Transcriptomic Analysis of *Klebsiella pneumoniae* Under High Glucose Condition and Analysis of Metabolic Related Genes

Reenu S P[1*], Ramesh Kumar Kushwaha[2], Ramanjaneyulu Golla[3], Thayaba Surur[4]

[1]PhD Research Scholar, Department of Microbiology, REVA UNIVERSITY Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bangalore, Karnataka, India, 560064. Email ID: spreenu8@gmail.com. ORCID ID: 0009-0006-4434-5301

[2]Assistant Professor, Department of Education, Bihar Government Email ID: kush0109@yahoo.com. ORCID ID: 0000-0003-0098-0398

[3]Assistant Professor, Department of Biochemistry, REVA UNIVERSITY Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bangalore, Karnataka, India, 560064. Email ID: ramanjaneyulu.golla@reva.edu.in. ORCID ID: 0000-0002-6847-5307

[4]Department of Microbiology, REVA UNIVERSITY Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bangalore, Karnataka, India, 560064. Email ID: thayabasurur333@gmail.com. ORCID ID: 0009-0006-9094-3783

***Corresponding author:**

Reenu S P, PhD Research Scholar, Department of Microbiology, REVA UNIVERSITY Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bangalore, Karnataka, India, 560064. Email ID: spreenu8@gmail.com. Phone number: +91 9845690117. ORCID ID: 0009-0006-4434-5301

Abstract

Introduction

Klebsiella pneumoniae infection is an important opportunistic pathogen associated with hospital-acquired infections and increasing antimicrobial resistance. Hyperglycaemic conditions, particularly in diabetic patients, may influence bacterial metabolism, virulence, and antibiotic resistance. This study aimed to investigate the transcriptomic response of *Klebsiella pneumoniae* under high-glucose conditions and identify metabolic, virulence-associated, and antimicrobial resistance (AMR) genes involved in bacterial adaptation.

Methodology

Klebsiella pneumoniae cultures were grown in control and high-glucose conditions, and bacterial growth kinetics were monitored using spectrophotometry. Total RNA was extracted, and RNA sequencing was performed on the Illumina HiSeq platform. High-quality reads were analysed by standard bioinformatics pipelines including FastQC, TrimGalore, Bowtie2, featureCounts and edgeR for differential gene expression (DEG) analysis. Differentially expressed genes were considered significantly differentially expressed if fold change ≥ 2 and adjusted p-value < 0.05. Functional annotation and pathway enrichment analyses were performed using KEGG, Gene Ontology (GO), STRING and ISMARA databases to identify biological pathways, virulence factors and AMR-associated genes.

Results

Transcriptional analysis revealed extensive changes in gene expression under high-glucose conditions, indicating significant metabolic reprogramming and adaptive responses of K. pneumoniae. A total of 63 virulence-associated genes and 40 antimicrobial resistance (AMR) genes were identified. The gene *yagZ/ecpA*, encoding an adhesin-associated pilus structural protein, was the most up-regulated among virulence genes with a log fold change of 13.003, whereas biofilm-associated genes *csgF* and *csgG* were down-regulated. Among the AMR genes, *mphK* associated with macrolide resistance was strongly up-regulated (log fold change 11.102), and *mdtA* was slightly down-regulated. Altered expression of β-lactamase genes and efflux pump-associated genes indicated enhanced antimicrobial resistance under glucose stress conditions.

Conclusion

Overall, high glucose exposure induces significant transcriptomic changes in *K. pneumoniae*, enhancing virulence potential and antimicrobial resistance. These findings provide insight into bacterial adaptation in hyperglycaemic conditions and may inform the development of targeted therapeutic strategies.

Keywords: Klebsiella pneumoniae; Transcriptomics; RNA sequencing (RNA-seq); Differential gene expression; High glucose stress.

How to cite this article: Reenu SP, Kushwaha RK, Golla R, Surur T. Transcriptomic Analysis of Klebsiella pneumoniae Under High Glucose Condition and Analysis of Metabolic Related Genes. Int J Drug Deliv Technol. 2026;16(7): 254-271. DOI: 10.25258/ijddt.16.7.28

Introduction

Klebsiella pneumoniae can acquire resistance to multiple antibiotics, such as broad-spectrum cephalosporins and β-lactams. It is an important pathogen in hospital settings, as it contributes to around 11.8% of all cases of hospital-acquired pneumonia globally. Fortunately, we can study *Klebsiella pneumoniae* in a laboratory due to its genetic tractability [1].

