MDPI

Type of the Paper (Article, Review, Communication, etc.)

Whole-genome sequencing and single acute toxicity of heavy metal to *Photobacterium kishitanii* FJ21

Shuzheng Yin , Zilong Yi, Jia Liu, Gang Liu, and Jun Fang*

Hunan Provincial Engineering Research Center of Applied Microbial Resources Development for Livestock and Poultry, College of Bioscience and Biotechnology, Hunan Agricultural University, Changsha 410128, China;

* Correspondence: fangjun1973@hunau.edu.cn (J.F.)

Abstract: In this study, the growth morphology of FJ21 strain was observed, and its 16S rRNA and whole genome were sequenced. Then, related software was used to make genome assembly, gene structure and function annotation, genome phylogenetic tree analysis, genome collinearity analysis and prediction of secondary metabolic gene cluster analysis. Finally, the single acute toxicity of five heavy metals to FJ21 strain was detected. There were *lux*C, *lux*D, *lux*A, *lux*B, *lux*F, *lux*E and *lux*G genes in FJ21, and the protein encoded by *lux* operon had certain hydrophilicity. The genome of this strain FJ21 contains a chromosome with a total length of 4853277bp and a GC content of 39.23%. The genome of FJ21 was compared with that of *Photobacterium kishitanii* ATCCBAA-1194, *Photobacterium phosphoreum* JCM21184, *Photobacterium aquimaris* LC2-065, *Photobacterium malacitanum* CECT9190, and *Photobacterium carnosum* TMW 2.2021. The average nucleotide identity(ANI), tetra nucleotide signatures (Tetra), comparative genome, and phylogenetic analysis proposed that FJ21 is a strain of *Photobacterium kishitanii*. In the acute toxicity test, the toxicity of heavy metals to the strain FJ21 is $Pb(NO_3)_2$ > $ZnSO_4{\cdot}7H_2O$ > $CdCl_2{\cdot}2.5H_2O$ > $CuSO_4{\cdot}5H_2O$ > $K_2Cr_2O_7$.

Keywords: luminescent bacteria; whole-genome sequencing; comparative analysis; phylogenetic analysis; *lux* operon; the acute toxicity test

1. Introduction

Luminescent bacteria are a group of Gram-negative bacteria that can emit blue-green fluorescence under normal conditions, and live in the marine environment mainly in the form of free organisms or parasites. The bioluminescence of luminescent bacteria is regulated by the enzyme-catalyzed reaction encoded by the lux operon. The *lux*A and *lux*B genes encode the α and β subunits of luciferase, respectively. *lux*C, *lux*D a,nd *lux*E constitute the fatty acid reductase complex, responsible for the synthesis of long Aldehyde substrate, luxG encodes flavin reductase[1]. Conserved genes *lux*C, *lux*D, *lux*A, *lux*B, *lux*E, and *lux*G exist in all luminescent bacteria that have been discovered[2], in addition, there are other genes such as *lux*I, *lux*R ,and *lux*F[3]. Although luminescent bacteria have the same *lux* operon, these bacteria show great differences in characteristics such as growth behavior, luminescence intensity, or bioluminescence regulation[4]. The luminescence process of luminescent bacteriis oxidizedze $FMNH_2$ and RCHO to FMN and RCOOH under the catalysis of intracellular specific luciferase and the participation of molecular oxyen , and at the same time release blue-green light. The luminescence reaction is as follows:

$FMNH_2+O_2+RCHO \rightarrow FMN+RCOOH+H_2O+light$ (450~490nm)

Luminescent bacteria can not only act as biosensors[5] but also produce antibacterial compounds[6], lipase[7], asparagine[8] ,and esterase[9]. 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[10], and the acute and chronic toxicity tests of heavy metal mixtures[11]. At present, the luminescent bacteria commonly used in water quality and environmental monitoring are *Photobacterium phosphoreum*, *Vibrio fischeri*, and *Vibrio Qinghaiensis*[12]. 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[13]. In recent years, 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.

In the study, 16S rRNA, genome phylogenetic tree analyses, comparative genomics, average nucleotide identity (ANI), and tetra nucleotide signatures (Tetra) were used to clarify the strain FJ21 to *Photobacterium kishitanii*. In addition, the single acute toxicity of heavy metals to strain FJ21 was detected, which provided some data support for ecological risk assessment of heavy metal pollution.

2. Materials and Methods

2.1. The growth morphology, 16S rRNA and lux operator analysis

Beef extract, tryptone, sodium chloride, potassium dihydrogen phosphate, disodium hydrogen phosphate, glycerol, and agar. After the glycerol-preserved strain was cultured at 25℃ for 24h, its morphological characteristics (size, shape, transparency, color, etc.) were observed.

Bacterial 16S rRNA universal primer was selected, and the PCR products were sent to Sangong Biotech (Shanghai) Co., Ltd. for sequencing, and the sequencing results were blast analyzed in the GenBank database. At the same time, the phylogenetic tree was constructed with mega-x software.

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

2.2. Genome Sequencing and Assembly

Nanopore sequencing technology[14] was used to complete the genome scanning and sequencing of the strain. Firstly, high-quality DNA was extracted with a Qiagen kit, anan d ID library was constructed. The DNA was sequencea d by single moleculethe 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(parameter： --plasmids --nano-raw), corrected with racon(parameter： default) combined with three-generation sequencing data, and corrected with pilon[15] or NextPolish[16] (parameter： default) combined with two-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 circulator[17] (parameter： fixstart), to obtain the final genome sequence.

2.3. Genome Structure and Function Annotation

Genome structure prediction includes coding gene prediction, non-coding gene prediction, CRISPR prediction, and gene island prediction. The coding gene was predicted by prodigal[18] (parameter： -p None -g 11), and the complete CDS was

retained. In the prediction of non-coding genes, RNAmmer[19] (parameter：-kingdom bac) and tRNAscan-SE2.0[20] (parameter：-B -I -m lsu,ssu,tsu) software were used to predict rRNA and tRNA in the genome, respectively. Other non-coding RNAs (ncRNAs) were predicted by the Infernal 1.1 [21] (parameter： --cut_ga --rfam --nohmmonly) search Pfam 13.0 database [22], and the predicted length was greater than 80 % of the sequence length in the database. CRISPR(parameter ： default) was predicted by minced, and gene island was predicted by Islander[23] (parameter r：default).

After extracting the genome-coded proteins, InterProScan 5[24] was used for annotation, and the annotation information of TIGRFAMs[25], Pfam[26], and GO[27]databases were extracted. Blastp was used to compare the encoded proteins to KEGG[28], Refseq[29], and COG[30] 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 bythe PHI database, and the antibiotic resistance genes were annotatethe d by CARD database.

2.4. Phylogenetic tree Analysis and Sequence-based method for species identification

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

The According to the recommended cut-off values for species determination (<95% for ANIb/ANIm and <0.989 for Tetra) [31,32], the calculation of average nucleotide identity based on BLAST(ANIb)/MUMmer(ANIm) and the correlation indexes of tetra nucleotide signatures (Tetra) were conducted using JspeciesWS (http://jspecies.ribohost.com/jspeciesws/#Analyse) [33,34].

