Bioinformatics Analysis of *Klebsiella pneumoniae* M1, from Microbial Flocculation Resources

Huan Diao,[a] Yan Xu,[a] Lv-mu Li,[b,*] and Qiang Tang [c,*]

The synthesis of a microbial flocculant is strictly controlled by its genetic genes, which may be the result of the expression of flocculant genes in the microbial genome. The whole genome of *Klebsiella pneumoniae* M1 was sequenced for obtaining the mechanism of flocculant synthesis and exploring the mechanism of flocculant production by flocculant bacteria. Therefore, it provided a basis for molecular genetics and functional genomic analysis of flocculants production bacteria. In this way, the mechanism of flocculant production by flocculating bacteria can be better explored. The whole genome sequence of flocculant strain M1 was determined using advanced second generation (Illumina) and third generation (PacBio) sequencing, which was screened from wheat alcohol wastewater. The genes related to flocculant characteristics produced by strain M1 were analyzed for combining with the analysis of flocculant structure characteristics. According to the *de novo* assembly, a total of 5,511,794 bp clean reads were generated and assembled into 24 contigs. The GC content was up to 58.39%. The genome contained approximately 5383 genes, but 5348 genes had obvious biological functions. A total of 437 genes were involved in carbohydrate metabolism and had coding genes of five carbohydrate-related enzymes. This result indicated that there were functional genes closely related to polysaccharide production in M1 genome. The main metabolic process of flocculant strain *Klebsiella pneumoniae M1* was closely related to the potential pathway of extracellular polysaccharide biosynthesis, in which five kinds of carbohydrate synthase genes were involved.

DOI: 10.15376/biores.17.3.4870-4885

Keywords: Klebsiella pneumoniae M1; Bioinformatics; Flocculation

Contact information: a: School of Pharmacy, Anhui Xinhua University, Hefei Anhui, 230088, China;
b: School of Animal Sciences and Technology, Anhui Agricultural University, Hefei, Anhui, 230036, China;
c: College of artificial intelligence and big data , Anhui Xinhua University, Hefei Anhui, 230088, China;
** Corresponding author: 39742727@qq.com;lilvmu@126.com*

INTRODUCTION

A flocculant agent is a substance used to promote flocculation by the formation of a floc of colloids and suspended particles in liquid. Flocculants are subdivided into traditional chemical flocculants and new microbial flocculants. A microbial flocculant (MBF) is a kind of metabolite produced by microorganisms or their secretions. It is a new type of high molecular biological agent. In recent years, close attention has been paid to microbial flocculants because of their advantages, including safety to humans, environmental friendliness, and acceptable removal performances. Microbial flocculants are macromolecular substances produced by microorganisms. Microbial flocculants are produced at specific stages of microbial growth. The flocculating activity is determined by specific bioactivators, such as proteins, polysaccharides, and glycoproteins (Diao *et al*.

2018, 2019). Microbial flocculants have been widely used in various industrial fields, such as wastewater treatment, microalgae harvest, activated sludge dewatering, heavy metal ion adsorption, and nanoparticle synthesis (Liu *et al.* 2021). Bioflocculant-producing bacteria can be screened from almost all types of terrestrial and aquatic environments such as soil, activated sludge, wastewater, and sea (Aida *et al*. 2022). The production of these flocculating active substances is mainly regulated by their genetic genes, and it is equal to the expression of specific flocculating genes. Therefore, the analysis of the whole genome of flocculant bacteria can explain the molecular mechanism of flocculation. Fu *et al.* (2016) analyzed the whole genome sequence of a flocculant-producing strain *Paenibacillus shenyangensis* by high-throughput sequencing technology. The total length of the genome was 5501467 bp; 4800 single genes were predicted. A total of 4393 annotations showed specific gene functions in NCBI NR database, and 3423 genes were found in the database of homologous lineage groups. Cell growth and metabolism were the main biological processes. Potential metabolic pathways from glucose to polysaccharides were starch and sucrose metabolic pathways. He (2012) studied the flocculating genetic background of flocculating yeast and isolated flocculating genes, then FL01, FL05, FL08, FL09, FL010 and FL011 strains were identified as flocculating genes of flocculant yeast. Sequence analysis was carried out to elucidate molecular mechanism of flocculation of the microbial flocculant. Chen *et al.* (2012) preliminarily confirmed that the flocculant gene of Panbacteria MBF03 strain was located on its chromosome, and the main component of flocculant was acidic extracellular polysaccharide by plasmid transformation and plasmid elimination tests. Xue (2013) constructed the recombinant expression vector by constructing the flocculating genetic engineering bacteria, and determined its flocculating activity. Compared with the original strain, the flocculating abilities of the recombinant expression vectors were increased by 1.48 and 2.85 times respectively. Wei *et al*. (2017) reported the whole genome sequence of *Microbacterium paludicola* CC3 producing polysaccharide bioflocculant. A series of genes encoding polysaccharide biosynthesis and modification proteins were found, and transporter genes were detected, which could promote polysaccharide biosynthesis. Therefore, the sequencing of the whole genome of flocculants can lay a foundation for further research on genetic information of biosynthesis.

Klebsiella is an important microbial flocculant-producing bacteria, and *Klebsiella* sp. is famous for its rich capsule in Gram-negative bacilli. At first, Li Xul (2007) screened two highly effective *Klebsiella* flocculants in 2007 for the application of *Klebsiella* in the field of flocculants, and their flocculants were distributed in extracellular fermentation broth. *Klebsiella* NIII2 with high glycoprotein yield was isolated from activated sludge (Ghosh *et al*. 2009; Guo *et al*. 2013; Xing 2013). The mutant strains NI-II2-1 and NIII2-2 were obtained by ultraviolet and chemical mutagenesis respectively. Adding appropriate amount of nitrogen source (sodium nitrate) to fermentation medium could promote flocculant production of NIII2 and its mutant strain NIII2-2. The maximum yield was 7.5 g/L, which was higher than 5 g/L using urea as nitrogen source. Liu *et al.* (2014) screened and isolated *Klebsiella* C11 from activated sludge, and the best carbon source, nitrogen source, and inorganic salt were glucose, $NaNO_3$, and $MgSO_4$, respectively. Their flocculation activity was closely related to the decomposable utilization characteristics of carbon source. Nitrate nitrogen was more easily utilized by *Klebsiella* than ammonium nitrogen. The latest report on *Klebsiella pneumoniae* showed that the microbial flocculant produced by *Klebsiella pneumoniae* had an effective flocculation effect on carbamazepine production wastewater (Xing *et al*. 2020). There are a few reports about *Klebsiella* strain

