GTs) have 12 types and 32 genes, accounting for 25.6%. GTs can participate in a variety of life activities in cells, transferring monosaccharides of active substances in vivo to proteins, lipids, sugars, and nucleic acids to form glycosylation. Carbohydrate esterases (CEs) and carbohydrate-binding modules (CBMs) accounted for 16%, respectively. The auxiliary modular enzymes (AAs) and polysaccharide lyases (PLs) genes were the least.

## Phylogenetic tree

Blast alignment of 16S rRNA sequence with NCBI database showed that FJ21 belonged to Proteobacteria, Gammaproteobacteria, Vibrionales, Vibrionaceae, and *Photobacterium*.

Then the whole genome sequence was constructed a phylogenetic tree (Figure 7), and the results showed that *Photobacterium kishitanii*, *Photobacterium phosphoreum*, *Photobacterium aquimaris*, *Photobacterium malacitanum*, and *Photobacterium carnosum* were clustered together in the phylogenetic tree.

## lux genes analysis

The sequencing results showed that *luxC*(1437bp), *luxD*(921bp), *luxA*(1074bp), *luxB*(987bp), *luxF* (696bp), *luxE* (1122bp), *luxG*(705bp) genes existed in *Photobacterium kishitanii* FJ21.

## Physicochemical properties of the protein encoded by lux genes

Online software ExPASy (<https://www.expasy.org/>) was used to predict the physicochemical properties of proteins encoded by *luxC*, *luxD*, *luxA*, *luxB*, *luxF*, *luxE*, and *luxG* genes (Table 5).

Table 5  
Prediction of physical and chemical properties of protein encoded by *lux* genes.

Type	<i>luxC</i>	<i>luxD</i>	<i>luxA</i>	<i>luxB</i>	<i>luxF</i>	<i>luxE</i>	<i>luxG</i>
Number of base pairs	1437	921	1074	987	696	1122	705
Number of amino acids	478	306	357	328	231	373	234
Relative molecular mass	54165.62	34417.89	40536.85	37494.22	26616.19	42929.59	26147.99
Theoretical isoelectric point	5.38	4.98	5.21	5.26	5.07	5.09	5.43
Residual number of negative charge	62	42	51	42	32	52	27
Residual number of positive charge	47	28	36	27	25	38	23
Extinction coefficient	77155 or76780	34045 or33920	47245 or46870	30620 or30370	29005 or28880	47705 or47330	31775 or31400
Instability coefficient	37.57 stability	38.63 stability	34.29 stability	30.53 stability	41.84 instability	36.90 stability	33.71 stability
hydrophobicity	-0.223	-0.153	-0.381	-0.32	-0.362	-0.389	-0.029

It can be seen that the positive charge residues carried by the protein encoded by the gene are less than the negative charge residues, and the isoelectric point is between 4.98 and 5.43, indicating that the protein is easy to precipitate between these values. Only the protein encoded by *luxF* gene is unstable, and the protein encoded by other genes is stable. In addition, the hydrophobicity is negative, indicating that these proteins have certain hydrophilicity.

## Prediction of secondary and tertiary structures proteins encoded by *lux* genes

For unknown proteins, their secondary and tertiary structures can be predicted by amino acid sequences. The online analysis software PSIPRED was used to predict the secondary structure of the protein encoded by *lux* genes (Figure 8). The amino acid sequence of the pink part was  $\alpha$  helix, and the amino acid sequence of the yellow part was  $\beta$  folding. SWISS-MODEL database (<http://swissmodel.expasy.org/repository/>) was used to predict the tertiary structure of proteins (Figure 9). The protein structure in this database was predicted by the homology modeling method. When the sequence similarity between the predicted protein and the template

protein exceeds 30%, the homology modeling method can generate the tertiary structure of the protein with a prediction accuracy of 90%.

Except for *luxG* gene, the similarity of sequences encoded by other genes was more than 30% after alignment, so the tertiary structure of the protein was closer to the real structure. From the predicted secondary and tertiary structures of *lux* genes encoding proteins, it can be seen that the  $\alpha$  subunit and  $\beta$  subunit of the luciferase encoded by *luxA* and *luxB* genes have  $\beta$  folding barrel structures, which may be these two genes play an important role in the luminescence activity of luminescent bacteria. Understanding the tertiary structure of proteins is of great significance for studying functional structures (such as molecular docking and virtual screening).

## Genome collinearity analysis

### The basic characteristics of genomes

The genome size of the six strains is similar, ranging from 4380538bp to 4853277bp, and the number of coding genes is 3739-4131 (Table 6). The genome characteristics of different strains of the same bacteria are closer, the stain FJ21 is closer to *Photobacterium kishitanii* in genome size. Compared with the number of coding genes and RNA of other strains, the results showed that the number of coding genes and RNA predicted were significantly higher than others.

Table 6  
Genomic information of strains.

Species name	Strain	Genome size (bp)	Coding gene	tRNAs	5S rRNA	16S rRNA	23S rRNA
<i>s__Photobacterium kishitanii</i>	ATCC:BAA-1194	4732354	4087	159	6	6	5
<i>s__Photobacterium phosphoreum</i>	JCM 21184	4550107	3840	73	3	0	1
<i>s__Photobacterium aquimaris</i>	LC2-065	4525475	3826	83	1	0	1
<i>s__Photobacterium malacitanum</i>	CECT 9190	4380538	3739	186	10	15	12
<i>s__Photobacterium carnosum</i>	TMW 2.2021	4559543	3984	122	8	11	8
<i>Photobacterium kishitanii</i>	FJ21	4853277	4131	226	24	22	22

### Collinearity analysis

MUMmer (version 3.23) software was used to compare the strain FJ21 with *Photobacterium kishitanii*, *Photobacterium phosphoreum*, *Photobacterium aquimaris*, *Photobacterium malacitanum*, and *Photobacterium carnosum*. The collinearity and structural variation of genomic sequences are shown (Figure 10), and there were 218, 744, 717, 748 and 708 contrast blocks between *s\_\_ Photobacterium kishitanii*, *s\_\_Photobacterium phosphoreum*, *s\_\_Photobacterium aquimaris*, *s\_\_Photobacterium malacitanum*,

s\_*Photobacterium carnosum* and the strain FJ21, respectively. They accounted for 86.92%, 70.27%, 64.82%, 65.16% and 56.32% of the genome of the strain, respectively. According to the results, the collinearity between genomes is good, but there are a small number of genome rearrangement events such as inversion and translocation. It can be seen that the six strains still have great differences in evolution.

## Prediction and Analysis of Secondary Metabolite Gene Clusters

The encoding genes of secondary metabolites are usually clustered in the genome, encoding complex enzymes with multiple functions. AntiSMASH ( version 6.0.0 ) software was used to predict the gene cluster of the assembled genome. Three types of secondary metabolite gene clusters, RiPP-like, betalactone, and arylpolyene, were predicted in FJ21 genome(Table 7). 90% similarity between arylpolyenes and APE Vf synthetic gene clusters.

Table 7  
Gene clusters of secondary metabolite of FJ21.