Transcriptome sequencing (RNA sequencing) is a recently developed analysis that uses high-throughput sequencing approaches for the next generation. The transcriptome is the collection of all RNA transcribed by an individual or a population of cells at a certain biological state. Transcriptome analysis of mRNA sequence using Illumina HI sequence methods to generate a million threads [2].

Transcriptomic analysis estimates the total number of mRNA transcripts to examine gene expression. It offers an understanding of the biological mechanisms and pathways that become activated in specific conditions. Transcriptomic analyses can be verified by an alternative method, such as quantitative PCR (qPCR), which is easily identifiable and can be statistically evaluated. Gene expression is compared to established benchmarks for both the gene being studied and the control genes [3].

The methods used to examine an organism's transcriptome—the totality of its RNA transcripts—are known as transcriptomics technology. An organism's information is encoded in its genome's DNA and expressed through transcription. In this case, mRNA acts as a temporary bridge molecule in the information network, while non-coding RNAs carry out a variety of other roles. A transcriptome is a temporal snapshot of all the transcripts that are present in a single cell. Transcriptomics technology provides a comprehensive picture of both active and inactive cellular processes. In molecular biology, comprehending how a single genome gives rise to a multitude of cells is a fundamental difficulty. The regulation of gene expression is another [4].

Transcriptomics involves examining the transcriptome, which refers to all the RNA present in a sample (such as a cell, tissue, or organ) at a specific moment. RNA plays various roles in the cell, and analysing the transcriptome offers a glimpse into gene activity and protein production within the organism [5].

In this study, we conducted a transcriptomic analysis of *Klebsiella penumoniae* grown under high glucose conditions (T2) to compare the changes in gene expression profiles with those grown under standard conditions (C2).

Trans genomic analysis was conducted to find the virulence genes, genes responsible for resistance against antibiotics. RNA sequencing was used to measure transcript levels, followed by bioinformatics analyses to identify differentially expressed genes and enriched pathways.

The study revealed a direct correlation between mRNA expression of these genes and the presence of antibiotic resistance; also corroborated by mutation analysis of the *AmpC* promoter region**.**

Materials and Methods

Bacterial Strains and Growth Conditions

All bacterial strains used in this study are represented in Table 1. *E. coli* strains were cultured at 37 C in Luria-Bertani (LB) (Becton & Dickinson, Sparks, MD, USA) broth or on Luria-Bertani Agar (Becton & Dickinson, Sparks, MD, USA). Wild-type (WT) strain refers to BW25113.

Growth Curves at High Glucose

An overnight culture of *K. pneumonia* was normalised to 1.0 McFarland using a Densimat (bioMérieux, Marcy-l'Étoile, France) and further diluted 1:10 in fresh LB. This bacterial culture was grown in a volume of 1 mL in a 24-well plate (ThermoFisher, Waltham, MA, USA) and incubated at 37 _C with 5 mm orbital shaking in an Infinite 200PRO plate reader (Tecan, Männedorf, Switzerland). Absorbance was measured every 10 min at an optical density of 595 nm (OD595nm). At the early exponential phase (OD595nm = 0.2), H2O2 was added from a 35% w/w (Acros Organics, Geel, Belgium) solution freshly diluted at 11-fold concentration in LB and serially diluted 1:2 in LB. A volume of 100 _L was added to the 1 mL bacterial culture (total volume per well: 1100 _L) to reach final concentrations ranging from 80 mM to 0.078 mM H2O2.

The concentration of H2O2 stock was assessed before the experiments by using a WPA Biowave DNA spectrophotometer (Biochrom, Cambridge, UK) at 240 nm with a quartz cuvette with a molar extinction coefficient of 43.6. The Doubling Time Software v3.1.0 (http://www.doublingtime.com, accessed on 18 February 2022) was used to calculate the growth rate after H2O2 addition.