2.5. Genome collinearity analysis

The whole-genome sequencing of strain FJ21 was analyzed by collinearity with other genome sequences with a similarity of 95 % in the NCBI database. MUMmer (version 3.23) was used to quickly compare the genomes of strain FJ21 and five closely related strains (*Photobacterium kishitanii* ATCCBAA-1194, *Photobacterium phosphoreum* JCM21184, *Photobacterium aquimaris* LC2-065, *Photobacterium malacitanum* CECT9190, and *Photobacterium carnosum* TMW 2.2021). 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.

2.6. 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.

2.7. For detection of heavy metal toxicity

The slant test-tube strains were eluted with 2ml of 3% NaCl solution, and after fully shaking, 200μL was taken into 50 ml of liquid culture medium, and cultured at 25 ℃ for 20-22h. Take 200μL bacterial solution into 40 ml of 3% NaCl solution and stir for 40 min for detection.

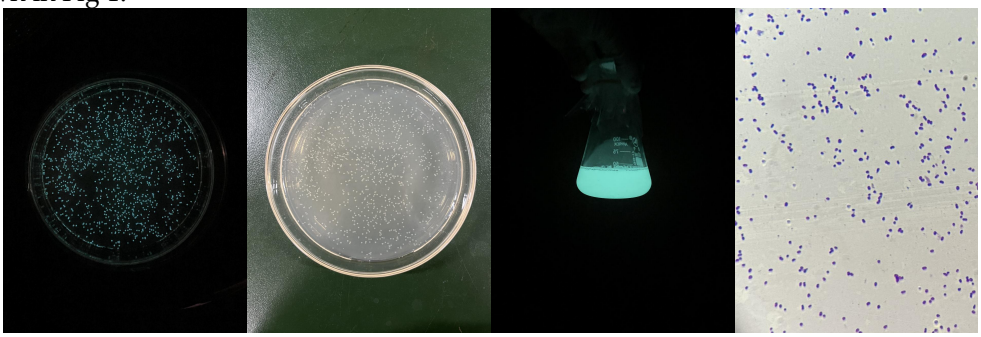
Dilute the solution of heavy metals($ZnSO4{\cdot}7H_2O$, $CuSO_4{\cdot}5H_2O$, $Pb(NO_3)_2$, $CdCl_2{\cdot}2.5H_2O$, $K_2Cr_2O_7$) to be measured with 3%NaCl into a series of concentration gradients. Add 0.1mL of the balanced bacterial solution into each sample tube, then take 1.9mL of the substance to be tested, and take 1.9mL of 3%NaCl as a blank. After 15min exposure, the luminous intensity was measured by portable Lux-T020 toxicity analyzer, and the dose-effect curve was drawn to obtain the half effect concentration (EC_{50}).

3. Results and Discussion

3.1. The growth morphology，16S rRNA and lux operator analysis

142 3.1.1. Growth morphology of strain

143 After the strain is cultured for 24 hours, the colony shape is round, slightly raised,
144 smooth, milky white and short rod, which emits bright blue-green light in the dark, as
145 shown in Fig 1.



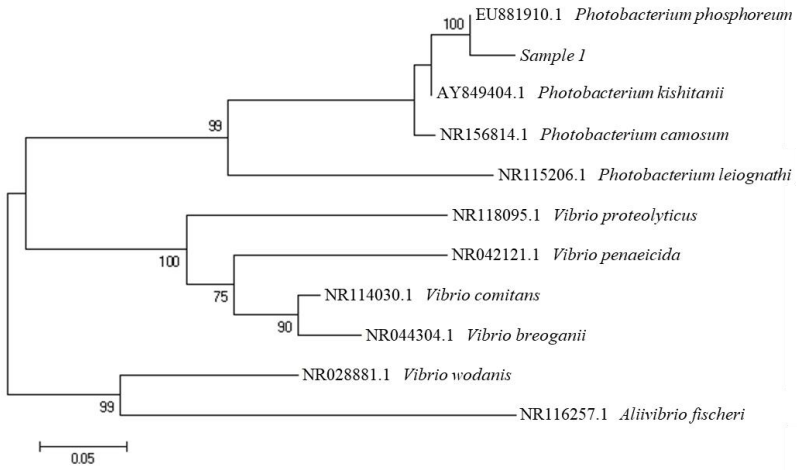
146
147 (a) (b) (c) (d)

148 **Figure 1**. Growth morphology of the strain. (a) Colony morphology on solid medium; (b) Colony
149 morphology on solid medium under dark conditions; (c) Colony morphology on liquid culture
150 medium under dark conditions; (d) Morphology of strains under oil microscope.

151 3.1.2. 16S rRNA analysis

152 Afte the 16S rRNA sequence was compared with NCBI database by Blast. The
153 results showed that strain FJ21 belonged to Proteobacteria, Gammaproteobacteria,
154 Vibrionales, Vibrionaceae and *Photobacterium*. Based on the phylogeny of 16S rRNA gene
155 sequence, FJ21clustered with the strain *Photobacterium phosphoreum and Photobacterium*
156 *kishitanii*.(Fig 2).

157
158
159
160
161
162
163
164
165
166
167
168
169
170



171

172 **Figure 2.** Neighbour Joining phylogenetic tree based on 16S rRNA gene sequences.

173 3.1.3. Physicochemical properties of the protein encoded by *lux* gene

174 The sequencing results showed that *lux*C(1437bp), *lux*D(921bp), *lux*A(1074bp),
175 *lux*B(987bp), *lux*F (696bp), *lux*E (1122bp), *lux*G(705bp) genes existed in strain FJ21.
176 Online software Expasy (https://www.expasy.org/) was used to predict the
177 physicochemical properties of proteins encoded by *lux*C, *lux*D, *lux*A, *lux*B, *lux*F, *lux*E,
178 and *lux*G genes (Table 1).

179 **Table 1.** Prediction of physical and chemical properties of the protein encoded by *lux* genes.

	*lux*C	*lux*D	*lux*A	*lux*B	*lux*F	*lux*E	*lux*G

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	77155or76780	34045or33920	47245or46870	30620or30370	29005or28880	47705or47330	31775or31400
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 the *lux*F 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.

3.1.4. 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 3). The amino acid sequence of the pink part was α-helix, and the amino acid sequence of the yellow part was β-sheet. SWISS-MODEL database (http://swissmodel.expasy.org/repository/) was used to predict the tertiary structure of proteins (Figure 4). 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 the *lux* G 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 α subunit and β subunit of the luciferase encoded by *lux*A and *lux*B genes have β-sheet 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

203 significance for studying functional structures (such as molecular docking).

204

205 **Figure 3.** Prediction of protein secondary structure encoded by *lux* genes. The amino acid
206 sequence of the pink part was α-helix, and the amino acid sequence of the yellow part was β-sheet.