as a flocculant, but the whole genome sequence of *Klebsiella* is focused on its pathogenicity and drug resistance. The complete genome sequence of several *Klebsiella pneumoniae* strains has been completed, including the first six *Klebsiella pneumoniae* strains HS11286, NTUH-K2044, MGH78578, 1084, KCTC2242, and 342 strains. In addition, Zhu *et al.* (2014) sequenced the whole genome of a *Klebsiella pneumoniae* JM45 strain with extensive drug resistance in 2014. Zhang (2013) completed the whole genome sequencing of a *Klebsiella pneumoniae* strain KG2 in 2013 by Solexa technology, which was widely used in high-throughput sequencing technology. Liu (2012) sequenced the whole genome of *Klebsiella pneumoniae* strain HS11286 in 2012, and identified or analyzed the mobile genetic elements by comparative genome analysis, which laid a foundation for the genetic diversity, pathogenesis, drug resistance, and horizontal transfer mechanism of the bacteria. Zeng (2012) adopted proteomics-related research methods, and the high-throughput surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) technology was tested the expression profile of *Klebsiella pneumoniae*. The study established a diagnostic model for strains of *Klebsiella pneumoniae* producing extended spectrum beta-lactamases (ESBLs). The test can quickly and conveniently detect such bacteria. It can be seen that the effective application of advanced methods and technologies in modern life science research can allow further understanding of the biological characteristics of pathogenic bacteria and the molecular mechanism of pathogenicity and immunity. However, there are few reports on the regulation of flocculation function genes on metabolic pathways. Therefore, a flocculant *Klebsiella pneumoniae M1* strain (Preservation Number CCTCC M 2018098, GenBank login number MG987011) with strong flocculating function was screened from sediment sludge in this study (Diao *et al*. 2018). The flocculant produced by the fermentation of this bacterium is mainly composed of polysaccharide and protein, with the contents of 65.9% and 19.74% respectively (Diao *et al*. 2019).

Whole genome sequencing (WGS) scans the whole genome, interprets all common and rare variation information within the whole genome, and comprehensively detects single nucleotide polymorphisms (SNPs), insertion deletion (indel), structure variation (SV), and copy number variation (CNV). Genome sequencing technology quickly obtains genetic information on organism DNA, which fully demonstrates the diversity and complexity of microbial genome (Metzker 2010). It plays a key role in the research of life science. In this study, the whole genome sequence of screening strain M1 was determined by the combination of the second generation (Illumina) and the third generation (PacBio) sequencing techniques, and the functional genes related to flocculation would be explored to analyzing the mechanism of extracellular polysaccharide synthesis of strain M1 at molecular level. The origin of flocculating genes and key enzymes were studied by analysis of various nutritional and metabolic variables, and the synthesis pathway of extracellular polysaccharides was preliminarily identified.

EXPERMENTAL

Strain Source

M1 strain was screened from sludge of wheat alcohol sewage sedimentation tank by Feed Biotechnology Laboratory of Anhui Agricultural University. In a previous study, it was confirmed that the flocculant had high flocculating activity for wheat alcohol wastewater.

Test Method

The whole genome of strain M1 was sequenced and entrusted to Shenzhen Huada Gene Co., Ltd. Using the latest three-generation sequencing technology, bacterial genome *de novo* sequencing, the bacterial genome was assembled from scratch after sequencing, and the genome components, functional annotations and comparative genomics were analyzed on the basis of assembly. By utilizing the latest high-throughput sequencing technology and powerful bioinformatics analysis capability, genome sequences of organisms can be efficiently determined and analyzed.

Bacterial genome *de novo* is the *de novo* assembly of bacterial genome after sequencing. On the basis of the assembly, genome components, functional annotation and comparative genomics are analyzed. The final assembly level is determined according to the needs of research and the characteristics of bacteria. According to the assembly level, it is divided into primary assembly, advanced assembly, and completion drawing assembly. The highest indicator is to complete the map assembly, that is, to assemble the complete sequence (chromosome and plasmid sequence information) in the bacterial genome. Bacterial *de novo* sequencing has replaced the traditional methods to become an important tool to study the genetic mechanism of bacterial evolution and key functional genes. It can be used for the identification of pathogenic genes, the study of intraspecific evolution, the transformation of engineering bacteria, the study of genetic theory model organisms, *etc*.

Sample Preparation

Strain M1 was cultured for 24 h and centrifuged at low temperature and high speed (8000 g/min) for 15 min under sterile conditions. Cells were collected and frozen for sampling.

Total DNA Extracting

Genomic DNA samples were extracted from cells, and bacterial DNA extraction kits were used to operate according to the instructions of the kits. A high quality DNA sample (A260/280=1.8 to 2.0) was obtained by fluorescence quantification of the total DNA extracted. A 200 to 300 bp fragment library was constructed.

Illumina Platform (Second Generation) Sequencing

After the DNA sample was received, the sample was tested. The next step was to construct the library with qualified samples: first, large DNA fragments were randomly broken by covaris or bioruptor, and a certain length of DNA fragments was generated. Then the sticky ends formed by breaking were repaired into flat ends by T4 DNA polymerase, Klenow DNA polymerase and T4 PNK. Then the base "a" was added at the 3'end to make the DNA fragments connect with the special connector with "t" base at the 3' end, the ligation products of the target fragments to be recovered were selected by electrophoresis, and then the DNA fragments with connectors at both ends were amplified by PCR. Finally, the qualified library was used for cluster preparation and sequencing.

PacBio Platform (Third Generations) Sequencing

Firstly, DNA was processed into appropriate size fragments by g-TUBE (a single-use device that shears genomic DNA into selected fragments sizes ranging from 6 kb to 20 kb). Then, fragments were repaired for damage and terminal repair; this was followed by joints, annealing, and finally sequencing. Both ends of the DNA fragments are connected to the joints of the hairpin structure to form a dumbbell structure, which is called

SMRTbell. The annealed SMRTbell is mixed with the polymerase at the bottom, which will be used for final sequencing.

Quality Shearing, Sequence Assembly, and Annotation of Original Sequencing Data

The predicted gene sequences were compared with functional databases such as COG, KEGG, GO, and NR to annotate gene functions for functional enrichment analysis.

Analysis of Common and Specific Genes

The gene sequences of *Klebsiella pneumoniae* strains (3 to 4 strains) were compared. The common gene and the specific gene (Pan gene) were analyzed between the strain M1 and representative strains, *K. pneumoniae* HS11286, *K. pneumoniae* MGH, 78578, and *K. pneumoniae* NTUHK2044, according to the homology comparison of NCBI.