Cluster type	Start gene and End gene	Ratio of genes show similarity (%)	Most similar known cluster (%)
RiPP-like	01645-01653	NZ_LN794352(100%) NZ_PYNJ01000046(100%) NZ_MSCC01000006(100%)	-
Betalactone	01929-01947	NZ_LN794352(100%) NZ_MSCC01000001(100%) NZ_MSCQ01000002(100%)	-
Arylpolyene	03136-03175	NZ_LN794352(97%) NZ_MSCQ01000002(94%) NZ_MSCC01000002(83%)	BGC0000837: APE Vf ( 90%) BGC0002008:arylpolyenes( 50%) BGC0000836:APE Ec( 42%)

## Discussion

Nanopore sequencing technology was used to obtain the whole genome sequence and bacterial complete map of *Photobacterium kishitanii* FJ21. Amino acid metabolism and transport gene abundance were the highest in COG functional annotation. Molecular sensor activity, antioxidant activity and transport activity in GO functional annotation accounted for a large number of genes, indicating that the molecular function of the strain was better. The number of genes related to metabolism in KEGG pathway is the largest, so the strain has a strong metabolic ability. Most of the genes annotated in the PHI database belonged to the reduced virulence genes and unaffected pathogenicity genes, indicating that the pathogenic ability of the strain was weak or nonpathogenic. Most luminous bacteria are nonpathogenic[18], while two subspecies of

*Vibrio harveyi* and *Photobacterium damsela* are pathogens of many aquatic organisms[19, 20]. Many species of *Vibrio* and *Photobacterium* have multiple resistance to many common antibiotics. Therefore, we annotated the antibiotic resistance gene of *Photobacterium kishitanii* FJ21 using the CARD database, and found that the strain was resistant to eight kinds of antibiotics, and the resistance mechanism was antibiotic efflux and target site change.

The 16S rRNA sequence of FJ21 was blastn compared with the NCBI database, and the results showed that FJ21 belonged to *Photobacterium*. Then the whole genome sequence was constructed a phylogenetic tree. According to the predicted secondary and tertiary structure of *lux* gene and its encoded protein, the strain contained *luxC*, *luxD*, *luxA*, *luxB*, *luxF*, *luxE* and *luxG* genes, and the  $\alpha$  subunit and  $\beta$  subunit of the fluorescent enzyme encoded by *luxA* and *luxB* genes had a  $\beta$  fold barrel structure, which was due to the important role of these two genes in the luminescence activity of luminescent bacteria. However, the function of *luxF* gene and the effect of *luxF* gene are still uncertain, which needs further study. Recent studies on the effect of *luxF* on light intensity in bacterial bioluminescence have shown that *luxF* has a significant effect on the maximum light intensity[21].

In the genome collinearity analysis, the genome sizes and basic characteristics of the six strains were similar, and the genome collinearity was good, but there were a small number of genome rearrangement events such as inversion and translocation. Inversions and translocations may result in the fusion of genes that further alter certain functions in species[22], and this may be one of the reasons why different strains of the same species differ. *Photobacterium kishitanii* FJ21 has higher homology with *Photobacterium kishitanii* ATCC:BAA-1194. By comparing the number of coding genes and RNA of strains, the results showed that the number of coding genes, tRNA, and rRNA predicted by *Photobacterium kishitanii* FJ21 were significantly higher than those of the other five strains, which may be related to strain evolution and environmental adaptation.

The completion of the whole-genome sequencing provides the whole genome information of *Photobacterium kishitanii* FJ21, and provides the basis for the further study of *lux* genes, luminous activities and other functional genes of the strain. Subsequently, recombinant luminescent bacteria can be explored or the *lux* genes can be recombined into the vector to realize the monitoring of acute toxic substances. For example, the vector is constructed and introduced into *E. coli*[23], so that *E. coli* can play the role of luminescent gene[24].

## Conclusions

Through whole genome sequencing, the genome collinearity analysis and bioinformatics analysis it was possible to determine the strain FJ21 is *Photobacterium kishitanii* and analyze the *lux* genes. The strain contains a chromosome with a total length of 4853277bp, and GC content of 39.23%. According to the predicted secondary and tertiary structure of *lux* gene and its encoded protein, the strain contained *luxC*, *luxD*, *luxA*, *luxB*, *luxF*, *luxE* and *luxG* genes. However, the function of *luxF* gene and the effect of *luxF* gene are still uncertain. Understanding the *lux* genes will help to understand luminescent activities and mechanism of *Photobacterium kishitanii* FJ21. Therefore, the strain can be used in water quality testing or monitoring of acute toxic substances.

# Materials And Methods

## Strain and Growth Conditions.

The strain FJ21 was isolated from the Microbiology Laboratory of College of Biological Science and Technology, Hunan Agricultural University. 2 ml 3% NaCl solution was added into the freeze-dried powder of *Photobacterium kishitanii* FJ21. After three generations of plate separation and streaking, it was transferred to liquid medium (Beef extract 0.5%, tryptone 0.5%, Na<sub>2</sub>HPO<sub>4</sub> 0.6%, KH<sub>2</sub>PO<sub>4</sub> 0.275%, glycerol 0.3%, and NaCl 3%). After 24 hours of culture at 25°C and 180 r / min, the glycerol-preserved strain was sent to the company for genome sequencing.

## Genome Sequencing and Assembly

Nanopore sequencing technology[25] was used to complete the genome scanning and sequencing of the strain. Firstly, high-quality DNA was extracted with Qiagen kit, and ID library was constructed. The DNA was sequenced by single molecule using Oxford Nanopore Technology sequencing instrument PromethION to obtain the original sequencing data. After quality control of the obtained sequencing data, the whole genome scanning of the strain was completed by bioinformatics analysis.

Assembly: the three-generation data after quality control were assembled with flye, corrected with racon combined with three-generation sequencing data. The corrected genome uses its script to detect whether the loop is formed. After removing redundant loops, the origin of the sequence is moved to the replication initiation site of the genome by the circlator[26], to obtain the final genome sequence.

Sequencing depth analysis: After the assembly was completed, the quality control was compared with minimap2[27] to the genome.

## Genome Structure and Function Annotation

Genome structure prediction includes coding gene prediction, non-coding gene prediction, CRISPR prediction and gene island prediction. The coding genes was predicted by prodigal[28]. In the prediction of non-coding genes, RNAmmer[29] and tRNAscan-SE2.0[30] software were used to predict rRNA and tRNA in the genome, respectively. Other non-coding RNAs (ncRNAs) were predicted by the Infernal 1.1[31] search Rfam 13.0 database[32], and the predicted length was greater than 80% of the sequence length in the database. CRISPR was predicted by minced, and gene island was predicted by Islander[33].

After extracting the genome-coded proteins, InterProScan 5[34] was used for annotation, and the annotation information of TIGRFAMs[35], Pfam[36], and GO[37]databases were extracted. Blastp was used to compare the encoded proteins to KEGG[38], Refseq[39] and COG[40] databases, and the best results with a coverage of more than 30% were retained as annotation results. The interaction genes between pathogen and host were annotated by PHI database, and the antibiotic resistance genes were annotated by CARD database.

## Phylogenetic tree Analysis

MEGA7.0 software was used to analyze the strain *Photobacterium kishitanii* FJ21 and construct a phylogenetic tree by the Neighbor-Joining method.

## **Prediction of protein structure encoded by luxABCDEF gene**

Prediction of protein secondary structure by protein analysis online software SWISS-MODEL.