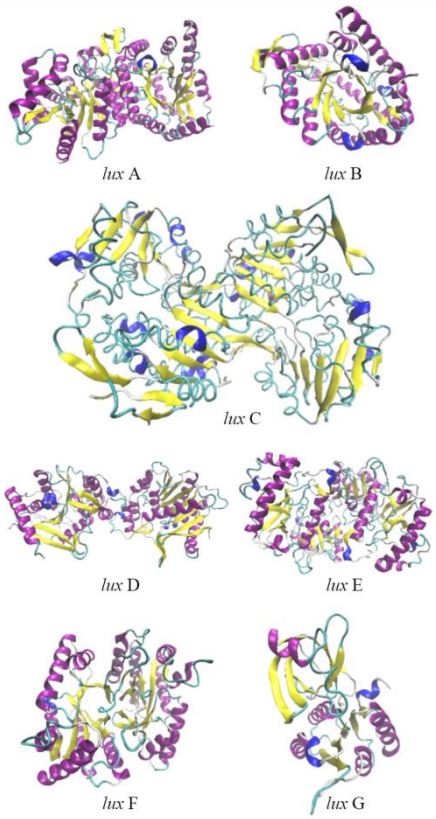


Figure 4. Prediction of protein tertiary structure encoded by *lux* genes, including *lux*A, *lux*B, *lux*C, *lux*D, *lux*E, *lux*F, *lux*G, α-helix (purple), β-sheet(yellow), turn (blue) and random coil(green).

3.2. Genome Sequencing and Assembly

The sequencing data are shown in Supplementary Table S1 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 % (Supplementary Table S2). The genome circle is shown in Figure 5. The genome sequence of strain FJ21 has been submitted to the GenBank database with accession number SRX10356131.

Table 2. Assembly result statistics.

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

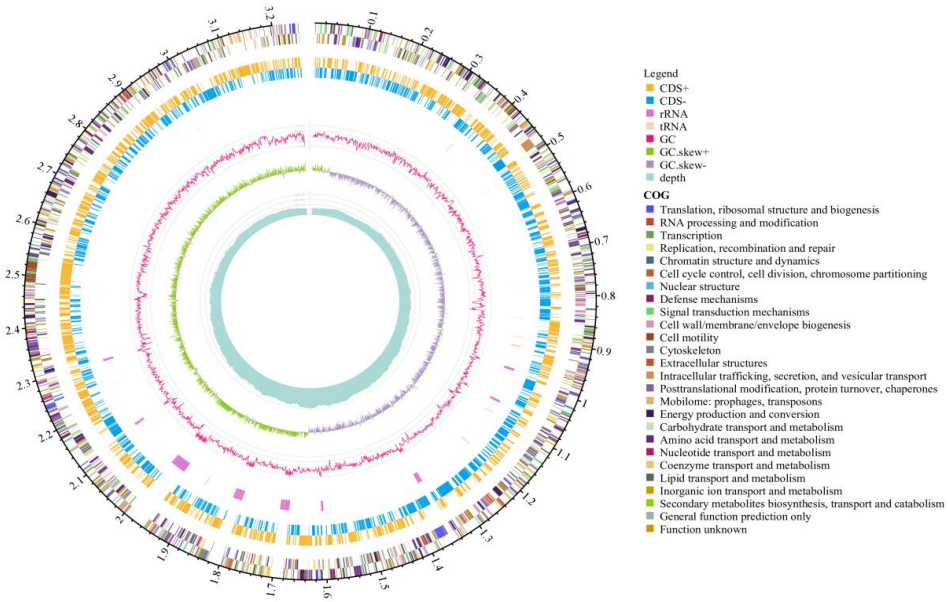


Figure 5. Genome circle diagram of strain FJ21. The characteristics of the marker are shown from outside to inside as follows: colored on clusters of orthologous groups (COG) functional categories, on the forward strand, coding sequences (CDSs); tRNA and rRNA; GC content (outward plots as positive values and inward plots as negative values); GC skew (G - /G + C, the leading chain and the lagging chain can be judged by the change of GC skew, generally the leading chain GC skew > 0, the lagging chain GC skew < 0) and Sequencing depth.

3.3. 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 Supplementary Table S3.

There are 1,769, 3,141, 2472, 4070, 3514, and 1413 genes that were annotated respectively relatethe d to KEGG pathway, COG category, GO, Refseq, Pfam and TIGRFAMs databases, and annotated results are shown in Supplementary Table S4.

3.3.1. COG function classification

In the COG category(Figure 6), there are 232 Energy production and conversion, 328 Amino acid transport, and metabolism, 210 Carbohydrate Transport and metabolism, 265 Translation, ribosomal structure and biogenesis, 266 Transcription, 267 Cell wall/membrane/envelope biogenesis and 194 Inorganic ion transport and metabolism. 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[35]. 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.