RESULTS AND DISCUSSION

Genome Sequencing

Sequencing results showed that the total length of the M1 genome was 5,511,794 bp. The GC content was 58.39%, and it contained approximately 5383 genes. The GC distribution of specific genes was shown in Fig. 1. The abscissa was the GC content, and the ordinate was the average depth. For data with unbiased GC content, the scatter plot was Poisson distribution.

The results of PacBio Sequencing Platform data are shown in Fig 2. The left corner of the graph was a length distribution. Polymerase Reads length was distributed in the upper left corner of the graph, and Subreads Length was distributed in the lower left corner. The abscissa represented the Polymerase Reads/Subreads length, and the ordinate represented the Polymerase Reads/Subreads quantity. The Polymerase Reads mass distribution map was in the upper right corner, and the mass Subreads distribution map was in the lower right corner, the abscissa represented the Polymerase Reads/Subreads quality, and the ordinate represented the Polymerase Reads/Subreads quantity. Whether the number or distribution of Read was measured, the trend was reasonable, indicating that the sequencing results were reliable.

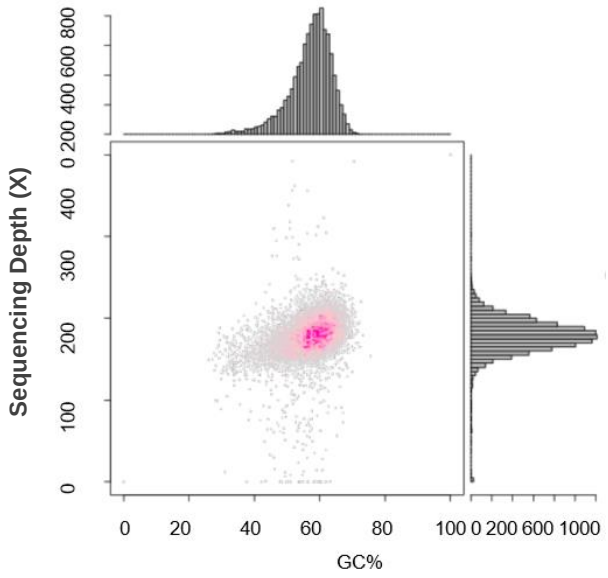

Fig. 1. GC% content distribution map

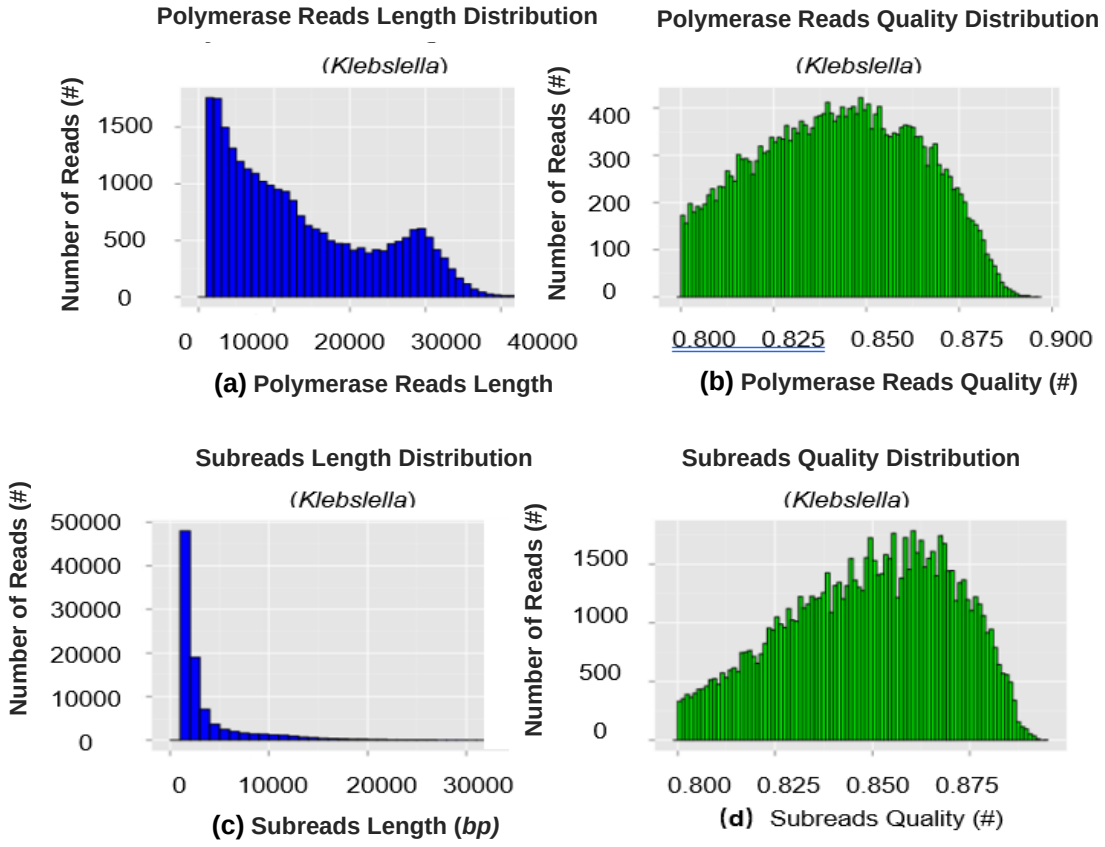

Fig. 2. Subreads data statistical chart

Prediction of Gene Components

Genes are the smallest DNA fragments with genetic effects. By analyzing the gene distribution on genome, the composition of functional genomes and the location of target genes of all types of sequencing strains was determined. As shown in Fig. 3, the M1 gene length of the strain was mostly concentrated between 100 and 1000 nt. Among them, the numbers of 100 to 200 nt and 900 to 1000 nt were more than 450 nt. There were not many genes between 1000 and 2000 nt length, but there were 261 genes far larger than 2000 nt.