## **Genome collinearity analysis**

The whole-genome sequencing of *Photobacterium kishitanii* FJ21 was analyzed by collinearity with other genome sequences with similarity of 95% in the NCBI database. MUMmer ( version 3.23 ) was used to quickly compare the genomes of *Photobacterium kishitanii* FJ21 and five closely related strains ( s\_\_*Photobacterium kishitanii*, s\_\_*Photobacterium phosphoreum*, s\_\_ *Photobacterium aquimaris*, s\_\_*Photobacterium malacitanum*, s\_\_*Photobacterium carnosum* ). Visualize each Contig of the genome using a brown box in the ggplot2 package in R language. Yellow lines between genomes represent Colinear and blue lines between genomes represent Inversion.

## **Prediction of secondary metabolite gene cluster**

The assembled genome was analyzed by using antiSMASH version 6.0.0, and the parameters were selected from taxon bacteria.

## **Declarations**

### ***Ethics approval and consent to participate***

No formal ethics approval was required in this particular case.

### ***Consent for publication***

Not applicable

### ***Availability of Data and materials***

The available complete genome sequence has been admitted to GenBank databases with accession number SRX10356131.

### ***Competing interests***

The authors declare that they have no competing interests in this paper.

### ***Funding***

This study was supported and funded by the National Natural Science Foundation of China(31672457) and the Project of Hunan Provincial Department of science and technology(2019TP2004, 2020NK2004, 2021JJ30008).

## ***Authors' contributions***

Gang Liu and Jun Fang planned and directed the project. Shuzhneg Yin, Zilong Yi and Jia LIU drafted the manuscript. All authors read and approved the final manuscript.

## ***Acknowledgments***

Not applicable

## **References**

1. Du ZJ, Wang XH, Li HF, Chi ZM. The study and application of luminescent bacteria. *Chinese High Technology Letters*. 2003;(12):103–106.
2. Brodl E, Winkler A, Macheroux P. Molecular Mechanisms of Bacterial Bioluminescence. *Comput Struct Biotechnol J*. 2018;16:551–564.
3. Gong L, Wu Y, Jian QJ, Yin CX, Li TT, Gupta VK, et al. Complete genome sequencing of the luminescent bacterium, *Vibrio qinghaiensis* sp. Q67 using PacBio technology. *Scientific Data*. 2018;5:170205.
4. Zavgelsky GB, Shakulov RS. Mechanisms and Origin of Bacterial Bioluminescence. *Mol Biol (Mosk)*. 2018;52(6):935–947.
5. Brodl E, Niederhauser J, Macheroux P. In Situ Measurement and Correlation of Cell Density and Light Emission of Bioluminescent Bacteria. *J Vis Exp*. 2018;(136):57881.
6. Meighen EA. Bacterial bioluminescence: organization, regulation, and application of the lux genes. *FASEB J*. 1993;7(11):1016–22.
7. Alloush HM, Lewis RJ, Salisbury VC. Bacterial Bioluminescent Biosensors: Applications in Food and Environmental Monitoring. *Analytical Letters*, 2006;39(8):1517–1526.
8. Oku N, Kawabata K, Adachi K, Katsuta A, Shizuri Y. Unnarmicins A and C, new antibacterial depsipeptides produced by marine bacterium *Photobacterium* sp. MBIC06485. *J Antibiot (Tokyo)*. 2008;61(1):11–7.
9. Ryu HS, Kim HK, Choi WC, Kim MH, Park SY, Han NS, et al. New cold-adapted lipase from *Photobacterium lipolyticum* sp. nov. that is closely related to filamentous fungal lipases. *Appl Microbiol Biotechnol*. 2006;70(3):321–6.
10. Yaacob MA, Hasan WANW, Ali MSM, Rahman RNZRA, Salleh AB, Basri M, et al. Characterisation and molecular dynamic simulations of J15 asparaginase from *Photobacterium* sp. strain J15. *Acta Biochim Pol*. 2014;61(4):745–52.
11. Shakiba MH, Ali MSM, Rahman RNZRA, Salleh AB, Leow TC. Cloning, expression and characterization of a novel coldadapted GDSL family esterase from *Photobacterium* sp. strain J15. *Extremophiles*. 2016;20(1):44–55.
12. Girotti S, Ferri EN, Fumo MG, Maiolini E. Monitoring of environmental pollutants by bioluminescent bacteria. *Anal Chim Acta*. 2008;608(1):2–29.

13. Adnan NA, Halmi MIE, Abd Gani SS, Zaidan UH, Adb shukor MY. Comparison of Joint Effect of Acute and Chronic Toxicity for Combined Assessment of Heavy Metals on *Photobacterium* sp. NAA-MIE. *Int J Environ Res Public Health*. 2021;18(12):6644.
14. Zhu WJ, Xu YT, Zhang QZ, Yang J, Ding GR, Wang XY. Application and development of monitoring with luminous bacteria for environmental pollutants. *Water Purification Technology*. 2010;29(04):54–59.
15. Li ZX, Yin Q, Zhao JN, Li GX, Liu YF, Li W, et al. Advance in the application of luminescent bacteria on detecting acute toxicity of water. *Journal of Hebei University of Science and Technology*. 2014;35(05):480–486.
16. Li EH, Yang HL, Wang XL, Wan LF, Pan HM, Zhu D. Whole-genome sequencing and analysis of inosine-producing strain *Bacillus subtilis* ATCC 13952. *Acta Microbiologica Sinica*. 2015;55(12):1560–1567.
17. Zhang BY, Zhu TH, Han S, Wang Y, Li SJ, Qiao TM. Whole-genome sequencing and analysis of *Streptomyces sampsonii* KJ40. *Acta Microbiologica Sinica*. 2018;45(04):805–818.
18. Feng GZ, Shi H, Liu B, Wu YT, Wang YL, Shi Y. Screening and Identification of Cellulase-producing Strains Isolated from Crayfish Intestine. *Biotechnology Bulletin*. 2020;36(02):65–70.
19. Brodl E, Csamay A, Horn C, Niederhauser J, Weber H, Macheroux P. The impact of LuxF on light intensity in bacterial bioluminescence. *J Photochem Photobiol B*. 2020;207:111881.
20. Moi IM, Roslan NN, Leow ATC, Ali MSM, Abd Rahman RNZR, Rahimpour A, et al. The biology and the importance of *Photobacterium* species. *Appl Microbiol Biotechnol*. 2017;101(11):4371–4385.
21. Ke HM, Prachumwat A, Yu CP, Yang YT, Promsri S, Liu KF, et al. Comparative genomics of *Vibrio campbellii* strains and core species of the *Vibrio Harveyi* clade. *Sci Rep*. 2017;7:41394.
22. Urbanczyk H, Ast JC, Dunlap PV. Phylogeny, genomics, and symbiosis of *Photobacterium*. *FEMS Microbiol Rev*. 2011;35(2):324–42.
23. Cheng XM, He J, Gan WD, Fang XH, Yang J, Zhu B, et al. Pseudocapsule of renal cell carcinoma associated with Xp11.2 translocation/TFE3 gene fusion: a clue for tumor enucleation? *Int J Clin Exp Pathol*. 2015;8(5):5403–10.
24. Wang YW, Ding W, Li P, Zhang LL, Wei XY. Prokaryotic expression of luciferase gene lux AB of *Vibrio qinghaiensis*. *Journal of Chinese Institute of Food Science and Technology*. 2015;15(01):60–65.
25. Luo AH, Feng X, Ning BA, Bai JL, Peng Y, Gao ZX. Comparison of the expression and luminescence of luxAB and luxCDABE genes from *Vibrio qinghaiensis* sp.-Q67 in *Escherichia coli*. *Journal of Food Safety & Quality*. 2015;6(12):4816–4821.
26. Senol Cali D, Kim JS, Ghose S, Alkan C, Mutul O. Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions. *Brief Bioinform*. 2019;20(4):1542–1559.
27. Hunt M, De Silva N, Otto TD, parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biology*. 2015;16:294.
28. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11:119.  
<https://doi.org/10.1186/1471-2105-11-119>.

29. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007,35(9):3100–3108.
30. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997,25(5):955–964.
31. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* 2013,29(22):2933–2935. <https://doi.org/10.1093/bioinformatics/btt509>.
32. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 2018,46(D1):D335-D342.
33. Hudson CM, Lau BY, Williams KP. Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res.* 2015,43(Database issue):D48-53.
34. Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014,30(9):1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
35. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Res.* 2013,41(Database issue):D387-395.
36. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016,44(D1):D279-285.
37. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 2004,32(suppl 1):D258-261.
38. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000,28(1):27–30.
39. O'Leary NA, Wright MW, Brister JR, Ciufu S, McVeigh DHR, Rajput B, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016,44(D1):D733-745.
40. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 2015,43(Database issue):D261-269.

## Figures

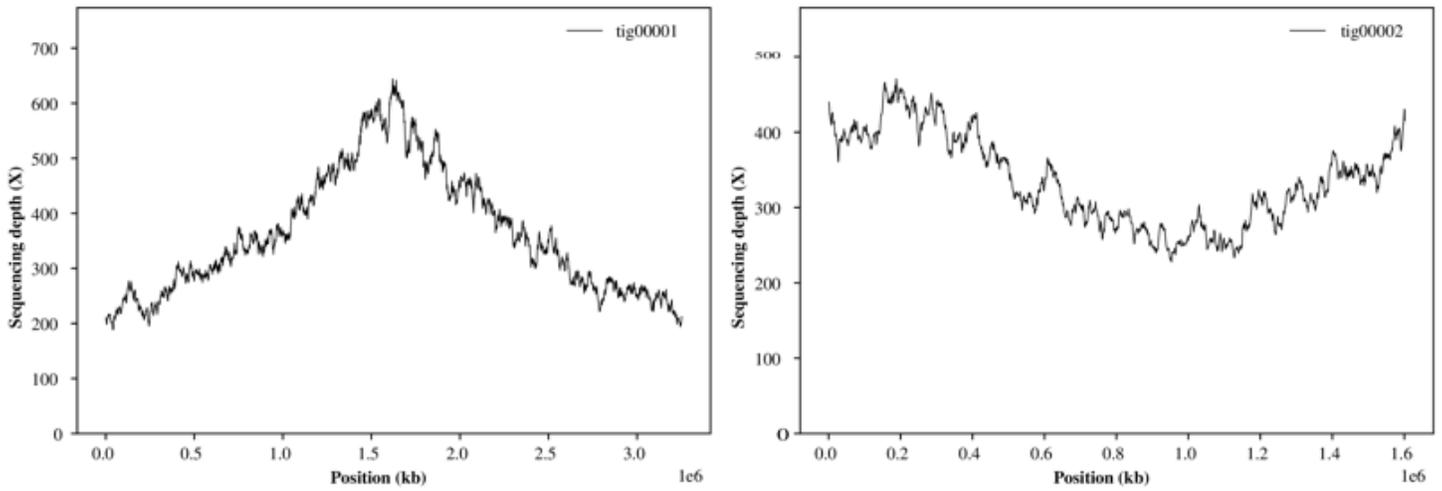


Figure 1

Sequencing depth distribution.