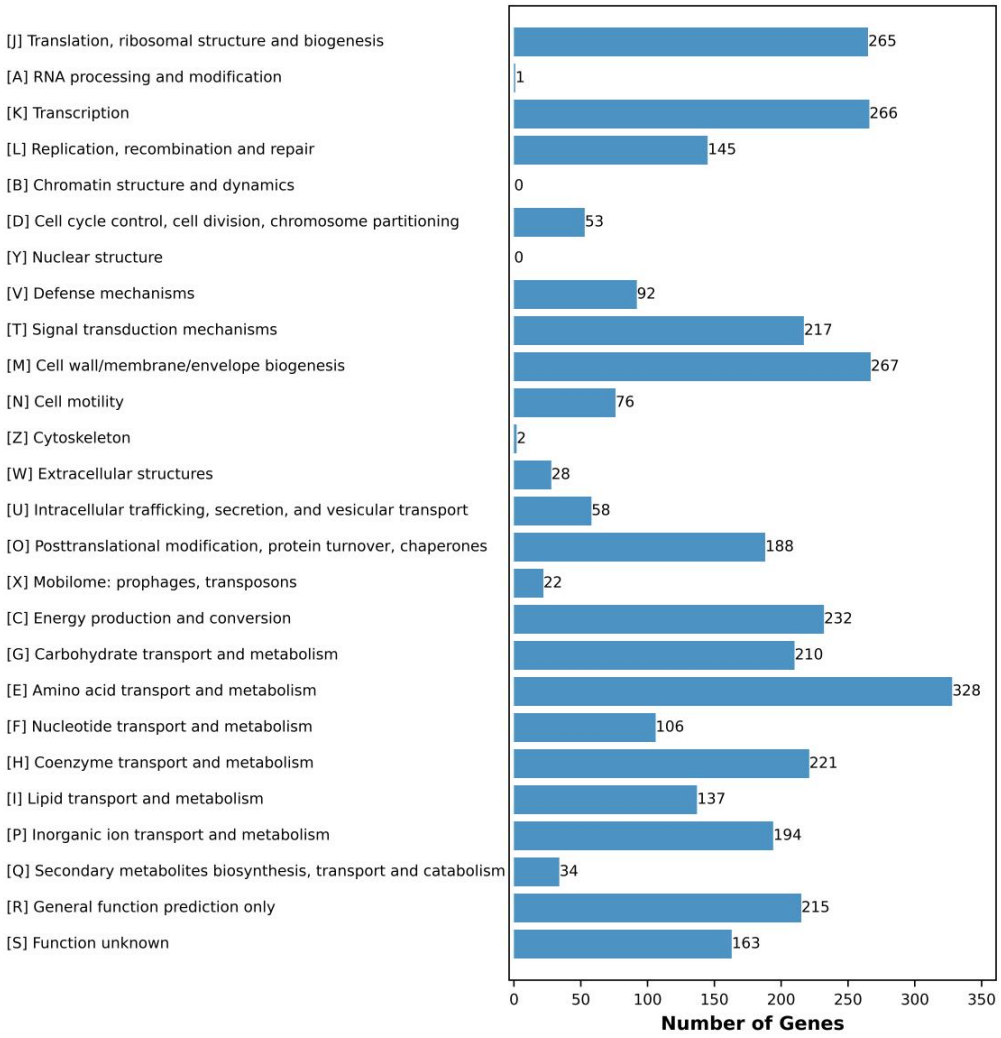


Figure 6. COG functional classification of strain FJ21, including Translation, ribosomal structure and biogenesis, RNA processing and modification, Transcription, Replication, recombination and repair, Cell cycle control, cell division, chromosome partitioning, Defense mechanisms, Signal transduction mechanisms, Cell wall/membrane/envelope biogenesis, Cell motility, Cytoskeleton, Extracellular structures, Intracellular trafficking, secretion, and vesicular transport, Posttranslational modification, protein turnover, chaperones, Mobilome: prophages, transposons, Energy production and conversion, Carbohydrate Transport and metabolism, Amino acid transport and metabolism, Nucleotide transport and metabolism, Coenzyme transport and metabolism, Lipid transport and metabolism, Inorganic ion transport and metabolism, Secondary metabolites biosynthesis, transport and catabolism.

3.3.2. GO function classification

As shown in Figure 7, 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[36]. 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, and 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.

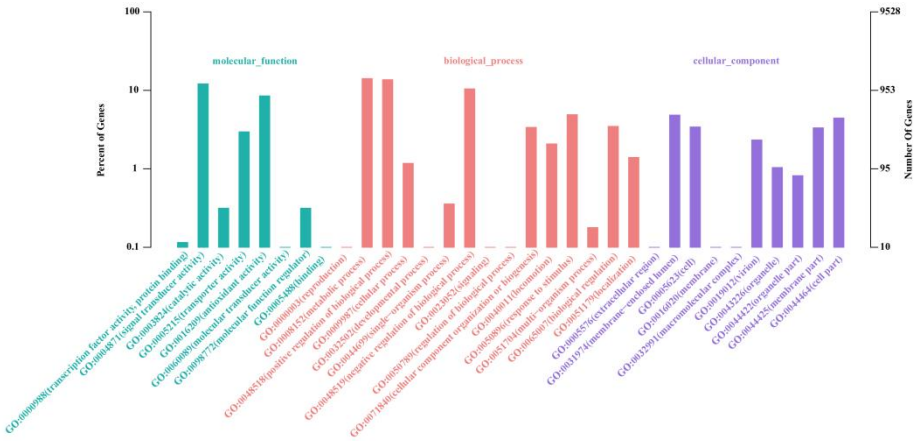


Figure 7. Go function classification diagram. There are three main sections: molecular function(including transcription factor activity, protein binding, signal transducer activity, catalytic activity, transporter activity, antioxidant activity, molecular transducer activity, molecular function regulator, binding), biological process(including reproduction, metabolic process, positive regulation of biological process, cellular process, developmental process, single-organism process, negative regulation of the biological process, signaling, regulation of biological process, cellular component organization or biogenesis, locomotion, response to stimulus, multi-organism process, biological regulation, localization) and cellular component(including extracellular region, membrane-enclosed lumen, cell, membrane, macromolecular complex, virion, organelle, organelle part, membrane part, cell part)

3.3.3. KEGG pathway analysis

The 2945 genes in the KEGG pathway were enriched in 208 metabolic pathways (figure 8), 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.

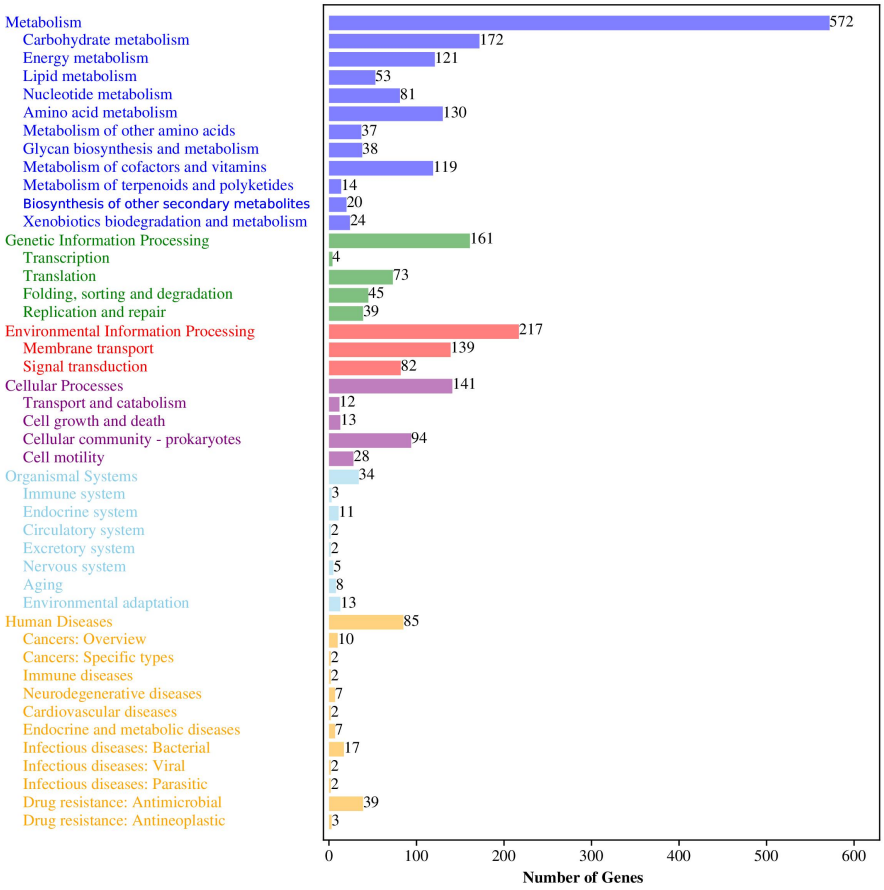


Figure 8. KEGG function classification diagram, including metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases; the number of genes (right).

3.3.4. Analysis of interaction genes between pathogen and host

Most luminous bacteria are nonpathogenic[38], while two subspecies of *Vibrio harveyi* and *Photobacterium damselae* are pathogens of many aquatic organisms[37,38]. 870 genes were annotated in the PHI database (Figure 9), 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 genand es, 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.