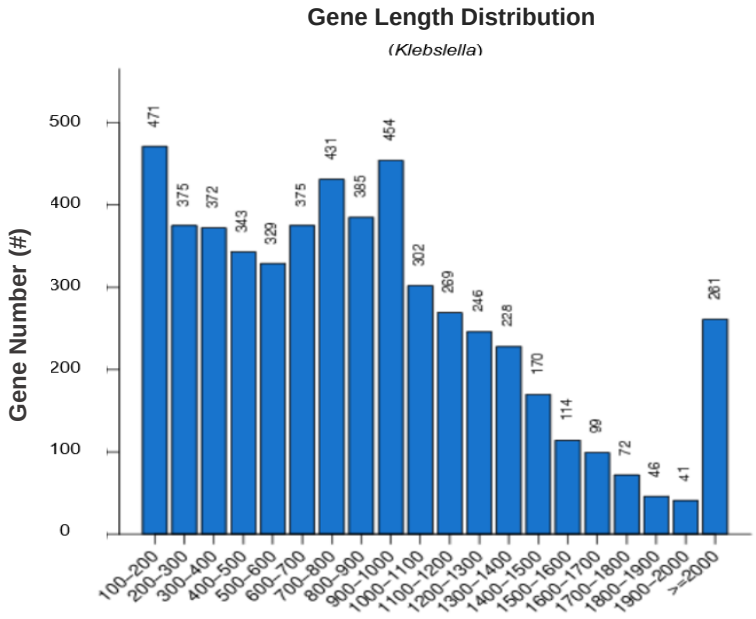

Fig. 3. Gene length distribution map

Gene Ontology

The distribution of enzymes that can catalyze carbohydrate degradation, modification, and biosynthesis were obtained by comparing the CAZy database (Carbohydrate-Active Enzymes Database, professional database related to carbohydrate enzymes), as shown in Table 1.

Table 1. Carbohydrate Enzyme-related Gene Statistics of Strain M1

Sample name	CEs number (#)	GHs number (#)	GTs number (#)	PLs number (#)	AAs number (#)	CBMs number (#)
Klebsiella pneumoniae M1	11	57	19	2	4	7

The genome of strain M1 contained five main classifications of carbohydrate-related enzymes: glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary activities (AAs). Carbohydrate-binding modules (CBMs) were also included. These results indicated that the genome of strain M1 was rich in carbohydrate-expressing genes and glycoside hydrolase-expressing genomes.

The classification statistics of strain M1 GO (Gene ontology) database are shown in Fig. 4. The number and types of functional genes constituting biological pathways in strain M1 genome were relatively larger, and they were mainly used for biosynthesis or biodegradation. The types and quantities of functional genes constituting cytological components were relatively small. Molecular functional genes were mainly reflected in the relatively large number of binding and catalytic activity genes.

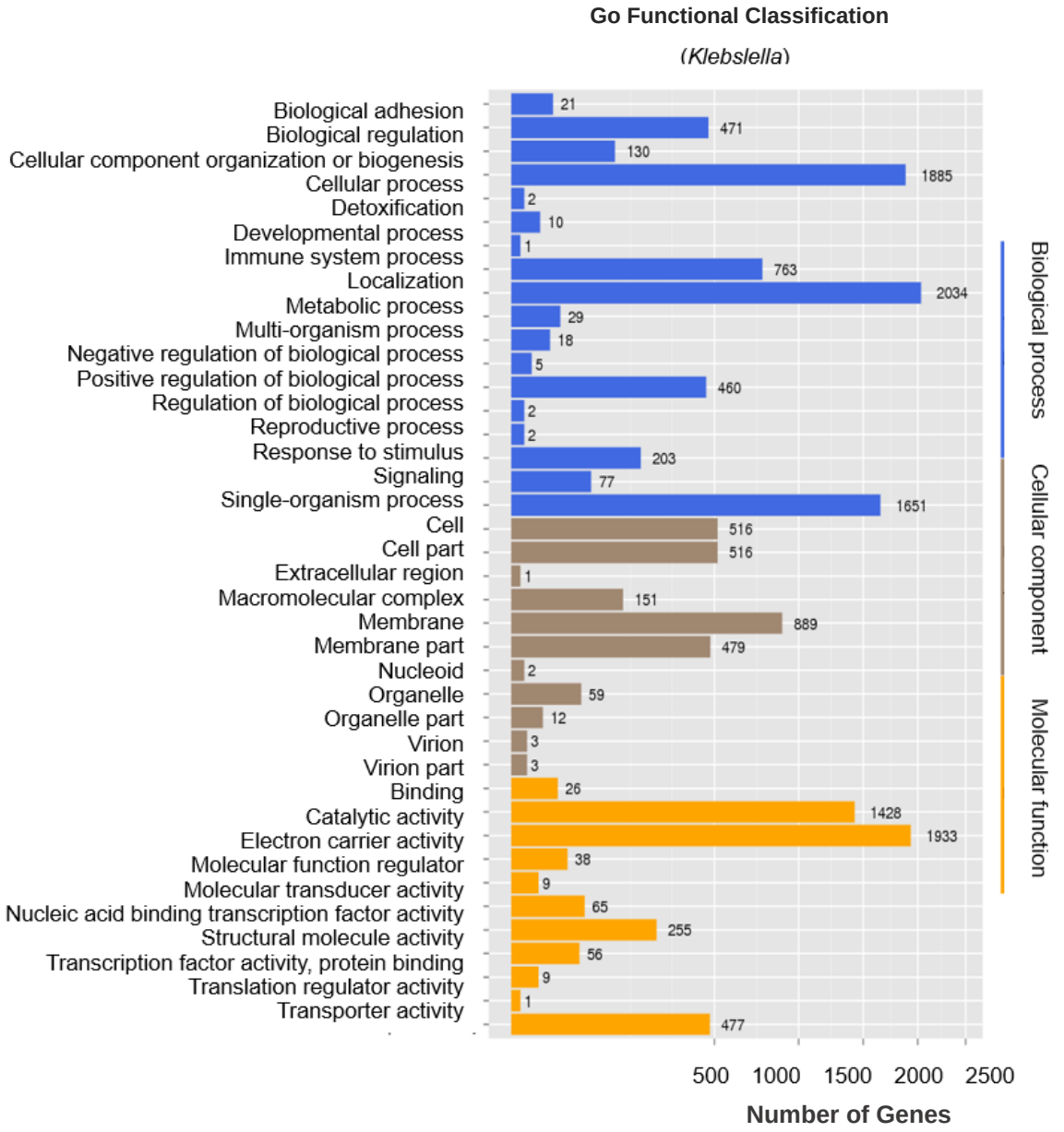

Fig. 4. Distribution of GO Functional Annotations

The COG (Clusters of Orthologous Groups of proteins) function prediction results are shown in Fig. 5. The protein sequences encoded by 5044 genes have corresponding COG numbers. The number of protein sequences involved in basic metabolism were 2473,

accounting for 49%. Carbohydrate transport and metabolic genes accounted for 11%, followed by amino acid transport and metabolic genes. Carbohydrate transporters and metabolic genes were dominant. Therefore, the study of the genes involved in the transport and metabolism of carbohydrates can provide a powerful basis for further elucidating the metabolic mechanism of polysaccharides.