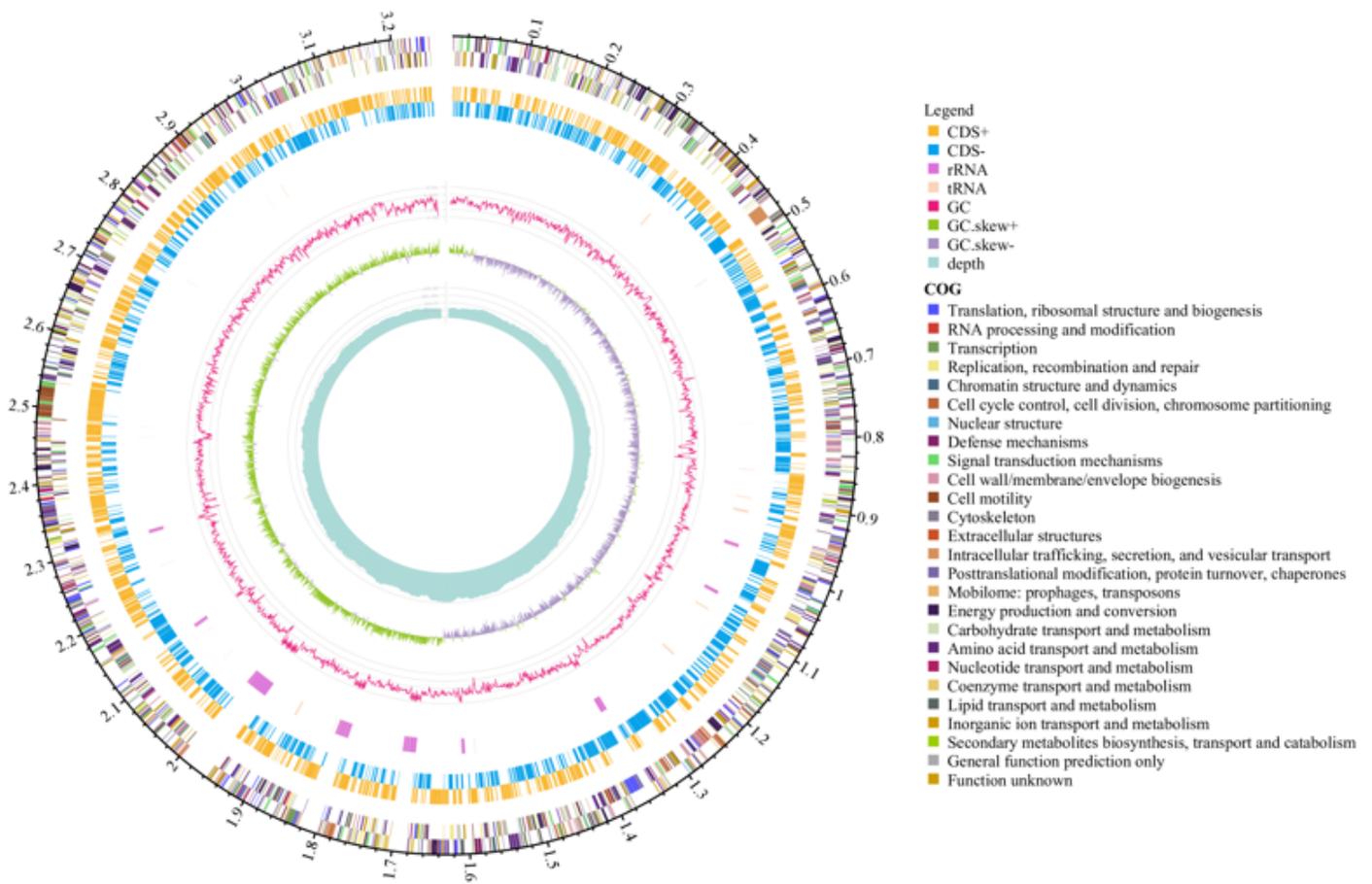
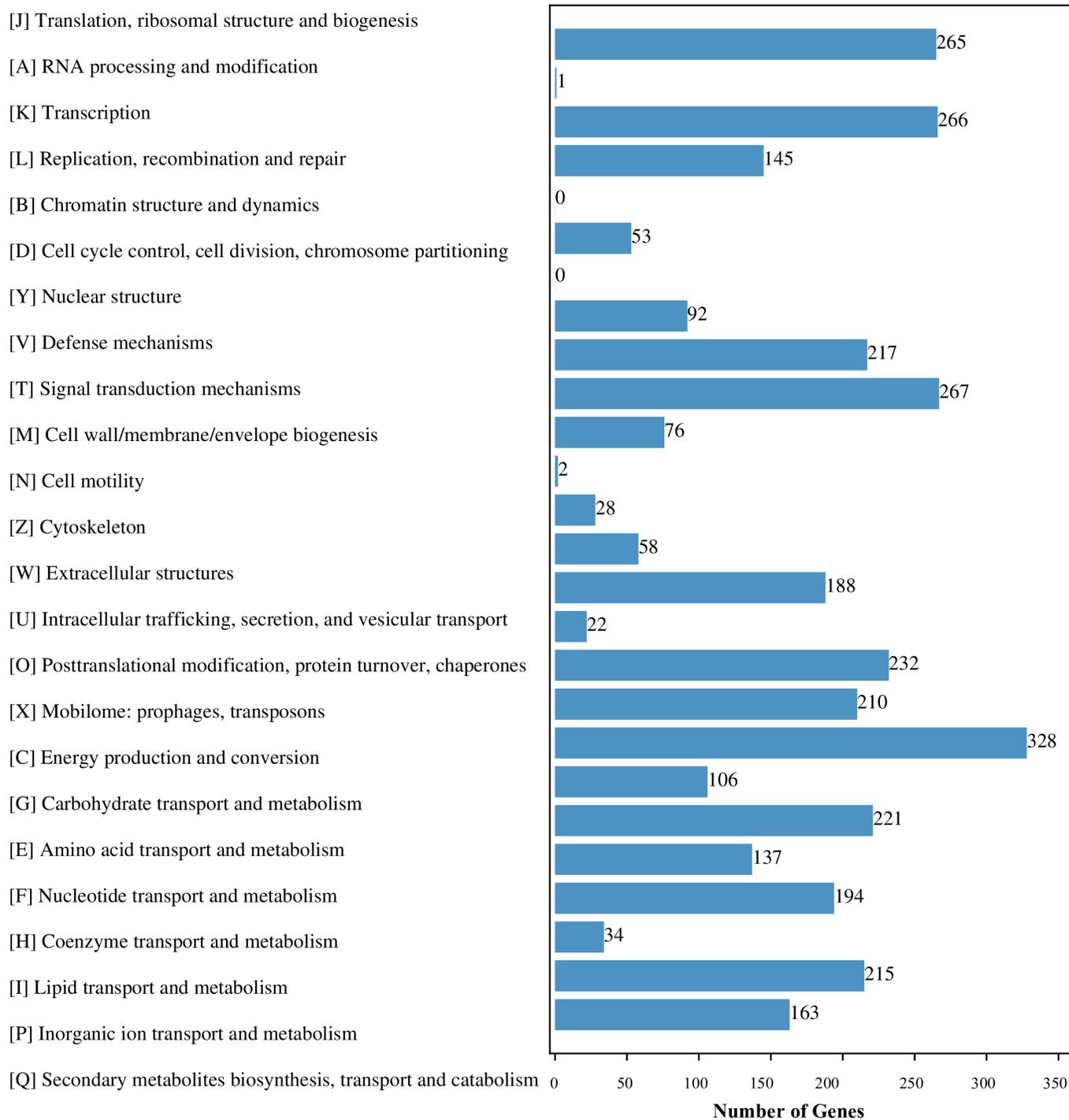


Figure 2

Genome circle diagram of *Photobacterium kishitanii* FJ21.



**Figure 3**

COG functional classification of *Photobacterium kishitanii* FJ21.