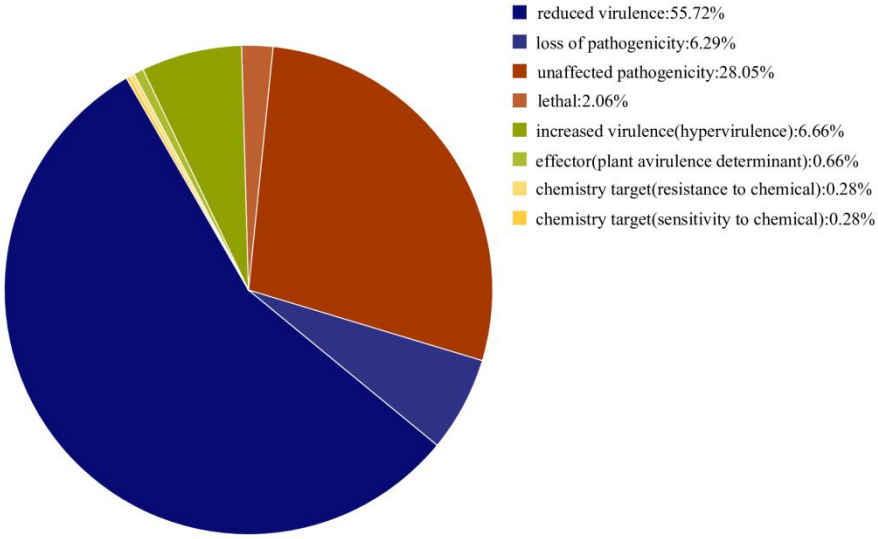


Figure 9. Gene analysis of pathogen host interaction. From left (blue) to right (yellow) as follows: reduced virulence genes; loss of pathogenicity genes; unaffected pathogenicity genes; lethal genes; increased virulence genes; effector genes; chemistry target (resistance to chemical) and chemistry target (sensitivity to chemical) genes.

3.3.5. Annotation of Resistance Genes in the CARD Database

The antibiotic resistance genes were annotated using the CARD database, and the information was shown in Table 3. 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 (Supplementary Table S5).

Table 3. Antibiotic resistance gene annotation of strain FJ21.

ARO	%Identities	Drug Class	Resistance Mechanism	AMR Gene Family
rsmA	89.29	A fluoroquinolone antibiotic, diaminopyrimidine antibiotic, phenicol antibiotic	antibiotic efflux	resistance-nodulation-cell division (RND)
vanWB	100	glycopeptide antibiotic	antibiotic target Alteration	antibiotic efflux pumpvanW, 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

3.3.6. 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)[39].

In this database, strain 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.

3.4. Phylogenetic tree, ANI and Tetra analysis

3.4.1. Genome phylogenetic tree analysis

Then the whole genome sequence(Supplementary Table S6) was constructed a phylogenetic tree (Figure 10). The results showed that *Photobacterium kishitanii* ATCCBAA-1194, *Photobacterium phosphoreum* JCM21184, *Photobacterium aquimaris* LC2-065, *Photobacterium malacitanum* CECT9190, and *Photobacterium carnosum* TMW 2.2021 were clustered together in the phylogenetic tree. We found that the whole genome sequence of strain FJ21 showed a great deal of similarity with the genome of *Photobacterium kishitanii* ATCCBAA-1194.

The difference between the two results may be due to the effect of gene transfer.

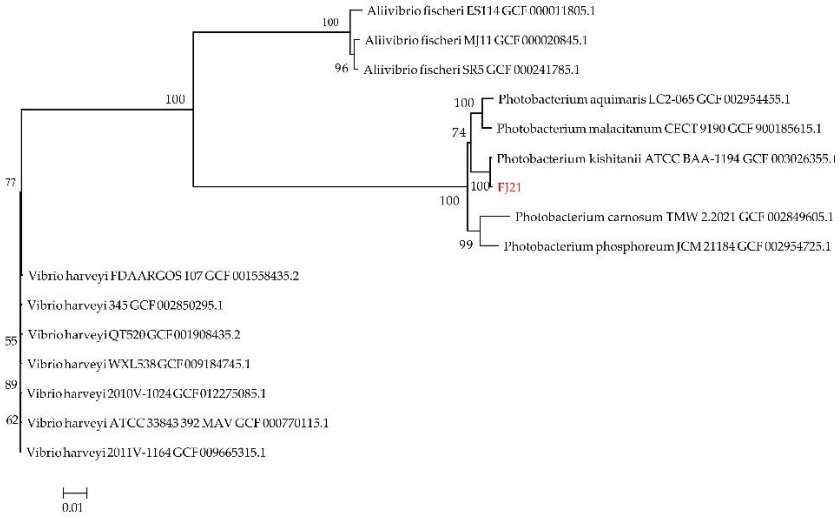


Figure 10. Phylogenetic tree based on whole-genome genes sequences.

3.4.2. ANI and TETRA analysis

Average nucleotide identity (ANI) and tetra nucleotide signatures (TETRA) between strain FJ21 and different *Photobacterium* strains were calculated (Table 4). The ANI value of strain FJ21 against *Photobacterium kishitanii* were 97.61% (ANIb, based on BLAST) and 97.55% (ANIm, based on MUMmer), respectively. Both were higher than the defined threshold (95%). In contrast, against the others, the ANI value was down to 84.68-87.32%, indicating strain FJ21 was phylogenetically close to *Photobacterium*

kishitanii. Therefore, strain FJ21 should be reclassified as *Photobacterium kishitanii* rather than *Vibrio fischeri* or *Photobacterium phosphoreum*. The results of Tetra also supported the conclusion.

Table 4 ANI and tetra nucleotide signatures (TETRA) analyses between strain FJ21 and *Photobacterium* strains.

The asterisk represented the result of self-comparison.

Species name	FJ21			*P.kishitanii*			*P.phosphoreum*			*P.aquimaris*			*P.malacitanum*			*P.carnosum*		
	ANIb	ANIm	Tetra	ANIb	ANIm	Tetra	ANIb	ANIm	Tetra	ANIb	ANIm	Tetra	ANIb	ANIm	Tetra	ANIb	ANIm	Tetra
FJ21	*	*	*	97.55	97.81	0.998	85.68	87.32	0.985	85.03	87.17	0.993	84.72	86.90	0.992	84.62	86.68	0.988
P.kishitanii	97.61	97.80	0.998	*	*	*	85.41	87.32	0.988	84.82	87.13	0.994	84.41	86.86	0.993	84.23	86.72	0.990
P.phosphoreum	85.64	87.32	0.985	85.64	87.32	0.988	*	*	*	84.89	86.86	0.991	84.48	86.57	0.990	85.66	87.69	0.992
P.aquimaris	85.11	87.17	0.993	85.16	87.13	0.994	85.03	86.86	0.991	*	*	*	92.61	92.97	0.998	84.10	86.87	0.989
P.malacitanum	84.68	86.9	0.992	84.62	86.86	0.993	84.56	86.57	0.990	92.44	92.96	0.998	*	*	*	83.23	86.09	0.986
P.carnosum	84.68	86.69	0.988	84.65	86.72	0.990	86.07	87.69	0.992	84.26	86.88	0.989	83.42	86.10	0.986	*	*	*

3.5. Genome collinearity analysis

3.5.1. 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 5). 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.

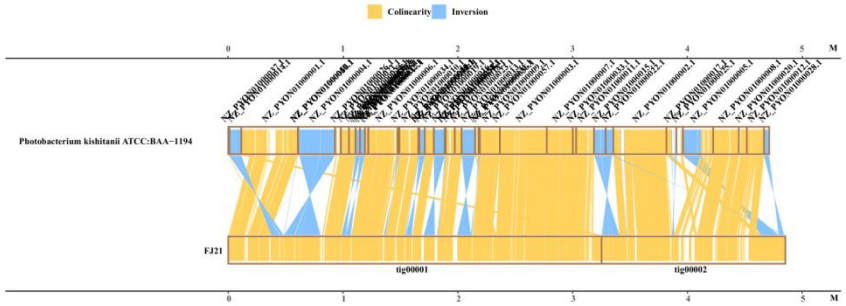
Table 5. Genome features of *P.kishitanii*, *P.phosphoreum*, *P.aquimaris*, *P.malacitanum*, *P.carnosum* and strain FJ21.