RNA Isolation, RNA-Seq

2 mL of bacterial cultures (see conditions above) were harvested 10 and 60 min after H2O2 addition. RNA was extracted using the TRIzol® Max™ Bacterial RNA Isolation Kit (ThermoFisher, Waltham, MA, USA) according to the manufacturer's instructions. After elution in RNA-free water, DNA digestion was performed in liquid by adding 2.5 L DNase I and 10 L Buffer RDD

(Qiagen) in a total volume of 100 L for 10 min at room temperature. Potential contaminants were removed by using the RNeasy mini kit plus (Qiagen, Germantown, MD, USA) according to the manufacturer's instructions. Analysis of total RNA quality and quantity was performed using Nanodrop and Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA).

For RNA sequencing, 1.2 _g of total RNA was ribo-depleted with the RiboMinus™ Bacteria 2.0 Transcriptome Isolation kit (ThermoFisher Scientific, Waltham, MA, USA). The TruSeq total RNA stranded kit (Illumina, San Diego, CA, USA) was then used for the library preparation. Library quantity was measured by Qubit, and quality was assessed with a Tapestation on a DNA High sensitivity chip (Agilent Technologies, Santa Clara, CA, USA). Libraries were pooled at equimolarity for clustering. Single-read sequencing (100 bases) was performed using the SBS chemistry on an Illumina HiSeq 4000 sequencer. Raw data have been deposited at ENA (https://www.ebi.ac.uk/ena, accessed on 4 March 2022) under the following accession number: PRJEB51098.

The RNA-seq data were analysed as follows. Fastq reads were mapped to the ENSEMBL reference genome (Ecolibw25113.48) using STAR (version 2.4.0j) with standard settings, except that any reads mapping to more than one location in the genome (ambiguous reads) were discarded (m = 1). All reads were reported using feature Counts version 1.4.6-p1. Levels of expression were reported as raw counts and, in parallel, normalised in RPKM to filter out genes with low expression values (<1 RPKM) before calling genes as differentially expressed. Library size normalisations and differential gene expression calculations were performed using the package edgeR designed for the R software. Only genes having a significant fold-change (Benjamini-Hochberg corrected p-value < 0.01) were considered for the rest of the RNA-seq analysis. Genes under the regulation (regulon) of selected transcription factors were investigated in the Ecocyc database and used to generate heatmaps. Genes were categorised in different expression patterns depending on whether they were significantly up-regulated, down-regulated, or unchanged at the 2 time points, leading to 8 different groups. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) web software was used to analyse physical or functional interactions in each of the 8 groups of genes.

GO term enrichment was performed using homemade scripts for the R software. Gene set enrichment analysis (GSEA) was performed, without a threshold of p-value or fold-change, using the KEGG metabolic pathways (KEGG http://www.genome.jp/kegg/, accessed on 15 December 2021) relative to Ecolibw25113.48 to generate gene sets. Genes were ranked by their calculated fold-changes (decreasing ranking). A gene set analysis using the GSEA package Version 2.2 from the Broad Institute (MIT, Cambridge, MA, USA) was used to analyse the pattern of differential gene expression between the two groups. Gene set permutations were performed 1000 times for each analysis. The Normalised Enrichment Score (NES) was calculated for each gene set. GSEA results with a nominal FDR < 0.05 and abs (NES) > 1 were considered significant. ISMARA analysis was performed to retrieve transcription factors implicated in causing changes in gene expression. Raw fastq data were uploaded to the ISMARA web server (Swiss Institute of Bioinformatics, Basel, Switzerland). Four replicates were averaged for each condition.

Softwares

GraphPad Prism v.9.3.1 for Windows (GraphPad Software, San Diego, CA, USA) was used for data processing, graph plotting and statistical analysis. Inkscape v.1.1.1 for Windows was used for image editing (https://inkscape.org).

Transcriptomic analysis

FASTA – the FASTA format is text based format for representing either nucleotide sequences or amino acid (protein) sequences, in which nucleotides and amino acids are represented using single-letter codes. A sequence begins with a greater-than character (">") followed by a description of the sequence (all in a single line). The lines immediately following the description line are the sequence representation, with one letter per amino acid or nucleic acid, and are typically no more than 80 characters in length.