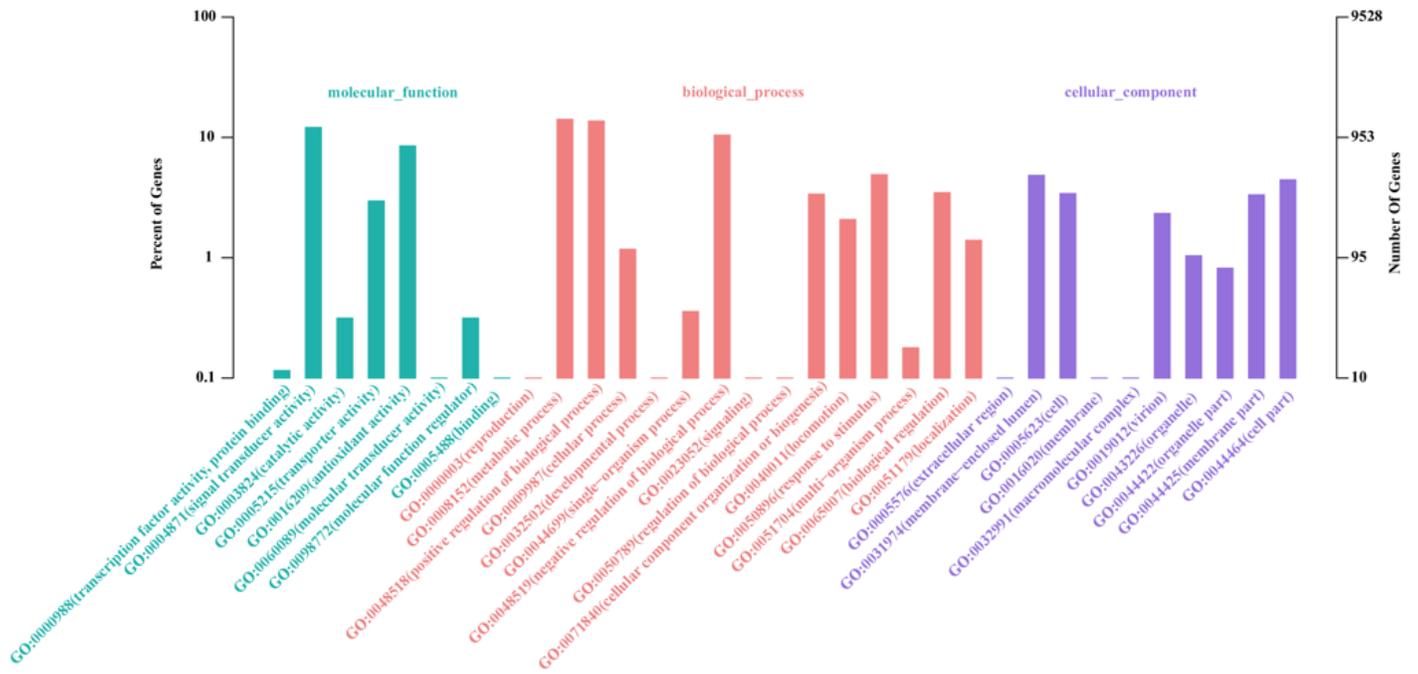
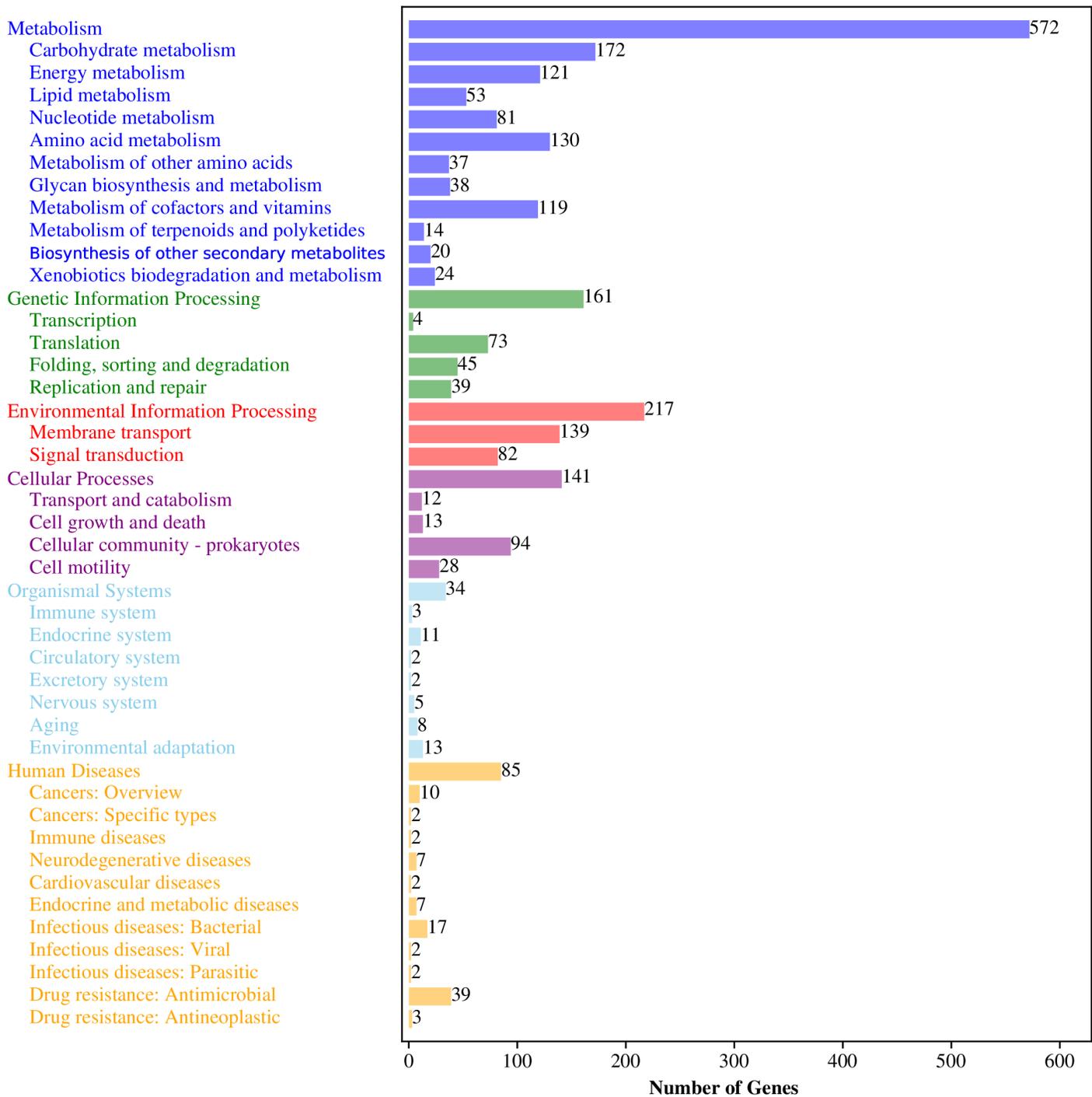


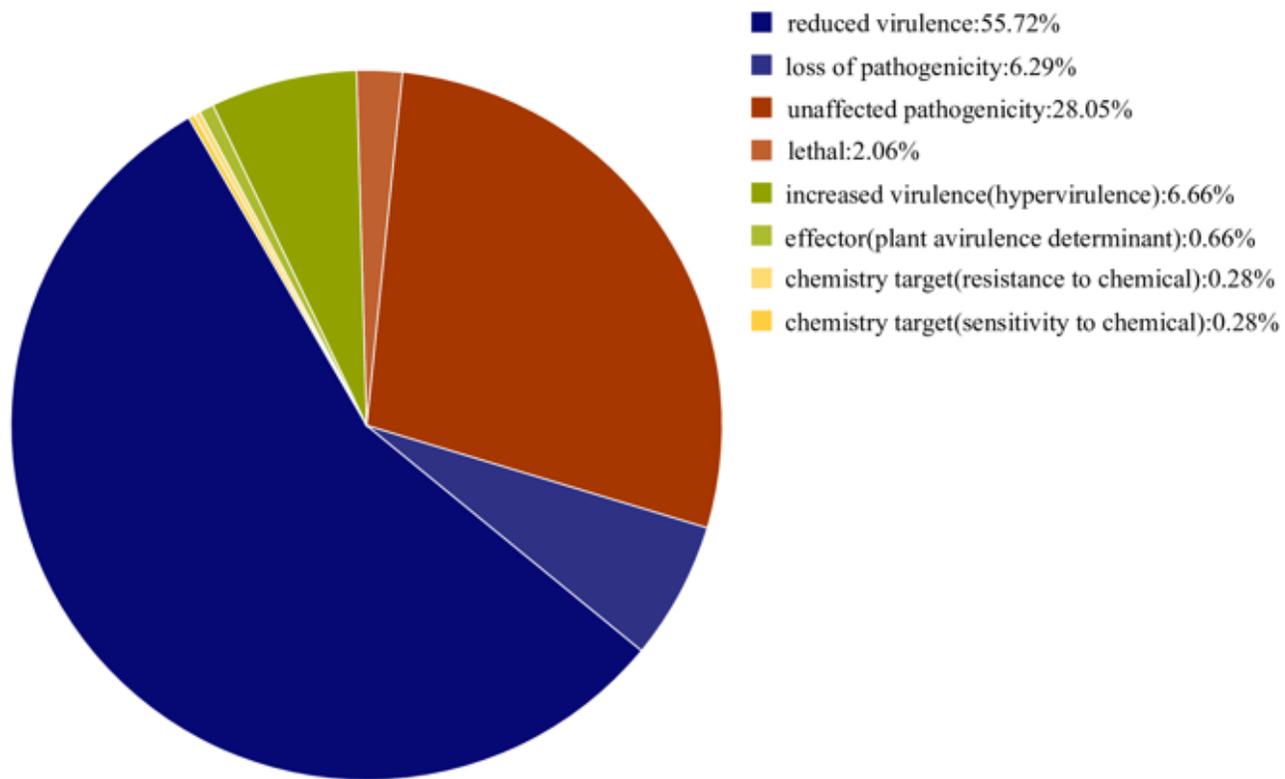
Figure 4

Go function classification diagram.



**Figure 5**

KEGG function classification diagram.



**Figure 6**

Gene analysis of pathogen host interaction.