Species name	Genome size (bp)	Coding gene	tRNAs	5S、16S、23S rRNA
P.kishitanii	4732354	4087	159	6, 6, 5
P.phosphoreum	4550107	3840	73	3, 0, 1
P.aquimaris	4525475	3826	83	1, 0, 1
P.malacitanum	4380538	3739	186	10, 15, 12
P.carnosum	4559543	3984	122	8, 1, 8
FJ21	4853277	4131	226	24, 22, 22

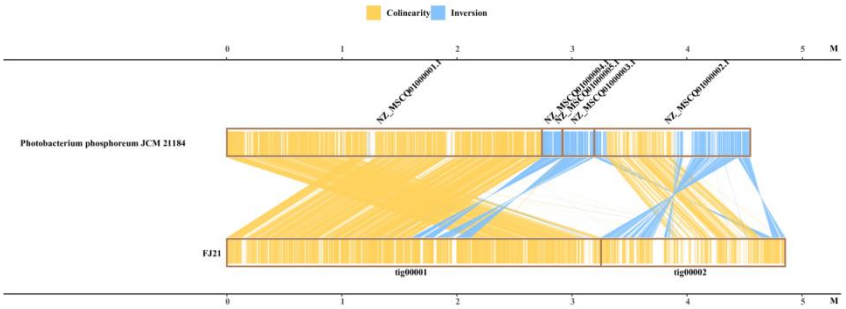
3.5.2. Collinearity analysis

MUMmer (version 3.23) software was used to compare the strain FJ21 with *Photobacterium kishitanii* ATCCBAA-1194, *Photobacterium phosphoreum* JCM21184, *Photobacterium aquimaris* LC2-065, *Photobacterium malacitanum* CECT9190, and *Photobacterium carnosum* TMW 2.2021. The collinearity and structural variation of genomic sequences are shown in Figure 11, and there were 218, 744, 717, 748, and 708 contrast blocks between *Photobacterium kishitanii* ATCCBAA-1194, *Photobacterium phosphoreum* JCM21184, *Photobacterium aquimaris* LC2-065, *Photobacterium malacitanum*

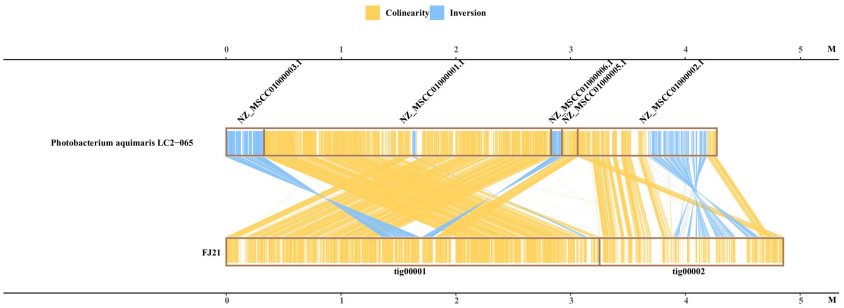
CECT9190, and *Photobacterium carnosum* TMW 2.2021 and the strain FJ21, respectively. They accounted for 86.92 %, 70.27 %, 64.82 %, 65.16 % and 56.32 % of the genome of the strain, respectively. According to the results, the collinearity between genomes is good, but there are a small number of genome rearrangement events such as inversion and translocation. It can be seen that the six strains still have great differences in evolution.



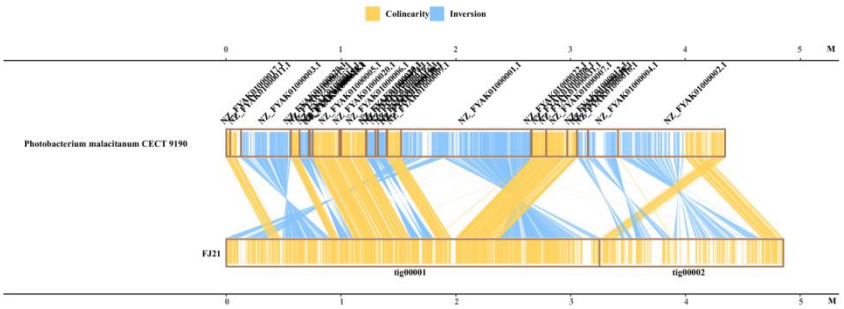
A



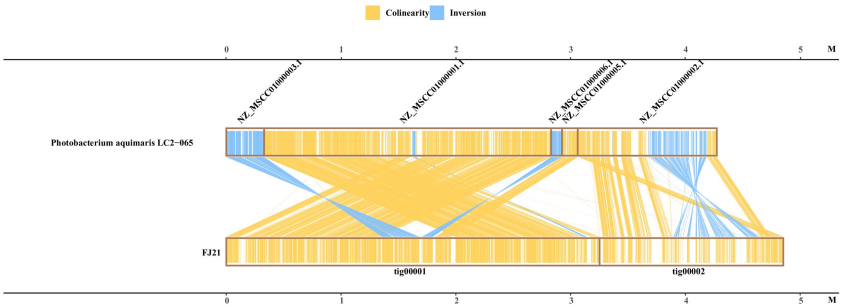
B



C



D



E

Figure 11. Collinearity analysis of strain FJ21, *P. kishitanii*, *P. phosphoreum*, *P. aquimaris*, *P. malacitanum* and *P. carnosum*. (A) Collinearity analysis of *P. kishitanii* and strain FJ21, (B) Collinearity analysis of *P. phosphoreum* and strain FJ21, (C) Collinearity analysis of *P. aquimaris* and strain FJ21, (D) Collinearity analysis of *P. malacitanum* and strain FJ21, (E) Collinearity analysis of *P. carnosum* and strain FJ21.

3.6. 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, beta lactone, and arylpolyene) were predicted in the FJ21 genome(Table 6), and siderophore only exists in *Photobacterium phosphoreum*.

Table 6. Gene clusters of a secondary metabolite of *Photobacterium kishitanii* ATCCBAA-1194, *Photobacterium phosphoreum* JCM21184, *Photobacterium aquimaris* LC2-065, *Photobacterium malacitanum* CECT9190, and *Photobacterium carnosum* TMW 2.2021 and strain FJ21.