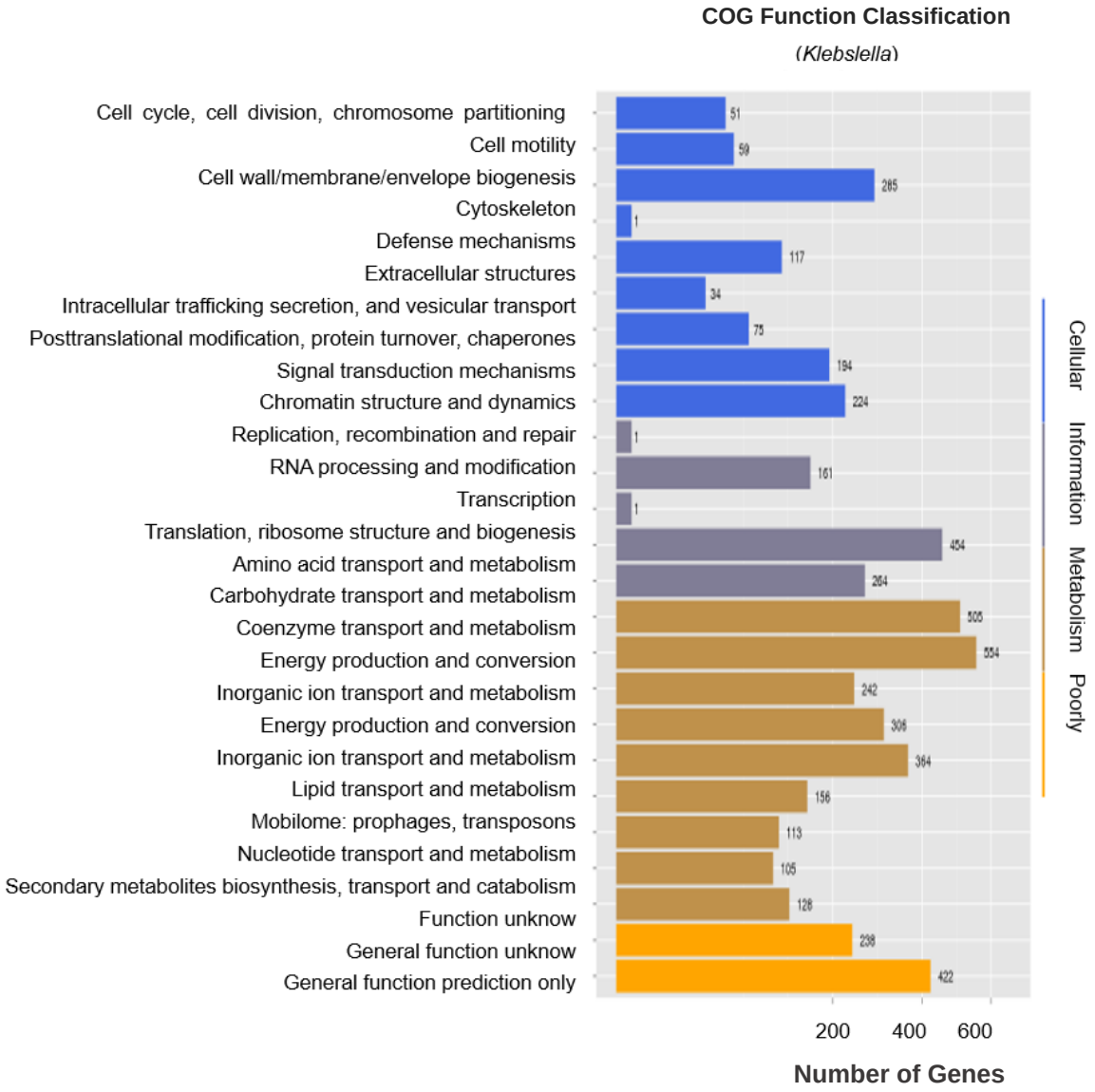

Fig. 5. Distribution of COG Functional Annotations

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a public database for genome research (Kanehisa and Goto 2000). The database shows the functional pathways of a gene ascription, including metabolic pathways, substance synthesis pathways, membrane transport processes, signal transduction, and disease-related pathways. The annotated results of KEGG database (Fig. 6) elucidate the biological functions and characteristics of genes more clearly at the system level. The BLAST algorithm was used to compare the predicted genes obtained with the KEGG database. A total of 3685 genes

were successfully mapped, and 1698 failed. They were distributed to 29 types of pathways respectively. The dominant pathway was metabolic pathway, which contained 2634 genes, accounting for 71% of all genes. The genomic metabolic pathways of the tested strains included globe and overview pathway, carbohydrate metabolism, energy metabolism, carbohydrate and lipid metabolism, nucleic acid and amino acid metabolism, cofactor, and vitamin metabolism. In the isolated bacteria, carbohydrate metabolism included typical glycolysis pathway, tricarboxylic acid cycle, and pentose phosphate pathway. These metabolic relating genes provided some data for fully exploring the biological characteristics of the bacteria.