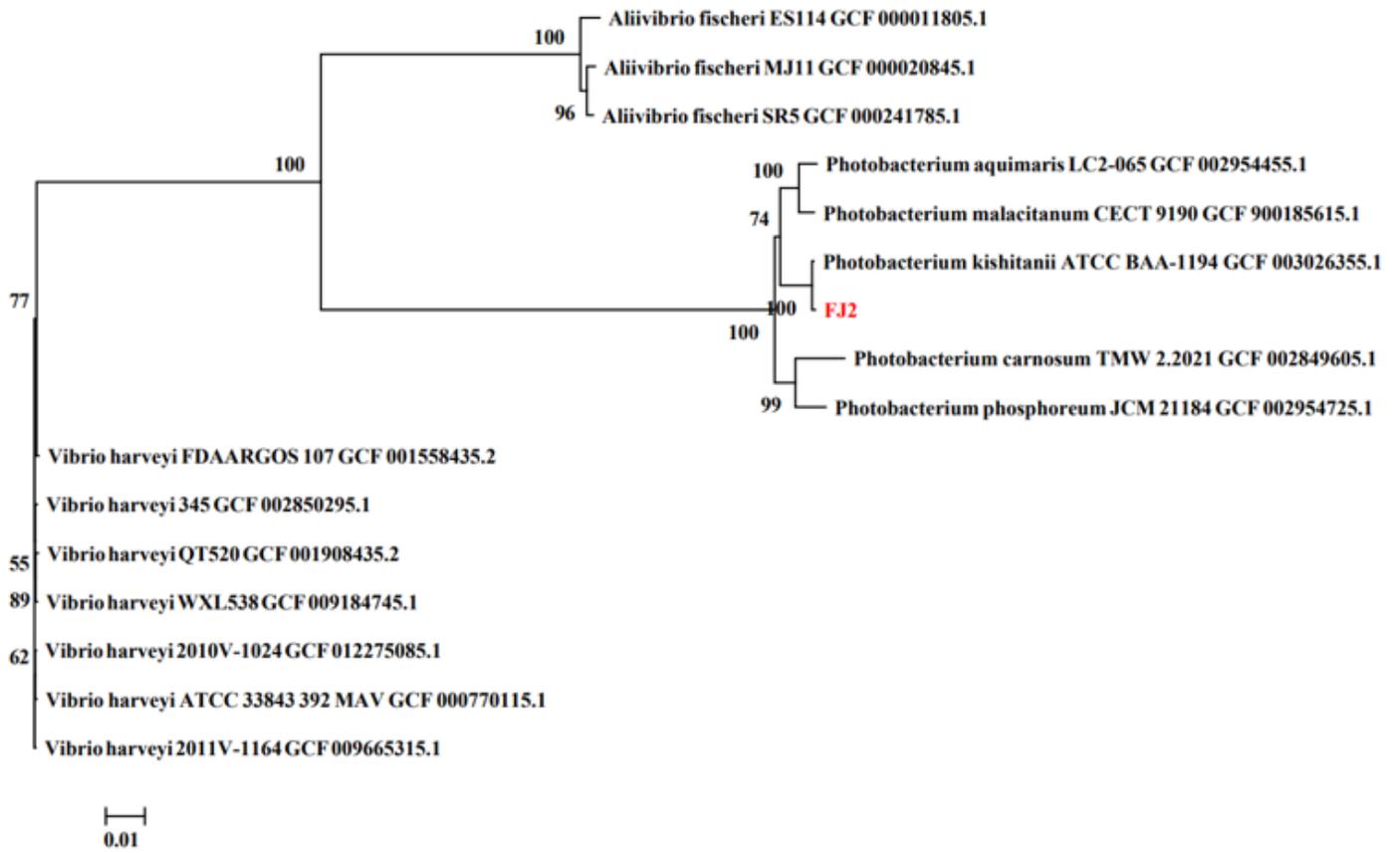


Figure 7

Phylogenetic analysis of *Photobacterium kishitanii* FJ21 and high evolutionary similar strains.