Example:

>MCHU - Calmodulin - Human, rabbit, bovine, rat, and chicken

MADQLTEEQIAEFKEAFSLFDKDGDGTITTKE
LGTVMRSLGQNPTEAELQDMINEVDADGN

GTID

FPEFLTMMARKMKDTDSEEEIREAFRVFDKD
GNGYISAAELRHVMTNLGEKLTDEEVDEMI

REA

DIDGDGQVNYEEFVQMMTAK*

FASTQ - FASTQ format is a text-based format for storing both a biological sequence

(usually nucleotide sequence) and its corresponding quality scores.

1. A FASTQ file has four line-separated fields per sequence:

• Field 1 begins with a '@' character and is followed by a sequence identifier and

an optional description (like a FASTA title line).

• Field 2 is the raw sequence letters.

• Field 3 begins with a '+' character and is optionally followed by the same sequence identifier

(and any description) again.

• Field 4 encodes the quality values for the sequence in Field 2, and must contain the same

number of symbols as letters in the sequence.

Fig : - Header Sequence Quality

FastQC - FastQC aims to provide a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines. It provides a modular set of analyses which you can use to give a quick impression of whether your data has any problems of which you should be aware before doing any further analysis.

The main functions of FastQC are

• Import of data from BAM, SAM or FastQ files (any variant)

• Providing a quick overview to tell you in which areas there may be problems

• Summary graphs and tables to quickly assess your data

• Export of results to an HTML-based permanent report

• Offline operation to allow automated generation of reports.

Flowchart of FASTA QC using Galaxy Australia search for galaxy Australia

in google open the software and register an account.

↓

Upload the FASTQ data to be checked.

↓

Search FastQC tool and run the test

↓

Check the FastQC reports.

↓

MultiQC is a tool to merge 2 or more FastQC data. Search for MultiQC

tool and select the FastQC reports and run the test.

↓

TrimGalore is a tool used to trim unwanted data and adapters.

Search for the TrimGalore tool, select the files and run the test.

↓

The next step is alignment. Alignment is a way to compare related DNA or protein sequences. They can be used to capture various facts about the sequences aligned, such as common evolutionary descent or common structural function. Download a reference FASTA data from NCBI in gcf format.

↓

Upload the downloaded data to Galaxy Australia.

↓

Once turned green, go for Bowtie 2.

↓

Select single or paired ends

↓

Select the trimmed files.

↓

Select the reference files.

↓

The results are in SAM (Sequence Alignment Map) format to summarize the results, run the Sam tools and flagstat tool.

*Note – since we used a reference data of a different bacterium, the mapping of the sequences was very low. If they were of the same species, the mapping percentage should be high.

↓

Run the feature counts tool. Feature counts is a program that counts how many reads map to genomic features, such as genes, exons, promoters, and genomic bins.

RNA SEQUENCE

RNA-Seq, short for RNA sequencing, is a cutting-edge methodology that employs next-generation sequencing to uncover the abundance and existence of RNA molecules within a biological specimen. By doing so, it offers a comprehensive overview of gene expression in the sample, commonly referred to as the transcriptome.