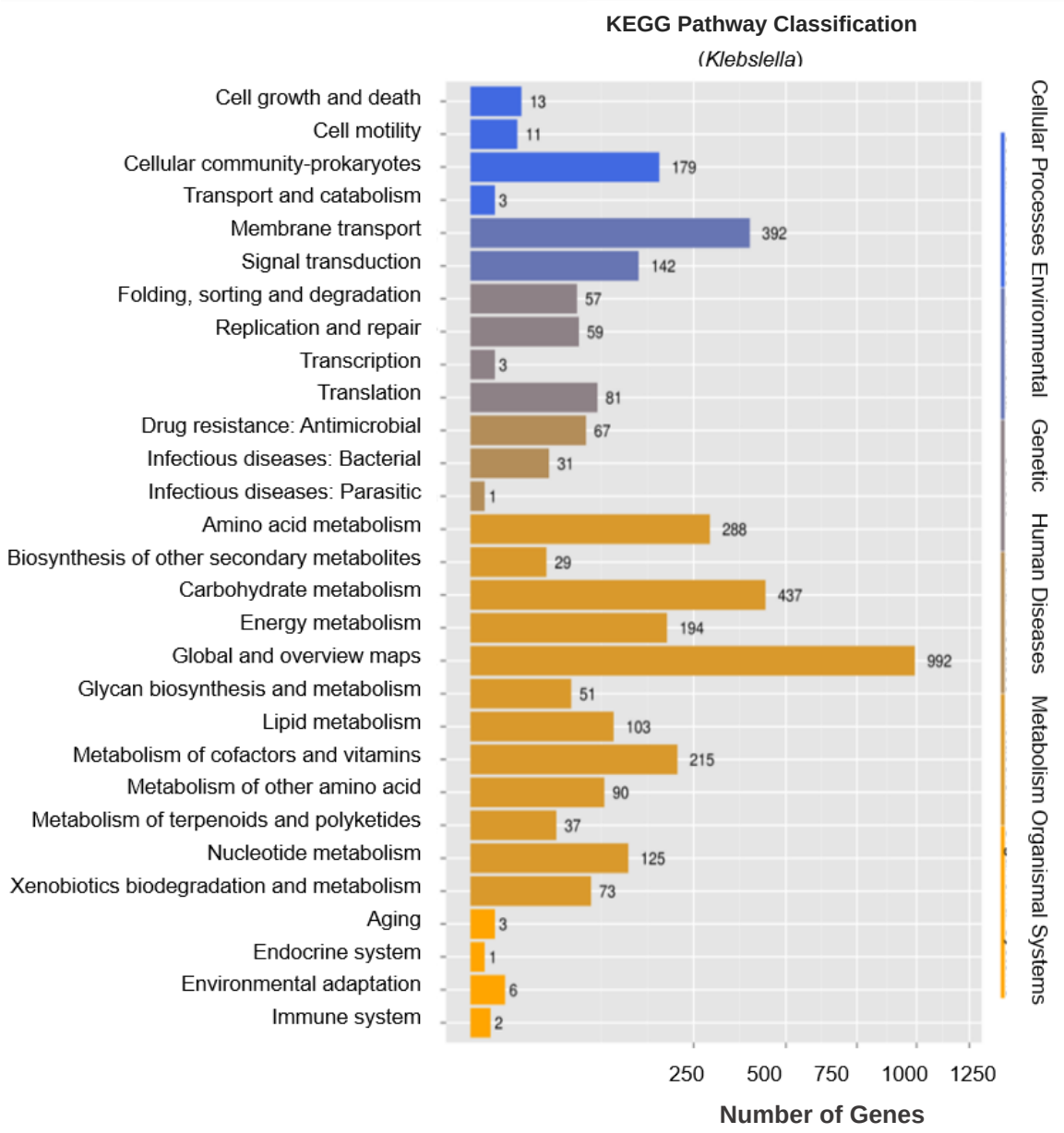

Fig. 6. Distribution of KEGG functional annotations

Common and Specific Genes

Common genes and specific genes among *K. pneumoniae* HS11286, MGH.78578, NTUHK2044, and M1 strains, gene set of each strain and statistical results of Core and Pan gene analysis were shown in Table 2.

Table 2. Statistical Results of Core and Pan Gene Analysis of Four Representative Strains

Core Gene Num (#)	Core Gene size (bp)	Pan Gene Num (#)	Pan Gene Size (bp)	Dispensable Num (#)	Dispensable size (bp)
3,851	1,237,283	7,306	1,990,133	1,182	286,262

The Venn diagram of genome pangenome set between four strains is shown in Fig. 7.

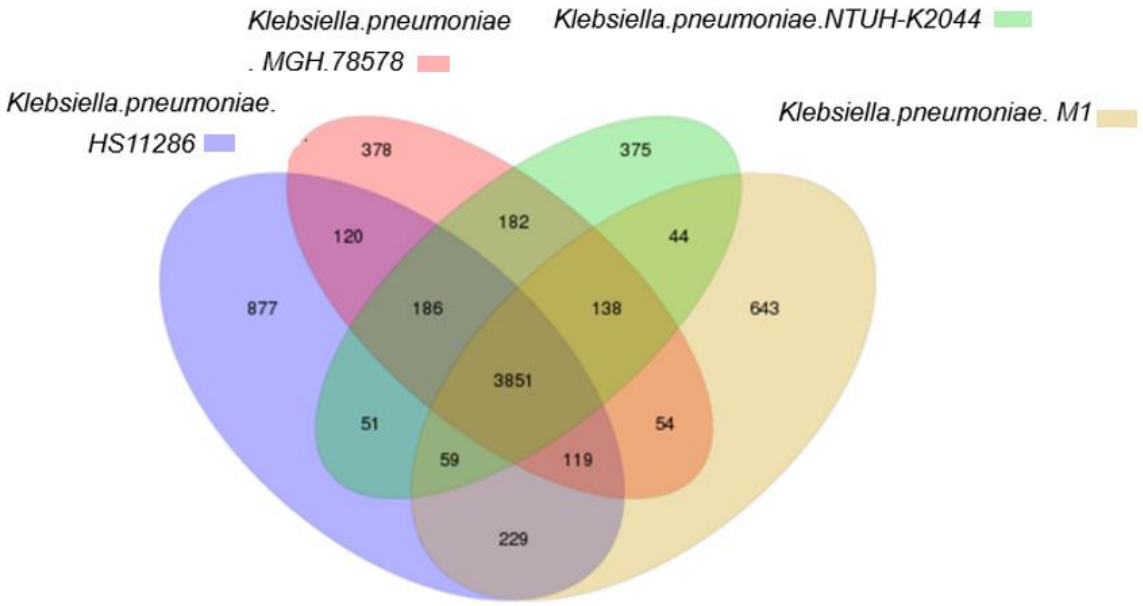

Fig. 7. Venn diagram of Pan gene set

As shown in Fig. 7, each strain was represented by an ellipse. Data on each region represented the number of clusters in the samples of the region. A cluster represented a group of genes with similarities greater than 50% and sequence length differences less than 0.3. The purpose of Pan gene research was to reveal the genetic diversity among individuals within bacterial species more comprehensively and to explore the genetic basis of phylogenetic relationships and phenotypic differences among individuals (Zhu and Zhu 2012).

Microbial flocculants has become a research hotspot because of its non-toxicity, high efficiency, and "green" environmental protection (Liu *et al.* 2013). It had been widely used in food, printing and dyeing, mineral, dairy, and other wastewater treatment (Wang *et al.* 2007; Okaiyeto *et al.* 2015;), but there was no report on microbial flocculation bacteria used in wheat starch alcohol wastewater treatment, and the specificity of flocculation bacteria was strong (Mayowa *et al.* 2017; Li *et al.* 2017). Microbial flocculants that can flocculate suspended matter in other sewage efficiently may not be as effective as

suspended matter in wheat starch alcohol wastewater. Therefore, it was necessary to develop the specific microbial flocculants. High-efficient microbial flocculant-producing bacteria were the core and key to obtain high-yield microbial flocculants. It was of great research value to understand flocculant bacteria from the molecular level and explore the gene sequence and functional genes of flocculant bacteria.