Species name	Cluster type	Start gene and End gene
FJ21	RiPP-like	01645-01653
	Betalactone	01929-01947
	Arylpolyene	03136-03175
P. kishitanii	Betalactone	58033-83578
	Arylpolyene	6204-49845
	RiPP-like	53445-64308
P. phosphoreum	siderophore	607412-619838
	Arylpolyene	1098069-1141719
	Betalactone	66857-92412
	RiPP-like	52861-63724
P. aquimaris	Betalactone	33216-58753
	Arylpolyene	203830-247492
	RiPP-like	52780-63643
P. malacitanum	Arylpolyene	884620-928282
	RiPP-like	53563-64426
	Betalactone	33744-59282
P. carnosum	Betalactone	32103-56200
	RiPP-like	25713-36576

3.7. For detection of heavy metal toxicity

The fitted concentration-response curve of the single toxicity of heavy metal to the strain FJ21 is shown in the figure 12, and the parameter values of concentration-dose effect curve are shown in the table 7. With the increase of heavy metal concentration, the inhibition rate of bacteria's luminescence gradually increased. R^2 values are all greater

than 0.98, which indicates that the concentration of heavy metals is positively related to its inhibition rate of luminescent bacteria. The EC_{50} value can reflect the toxicity of pollutants. The smaller the EC50 value, the greater the toxicity. According to the table, the EC_{50} values of $Pb(NO_3)_2$, $ZnSO_4{\cdot}7H_2O$, $CdCl_2{\cdot}2.5H_2O$, $CuSO_4{\cdot}5H_2O$ and $K_2Cr_2O_7$ to luminescent bacteria are 1.11mg/L, 1.57mg/L, 28.63mg/L, 77.35mg/L and 201.80mg/L, respectively. The toxicity is $Pb(NO_3)_2$ > $ZnSO_4{\cdot}7H_2O$ > $CdCl_2{\cdot}2.5H_2O$ > $CuSO_4{\cdot}5H_2O$ > $K_2Cr_2O_7$.

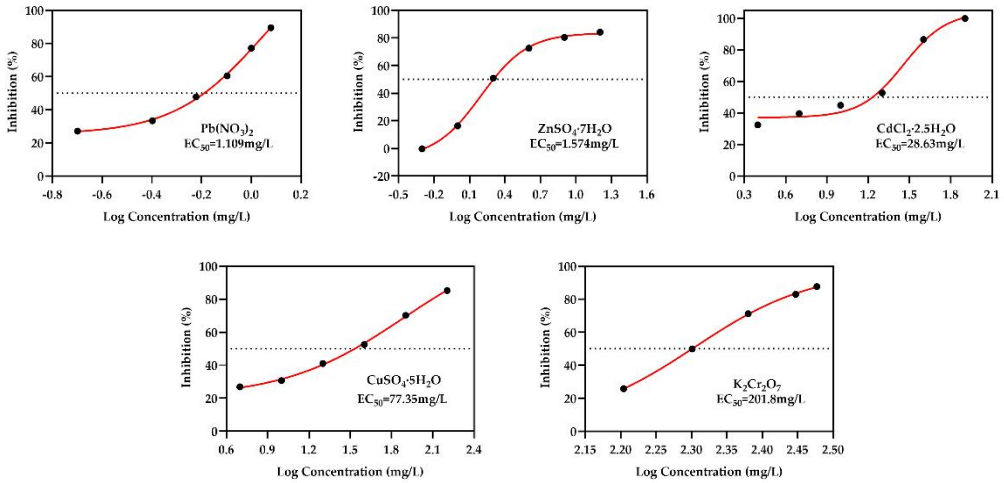


Figure 12. The fitted concentration-response curve of the single toxicity of heavy metal to the strain FJ21. The abscissa represents the logarithm of heavy metal concentration, and the ordinate represents the luminescence inhibition rate of the strain FJ21.

Table 7. parameter values of concentration-dose effect curve.

Heavy metal	EC_{50}(mg/L)	Ion mass concentration	R^2
$ZnSO_4{\cdot}7H_2O$	1.57	0.36	0.9994
$CuSO_4{\cdot}5H_2O$	77.35	19.80	0.9988
$Pb(NO_3)_2$	1.11	14.09	0.9962
$CdCl_2{\cdot}2H_2O$	28.63	0.69	0.9848
$K_2Cr_2O_7$	201.80	35.67	0.9992

4. Conclusions

The colony shape is round, slightly raised, smooth, milky white and short rod, which emits bright blue-green light in the dark. Based on 16S rRNA the strain FJ21 was phylogenetically close to the strain *Photobacterium phosphoreum and Photobacterium kishitanii.* While the whole genome, ANI, TETRA analyses and collinearity analysis further verified the relationship between strain FJ21 and the species of *Photobacterium kishitanii* ATCCBAA-1194. Therefore, we argue that strain FJ21 should be classified as a strain of *Photobacterium kishitanii*. This indicates that the whole genome analytic method is very important for species identification.

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 the *lux* gene and its encoded protein, the strain contained *lux*C, *lux*D, *lux*A, *lux*B, *lux*F, *lux*E, and *lux*G genes. However, the function of *lux*F gene is still uncertain. Understanding the *lux* genes will help to understand luminescent activities and the mechanism.

In the toxicity test, the toxicity of heavy metals to strain FJ21 is as follows: $Pb(NO_3)_2$ > $ZnSO_4{\cdot}7H_2O$ > $CdCl_2{\cdot}2.5H_2O$ > $CuSO_4{\cdot}5H_2O$ > $K_2Cr_2O_7$.

Supplementary Materials: Table S1: Sequencing data statistics. Table S2: Assembly results statistics. Table S3: Genome structure prediction statistics. Table S4: Functional annotation statistics of genome-encoded proteins. S5: Annotation of Resistance Genes in the CARD database. Table S6: The list of genomes used in the study. Word: the strain FJ21 genomic.

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

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

Acknowledgments: Authors would like to thank the Hunan Provincial Engineering Research Center of Applied Microbial Resources Development for Livestock and Poultry for technical supporting.

Conflicts of Interest: The authors declare that they have no competing interests in this paper.