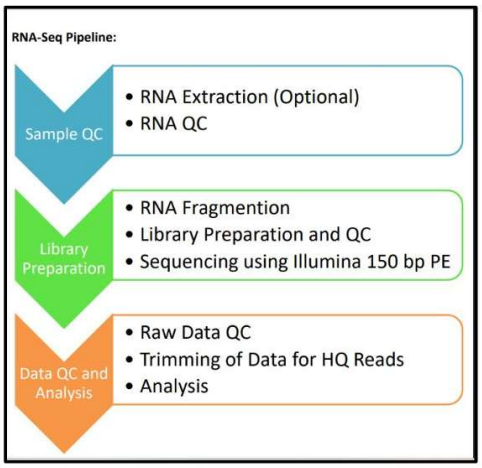


Fig :- RNA-Seq Pepline

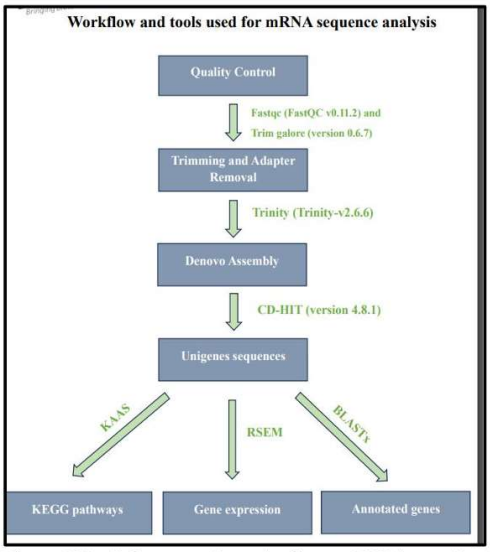


Fig ; -Workflow and tools for m RNA analysis

Fig :- data qc

De novo sequencing

De novo sequencing refers to the process of assembling sequences from scratch, without the aid of a reference genome. This is particularly important in genomics when sequencing new species or organisms for which no reference genome exists.

De Novo Sequencing offers numerous benefits:

• Including the production of precise reference sequences for intricate or polyploid genomes. It also supplies valuable data for the mapping of genomes from unfamiliar organisms or the completion of genomes from recognised organisms.

• Additionally, it aids in the clarification of highly similar or repetitive regions to facilitate

accurate de novo assembly. Furthermore, this sequencing method is capable of identifying structural variants and intricate rearrangements, such as deletions, inversions, or translocations.

• Two common types of de novo assemblers are greedy algorithm assemblers and De Bruijn graph assemblers.

De novo Assembly of reads - De novo assembly of high-quality reads was accomplished using Trinity software. Transcript sequences generated are provided in the data deliverables folder entitled “Primary assembly”.

Description	
Total trinity transcripts	79422
Total assembled bases	54147099
Average contig	681.76
Contig N50	757

Transcripts were further processed for Unigene prediction using the CD-HIT package. Fasta sequences of these Unigenes predicted are provided in the data deliverables folder entitled "Unigene".

Basic statistics for Unigene predicted is given in table below:

Description	
Total transcripts	73288
Total assembled bases	45344342
Average contig	618.71
Contig N50	605

Summary of transcripts, length distribution and unigenes were assembled

Length	No. of transcripts	No. of Unigenes
301-400	43832	42114
401-500	13013	12057
501-600	5900	5369
601-700	3356	2996
701-800	2309	2053
801-900	1559	1388
901-1000	1096	964
>1000	8357	6347

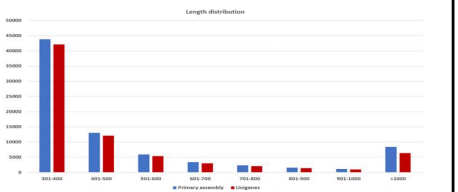
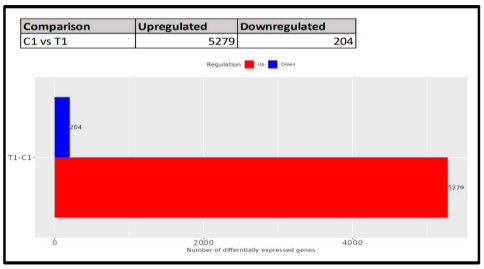


Fig. 2: Transcript length distribution of primary assembly and unigenes where number of Transcript is plotted in Y-axis and length range of transcript is plotted in X-axis.

DEG stats

Note: DEGs have been filtered using a minimum fold change threshold of 2 and an adjusted p-value threshold of 0.05. Samples: C1, T1 Comparisons:

Comparison	Upregulated	Downregulated
C1 vs T1	5279	204



PCA - The PCA (Principal Component Analysis) plot shows how the samples in the dataset are related to each other based on their gene expression patterns. Variation between the expression profiles of samples and their counterpart groups is explained by the first (x-axis) and second (y-axis) principal components. In the PCA plot, each data point represents a sample, and the position of the points in the plot is determined by the expression levels of genes. The first principal component represents the largest source of variability in the data, and the second component represents the second-largest source of variability, and so on. The distance between two data points in a PCA plot reflects the dissimilarity between their gene expression profiles.

When samples are diagonally opposite, it means that they are farthest away from each other in terms of gene expression, indicating a substantial difference between the two samples, and they are influenced by different sets of genes or expression patterns that are nearly orthogonal to each other.