Whole genome sequencing technology is another important means to understand the microbial genome characteristics at a molecular level. Genomic sequencing has become a tool for people to study organisms at the molecular level, especially after the rise of third-generation sequencing technology, not only improve the speed of sequencing, but also reduce the cost of sequencing (Zhang and Yuan 2013). The genome of *Klebsiella pneumoniae M1* strain was studied by whole genome sequencing technology, which can predict the components of the genome, annotate the corresponding functional genes, and explore its metabolic pathways. Through the results of genome sequencing and its subsequent analysis, not only can obtain more abundant data resources, but also have an updated understanding of microorganisms at the molecular level.

The whole genome sequence of a flocculant strain *Klebsiella pneumoniae M1* from wheat alcohol wastewater was determined by PacBio RS II sequencing platform for the first time in this study. Its basic characteristics (sequence length, GC content, gene number, functional annotation) were analyzed, and the functional genes related to flocculation were explored. At present, there is no relevant report on the whole genome sequencing of flocculant bacteria for the treatment of wheat alcohol wastewater. By sequencing experiments, the genome length of strain M1 was 5,511,794 bp. The content of GC was 58.39%, and relatively high. A total of 5348 genes had obvious biological functions. However, 437 genes were involved in carbohydrate metabolism, and they had coding genes for five carbohydrate-related enzymes. This result indicated that there were functional genes closely related to polysaccharide production in strain M1 genome. Through comparative analysis of COG, GO, and KEGG databases, it was found that M1 strain genes mainly involved in genetic information processing, cell transport, and metabolic processes. In the metabolic process, M1 strain genes mainly participated in carbohydrate, amino acid, nucleotide, energy metabolism, and other metabolic pathways. It provided a theoretical basis for further understanding the flocculation mechanism, and follow-up study of functional genomics. Studying the regulation mechanism of flocculant active substances will be helpful to the application of microbial flocculants in industrial wastewater treatment. Whole genome sequencing technology is only the starting point of genomics research, but it cannot explain many biological phenomena and mechanisms. On the basis of the genomic information of bacteria, it is still a long way to go to understand, digest and even use the huge information.

DISCUSSION

In past decades, the whole genome sequencing technology has made great progress. The once time-consuming and labor-intensive sequencing projects can now be completed in a relatively short time. Limited by the complexity of data storage and data analysis strategies, large sample data analysis and complex data operations still need to rely on the Linux operating system, which requires researchers in this field to master certain bioinformatics skills. The research team can develop some easy-to-use online data analysis

platforms and visualization tools for locating the flocculation function gene and constructing the gene map of microbial flocculating bacteria. *Klebsiella pneumoniae* M1 is a newly discovered bacterium with high flocculating activity. Until the present study, there was little information that demonstrated its genetic information and metabolic processes. Therefore, it is fundamental to investigate and preserve the genetic information of the *Klebsiella pneumoniae* based on Illumina sequencing technology. The flocculating biological resources produced by *Klebsiella pneumoniae* M1 are mainly extracellular polysaccharides (Diao *et al.* 2019). Bacterial extracellular polysaccharide synthesis is a multi-step reaction process, which requires more proteins with special functions to participate in the synthesis of nucleotide sugar precursors, assembly of repeat units, polymerization, and transport. The genes encoding these proteins are usually arranged on some gene clusters (Rossi and De 2016). The 16 open reading frames of exopolysaccharide gene cluster of *Bacillus licheniformis* CGMCC 2876 are related to the synthesis of polysaccharide bioflocculant (Yan *et al*. 2013). There are few reports on the flocculation function gene of flocculating bacterium *Klebsiella*, and we need more researchers from different disciplines to work together to explore the gene location with flocculation function.

CONCLUSIONS

1. The M1 whole genome was 5,511,794 bp in length, and the GC content was up to 58.39%. The genome contained about 5383 genes, but 5348 genes had obvious biological functions. 437 genes were involved in carbohydrate metabolism and had coding genes of five carbohydrate-related enzymes. It was indicated that there were functional genes closely related to polysaccharide production in M1 genome.
2. The main metabolic process of flocculant strain *Klebsiella pneumoniae M1* was closely related to the potential pathway of extracellular polysaccharide biosynthesis, in which five kinds of carbohydrate synthase genes were involved. Likewise, this methodology is valuable for future research on the gene expression, genomics, and functional genomics of this species.

ACKNOWLEDGMENTS

This study was supported by a grant from the Anhui Key Research and Development Program (1804g07020169); Anhui Provincial Quality Engineering projects (2020jyxm0782; 2019mooc351; 2020kcszx04); and Key scientific research project of Anhui Provincial Department of Education (KJ2020A0790); Start up fund for scientific research of Anhui Xinhua University (2019kyqd01); Anhui College Students' innovation and entrepreneurship training program (No. AH 202012216096, AH 202012216097).

REFERENCE CITED

Aida, N. R., Amara, N. K. W. S., and Luqman, H. A. R. (2022). "The potency of microbial flocculant produced by *B. licheniformis* using molasses as the carbon source and its application in food industry wastewater treatment," *Materials Today: Proceedings* (63), s244-s247. DOI: 10.1016/j.matpr.2022.02.433

Chen, J., Zhan, Z. J., Chen, Z., Wang, T., and Yang, Z. F. (2012). "Identification, gene mapping and flocculant composition analysis of a flocculant," *Biotechnology Bulletin* 4, 122-126.

Diao, H., Li, L. M., Liang, J., and Ding, X. L. (2018). "Screening of high-performance flocculant-producing bacteria and optimization of the conditions for flocculation of wheat distillery wastewater," *BioResources* 13(4), 7738-7757. DOI: 10.15376/biores.13.4.7738-7757

Diao, H., Li, L. M., Liang, J., and Ding, X. L. (2019). "Optimized extraction process and compositional analysis of bioflocculant produced by *Klebsiella* M1," *BioResources* 14(2), 3146-3167. DOI: 10.15376/biores.14.2.3146-3167

Fu, L. L., Jiang, B. H., Liu, J. L., Zhao, X., Liu, Q., and Hu, X. M. (2016). "Genome sequence analysis of a flocculant-producing bacterium, *Paenibacillus Shenyang-ensis*," *Biotechnology Letters* 38, 447-453. DOI: 10.1007/s10529-015-1990-2

Ghosh, M., Pathak, S., and Ganguli, A. (2009). "Effective removal of *Cryptosporidium* by a novel bioflocculant," *Water Environment Research* 81(2), 160-164. DOI: 10.2175/106143008X325818

Guo, A. J., Nie, M. X., Xing, J., Wu, X. Z., and Chen, X. D. (2013). "Effects of nitrogen sources on flocculant secretion characteristics of *Klebsiella* NIII2," *Journal of Environmental Engineering* 7(5), 1996-2000.