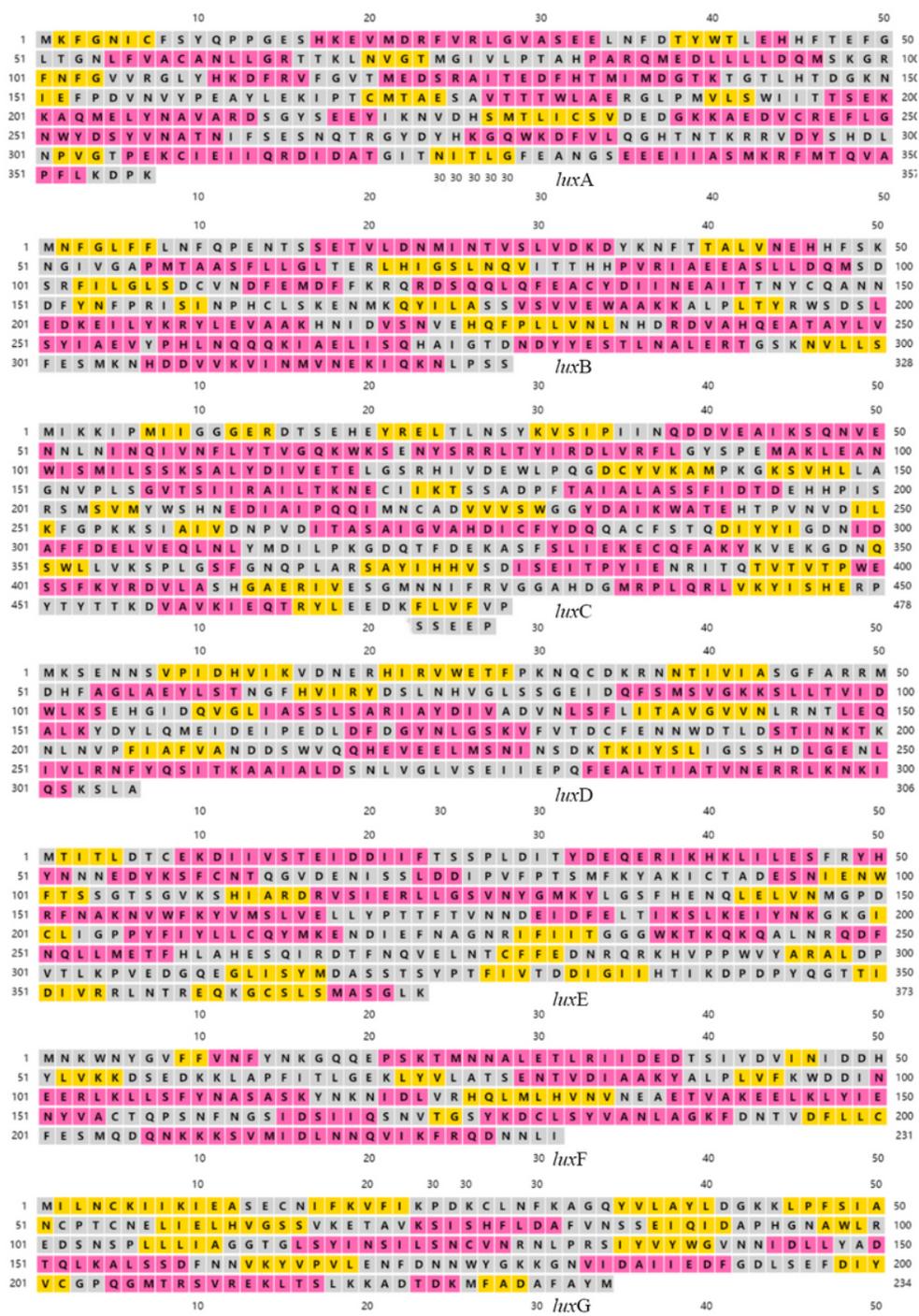
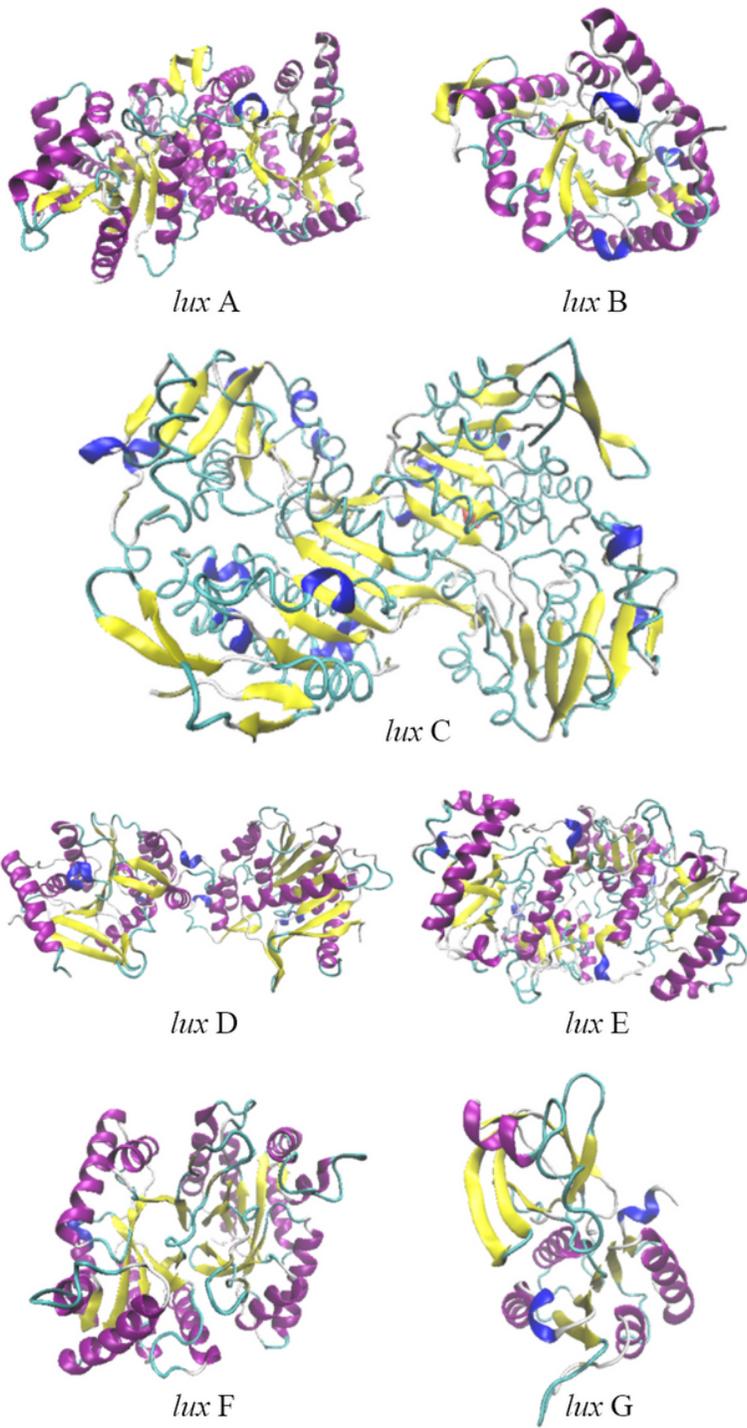


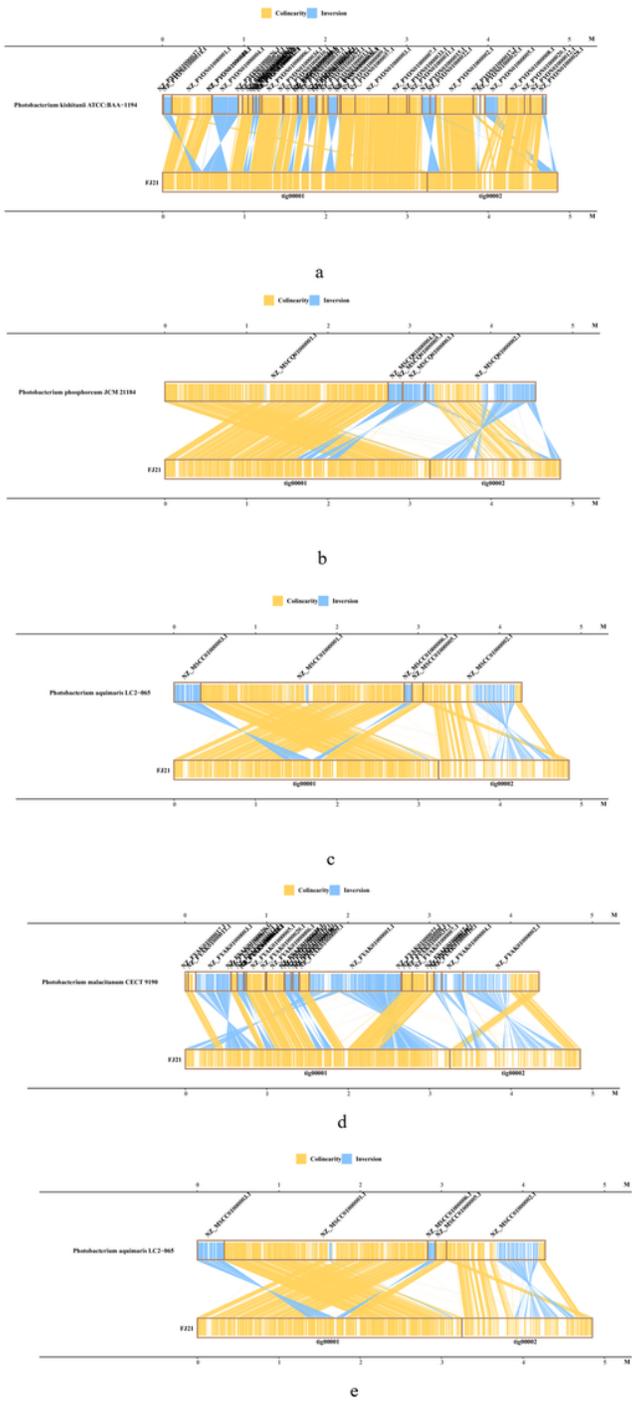
Figure 8

Prediction of protein secondary structure encoded by *lux* genes.



**Figure 9**

Prediction of protein tertiary structure encoded by *lux* genes.



**Figure 10**

Global genome alignment of *Photobacterium*. Note: a: Global genome alignment of s\_ *Photobacterium kishitani* and FJ21 b: Global genome alignment of s\_ *Photobacterium phosphoreum* and FJ21 c: Global genome alignment of s\_ *Photobacterium aquimaris* and FJ21 d: Global genome alignment of s\_ *Photobacterium malacitanum* and FJ21 e: Global genome alignment of s\_ *Photobacterium carnosum* and FJ21