References

1. Brodl, E.; Winkler, A.; Macheroux, P. Molecular Mechanisms of Bacterial Bioluminescence. *Comput. Struct. Biotechnol. J.* **2018**, *16*, 551-564. [PubMed]
2. Gong, L.; Wu, Y.; Jian, Q.J.; Yin, C.X.; Li, T.T.; Gupta, V.K.; et al. Complete genome sequencing of the luminescent bacterium, *Vibrio qinghaiensis* sp. Q67 using PacBio technology. *Scientific. Data*. **2018**, 170205
3. Zavilgelsky, G.B.; Shakulov, R.S. Mechanisms and Origin of Bacterial Bioluminescence. *Mol Biol)*. **2018**, 52, 935-947. [PubMed]
4. Brodl, E.; Niederhauser, J.; Macheroux, P. In Situ Measurement and Correlation of Cell Density and Light Emission of Bioluminescent Bacteria. *J. Vis. Exp.* **2018**, 57881. [PubMed]
5. Alloush, H.M.; Lewis, R.J.; Salisbury, V.C. Bacterial Bioluminescent Biosensors: Applications in Food and Environmental Monitoring. *Analytical. Letters*. **2006**, 39, 1517–1526. [PubMed]
6. 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.* **2008**, *61*, 11-17. [PubMed]
7. Ryu, H.S.; Kim, H.K.; Choi, W.C.; Kim, M.H.; Park, S.Y.; Han, N.S.; Oh, T.K.; Lee, J.K. New cold-adapted lipase from Photobacterium lipolyticum sp. nov. that is closely related to filamentous fungal lipases. *Appl. Microbiol. Biotechnol*. **2006**, 70, 321-326. [PubMed]
8. Yaacob, M.A.; Hasan, W.A.; Ali, M.S.; Rahman, R.N.; Salleh, A.B.; Basri, M.; Leow, T.C. Characterisation and molecular dynamic simulations of J15 asparaginase from Photobacterium sp. strain J15. *Acta. Biochim. Pol.* **2014**, *61*, 745-752. [PubMed]
9. Shakiba, M. H.; Ali, M. S.; Rahman, R. N.; Salleh, A. B.; & Leow, T. C. Cloning, expression and characterization of a novel cold-adapted GDSL family esterase from Photobacterium sp. strain J15. *Extremophiles*. **2016**, *20*, 44–55. [PubMed]
10. Girotti, S.; Ferri, E.N.; Fumo, M.G.; Maiolini, E. Monitoring of environmental pollutants by bioluminescent bacteria. *Anal. Chim .Acta*. **2008**, *608*, 2-29. [PubMed]
11. Adnan, N.A.; Halmi, M.I.E.; Abd Gani, S.S.; Zaidan, U.H.; Abd Shukor, M.Y. 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*, 6644. [PubMed]
12. Zhu, W.J.; Xu, Y.T.; Zhang, Q.Z.; Yang, J.; Ding, G.R.; Wang, X.Y. Application and development of monitoring with luminous bacteria for environmental pollutants. *Water. Purification. Technolo*. **2010**, 29, 54-59.
13. Li, Z.X.; Yin, Q.; Zhao, J.N.; Li, G.X.; Liu, Y.F.; Li, W. Advance in the application of luminescent bacteria on detecting acute toxicity of water. *Hebei University of Science and Technology*. **2014**, 35, 480-486.
14. Senol Cali, D.; Kim, J.S.; Ghose, S.; Alkan, C.; Mutlu, O. Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions. *Brief. Bioinform*. **2019**, *20*, 1542–1559. [PubMed]
15. Walker, B. J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; Earl, A. M. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one*, . **2014**, *9*, e112963. [PubMed]
16. Hu, J.; Fan, J.; Sun, Z.; Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*. **2020**, *36*, 2253–2255. [PubMed]
17. Hunt, M.; Silva, N.D.; Otto, T. D.; Parkhill, J.; Keane, J.A.; Harris, S.R. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome. biology*. **2015**, *16*, 294. [PubMed]
18. Hyatt, D.; Chen, G.L.; Locascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC. Bioinformatics*. **2010**, *11*, 119. [PubMed]

19. Lagesen, K.; Hallin, P.; Rødland, E.A.; Staerfeldt, H. H.; Rognes, T.; Ussery, D.W. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic. Acids. Res.* **2007**, *35*, 3100–3108. [PubMed]
20. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. **1997**, *25*, 955–964. [PubMed]
21. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. **2013**, *29*, 2933–2935. [PubMed]
22. Kalvari, I.; Argasinska, J.; Quinones-Olvera, N.; Nawrocki, E.P.; Rivas, E.; Eddy, S.R.; Bateman, A.; Finn, R.D.; Petrov, A. I. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic. Acids. Research*. **2018**, *46*, D335–D342. [PubMed]
23. Hudson, C.M.; Lau, B.Y.; Williams, K.P. Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic. Acids. Res*. **2015**, *43*, D48–D53. [PubMed]
24. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; Pesseat, S.; Quinn, A.F.; Sangrador-Vegas, A.; Scheremetjew, M.; Yong, S.Y.; Lopez, R.; Hunter, S. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. **2014**, *30*, 1236–1240. [PubMed]
25. Haft, D.H.; Selengut, J.D.; Richter, R.A.; Harkins, D.; Basu, M.K.; Beck, E. TIGRFAMs and Genome Properties in 2013. *Nucleic. Acids. Res.* **2013**, *41*, D387–D395. [PubMed]
26. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; Salazar, G.A.; Tate, J.; Bateman, A. The Pfam protein families database: towards a more sustainable future. *Nucleic. Acids. Res.* **2016**, *44*, D279–D285. [PubMed]
27. Harris, M. A. ; Clark, J.; Ireland, A.; Lomax, J.; Ashburner, M.; Foulger, R.; Eilbeck, K.; Lewis, S.; Marshall, B.; Mungall, C.; Richter, J.; Rubin, G. M.; Blake, J. A.; Bult, C.; Dolan, M.; Drabkin, H.; Eppig, J.T.; Hill, D.P.; Ni, L.; Ringwald, M. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic. Acids. Res.* **2004**, *32*, D258–D261. [PubMed]
28. Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic. Acids. Res.* **2000**, *28*, 27–30. [PubMed]
29. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C.M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina, W.; Joardar, V.S.; Kodali, V.K.; Li, W.; Maglott, D.; Masterson, P.; McGarvey, K.M.; Murphy, M.R.; O'Neill, K.; Pujar, S.; Rangwala, S.H.; Rausch, D.; Riddick, L.D.; Schoch, C.; Shkeda, A.; Storz , S.S.; Sun, H.; Thibaud-Nissen, F.;, Tolstoy, I.; Tully, R.E.; Vatsan, A.R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M.J.; Kimchi, A ; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T.D.; Pruitt, K.D. Reference sequence(RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic. Acids. Res.* **2016**, 44, D733-745. [PubMed]
30. Galperin, M.Y.; Makarova, K.S.; Wolf, Y.I.; Koonin, E.V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic. Acids. Research*. **2015**, *43*, D261–269. [PubMed]
31. Parks, D.H.; Chuvochina, M.; Chaumeil, P.A.; Rinke, C.; Mussig, A.J.; Hugenholtz, P. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* **2020**, *38*, 1079-1086. [PubMed]
32. Tindall, B.J.; Rossello-Mora, R.; Busse, H.J.; Ludwig, W.; Kampfer, P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* **2010,** *60*, 249-266. [PubMed]
33. Kim, M.; Oh, H.S.; Park, S.C.; Chun, J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 346-351. [PubMed]
34. Colston, S.M.; Fullmer, M.S.; Beka, L.; Lamy, B.; Gogarten, J.P.; Graf, J. Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using Aeromonas as a test case. *mBio.* **2014**, *5*, e02136. [PubMed]
35. Li, E.; Yang, H.; Wang, X.; Wan, L.; Pan, H.; Zhu, D. Whole-genome sequencing and analysis of inosine- producing strain Bacillus subtilis ATCC 13952. *Wei Sheng Wu Xue Bao.* **2015**, *55*, 1560-1567. [PubMed]
36. 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. [PubMed]
37. Moi, I.M.; Roslan, N.N.; Leow, A.; Ali, M.; Rahman, R.; Rahimpour, A.; Sabri, S. The biology and the importance of Photobacterium species. *Appl. Microbio.l Biotechnol*. **2017**, *101*, 4371–4385. [PubMed]
38. Ke, H.M.; Prachumwat, A.; Yu, C.P.; Yang, Y.T.; Promsri, S.; Liu, K.F.; Lo, C.F.; Lu, M.J.; Lai, M.C.; Tsai, I.J.; Li, W.H. Comparative genomics of Vibrio campbellii strains and core species of the Vibrio Harveyi clade. *Sci. Rep.* **2017**, *7*, 41394. [PubMed]
39. Zhang, B.Y.; Zhu, T.H.; Han, S.; Wang, Y.; Li, S.J; Qiao, T.M. Whole-genome sequencing and analysis of Streptomyces sampsonii KJ40. *Acta. Microbiologica. Sinica*. **2018**, 45, 805-818.