He, L. Y. (2012). *Isolation and Functional Study of Flocculating Gene of Flocculating Yeast SPSC01*, Master's Thesis, Dalian University of Technology, Dalian, China.

Kanehisa, M., and Goto, S. (2000). "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Research* 28(1), 27-30. DOI: 10.1093/nar/28.1.27

Li, X. (2007). *Fermentation Culture of Microbial Flocculant-producing Bacteria and its Application in Sludge Dewatering*, Master's Thesis, Hunan University, Changsha, China.

Li, J., Yun, Y. Q., Xing, L., and Song, L. (2017). "Novel bioflocculant produced by salt-tolerant, alkaliphilic strain *Oceanobacillus polygoni* HG6 and its application in tannery wastewater treatment," *Bioscience, Biotechnology, and Biochemistry* 81(5), 1018-1025. DOI: 10.1080/09168451.2016.1274635

Liu, P.L. (2012). *Study on the Mobile Genome of Klebsiella pneumoniae*, Master's Thesis, Shanghai Jiaotong University, Shanghai, China.

Liu, Z. Y., Hu, Z. Q., Wang, T., Chen, Y. Y., Zhang, J., Yu, J. R., Zhang, T., Zhang, Y. F., and Li, Y. L. (2013). "Production of novel microbialflocculants by *Klebsiella* sp.TG-1 using waste residue from the food industry and its use in defecating the trona suspension," *Bioresource Technology* 139, 265-271. DOI: 10.1016/j.biortech.2013.03.165

Liu. J. W., Ma, J. W., Liu, Y. Z., Yang, Y., Yue, D. B., and Wang, H. T. (2014). "Culture optimization of flocculant M-C11 produced by *Klebsiella* and its application in sludge dewatering," *Environmental Science* 35(3), 1183-1190. DOI: 10.13227/j.hjkx.2014.03.052

Liu, C., Sun, D., Liu, J. W., Zhu, J. R., and Liu, W. J. (2021). "Recent advances and perspectives in efforts to reduce the production and application cost of microbial flocculants," *Bioresour. Bioproc.* 8(1), 1-20. DOI: 10.1186/s40643-021-00405-2.

Mayowa, O. A., Esta, V. H., Carolina, H. P., and Anofi, T. A. (2017). "Flocculating performance of a bioflocculant produced by *Arthrobacter humicola* in sewage waste water treatment," *BMC Biotechnology* 17(1), 51-61. DOI: 10.1186/s12896-017-0375-0

Metzker, M. L. (2010). "Sequencing technologies-the next generation," *Nature Reviews Genetics* 11(1), 31-46. DOI: 10.1038/nrg2626

Okaiyeto, K., Nwodo, U. U., Mabinya, L.V., and Antholy, I. O. (2015). "*Bacillus toyonensis* strain AEMREG6, a bacterium isolated from South African marine environment sediment samples produces a glycoprotein bioflocculant," *Molecules* 20(3), 5239-5259. DOI: 10.3390/molecules20035239

Rossi, F., and De, P. R. (2016). "The Physiology of microalgae," in: *Exocellular Polysaccharides in Microalgae and Cyanobacteria: Chemical features, Role and Enzymes and Genes Involved in Their Biosynthesis,* M. A. Borowitzka, J. Beardall, J. A. Raven (ed.), Springer International Publishing, Switzerland, pp.565-590.

Wang, S. G., Gong, W. X., and Liu, X. W. (2007). "Production of a novel bioflocculant by culture of *Klebsiella mobilis* using dairy wastewater," *Biochemical Engineering Journal* 36(2), 81-86. DOI: 10.1063/1.1453360

Wei, J. L., Cong, L., and Di, S. (2017). "Complete genome sequence of a novel bioflocculant-producing strain, *Microbacterium paludicola CC3*," *American Society for Microbiology* 38(5), 1008-1017. DOI: 10.1128/genomeA.01008-17

Xing, J. (2013). *Screening and Characterization of Ultraviolet Mutant Strains of Klebsiella spp. with High Yield of Flocculant*, Master's Thesis, Xi'an University of Architecture and Technology, Xi'an, China.

Xing, J., Song, N. Z., Chen, X. W., Li, A., and Ni, H. W. (2020). "Flocculation efficiency and mechanism of carbamazepine by microbial flocculant extracted from *Klebsiella pneumoniae J1*," *Archaea* 2020(11), 1-8. DOI: 10.1155/2020/8811516.

Xue, J. (2013). *Construction of Flocculant Genetic Engineering Bacteria and Screening of Flocculant-producing Bacteria*, Master's Thesis, Inner Mongolia University, Hohhot, China.

Yan, S., Wang, N., Chen, Z., Wang, Y. P., He, N., Peng, Y. J., Li, Q. B., and Deng, X. (2013). "Genes encoding the production of extracellular polysaccharide bioflocculant are clustered on a 30—kb DNA segment in *Bacillus licheniformis*," *Functional & Integrative Genomics* 13(4), 425-434. DOI: 10.1007/s10142-013-0333-4

Zeng, B. H. (2012). *Proteomic Analysis of Klebsiella pneumoniae*, Master's Thesis, Luzhou Medical College, Luzhou, China.

Zhang, Q. L., and Yuan, M. L. (2013). "Advances in insect transcriptionomics based on the new generation sequencing technology," *Journal of Entomology* 56, 1489-1508.

Zhang, S. C. (2013). *Genome Splicing of Klebsiella pneumoniae Based on High Throughput Sequencing*, Master's Thesis, Shanghai Normal University, Shanghai, China.

Zhu, J. M., Jiang, R. J., Wu, K. L., Zhang, F., Weng, X. B., and Kong, H. S. (2014). "Molecular evolution analysis of single gene and genomic systematics of *Klebsiella pneumoniae*," *Chinese Journal of Hospital Infectiology* 12, 2861-2864. DOI: 10.11816/cn.ni.2014-132770

Zhu, X.R., and Zhu, Y. Z. (2012). "Research on bacterial pangenomics," *Journal of Shanghai Jiaotong University (Medical Edition)* 32(1), 1440-1443. DOI: 10.3969/j.issn.1674-8115.2012.11.010

Article submitted: April 18, 2022; Peer review completed: June 19, 2022; Revisions accepted: July 1, 2022; Published: July 6, 2022.
DOI: 10.15376/biores.17.3.4870-4885