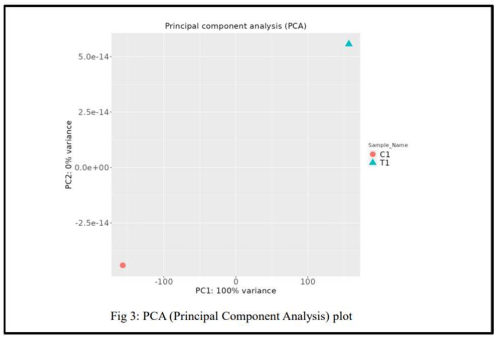


Fig 3: PCA (Principal Component Analysis) plot

Volcano Plot:

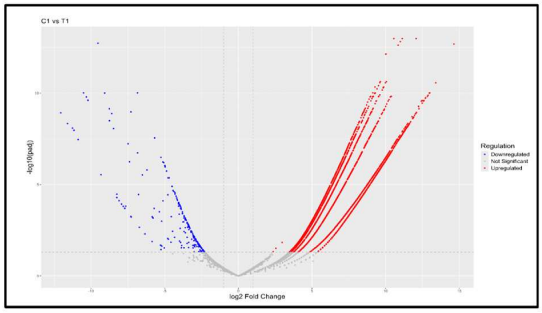
In the volcano plot, each gene is represented by a data point, and two key measurements are used to position these points on the graph:

• Fold Change: The horizontal axis shows the fold change, which is the ratio of gene expression levels between the two groups. Genes that have a fold change greater than 1 are upregulated, and those with a fold change less than 1 are downregulated.

• Statistical Significance (negative logarithm of adjusted p-value - log10(padj)): The vertical axis illustrates the statistical significance of the gene expression change, often portrayed as the negative logarithm of the adjusted p-value. The adjusted p-value accounts for multiple testing and helps control the false discovery rate.

• Smaller adjusted p-values indicate more significant changes. Thus, genes with low adjusted p-values are considered statistically significant.

• In the volcano plot, genes that exhibit significant upregulation in the experimental condition will appear on the right side of the plot, displaying high fold change values and low adjusted p-values. Conversely, genes significantly downregulated will appear on the left side, characterised by low fold change values and low adjusted p-values. The adjusted p-value, reflecting the probability of observing the observed gene expression difference given multiple testing, provides a more stringent measure of statistical significance in RNA-seq analyses.



• In the volcano plot, genes that exhibit significant upregulation in the experimental condition will appear on the right side of the plot, displaying high fold change values and low adjusted p-values. Conversely, genes significantly downregulated will appear on the left side, characterised by low fold change values and low adjusted p-values. The adjusted p-value, reflecting the probability of observing the observed gene expression difference given multiple testing, provides a more stringent measure of statistical significance in RNA-seq analyses.

The x-axis shows the fold change, which is a measure of how much a gene's expression level changes between two groups (e.g., disease vs. control).

Values to the right indicate upregulation, while values to the left indicate downregulation. The y-axis represents the statistical significance of the adjusted p-value.

Lower values (closer to the top) indicate higher significance, suggesting that the observed change in gene expression is less likely due to chance. In other words, as you move upwards on the y-axis, focusing on genes where the evidence for a real change in expression is stronger, even after accounting for the number of comparisons made during the analysis. Explanation of the colors of the dots: (Here C1 is the control group and T1 is the test condition in which up and down regulated genes have been identified.)

Red data points: represent genes that are significantly upregulated in the experimental condition. These genes have both a substantial fold change and a low p-value, indicating a strong association with the condition.

• Blue data points: represent genes that are significantly downregulated in the experimental condition.

Like the red dots, they also have a substantial fold change and a low p-value, suggesting a robust connection to the condition, but in the opposite direction.

• Gray data points: represent genes that do not show significant differential expression between the two groups. These genes have fold change values closer to 1 and higher p-values, indicating that their expression levels are relatively similar in both conditions.

Gene name	Product	Gene expression
yagZ/ecpA	*E. coli* common pilus structural subunit	upregulation with the log fold value of 13.003
yagY/ecpB	E. coli common pilus chaperone	upregulation with the log fold value of 9.103
yagX/ecpC	E. coli common pilus usher	upregulation with the log fold value of 9.103
yagV/ecpE	E. coli common pilus chaperone	upregulation with the log fold value of 6.594
aslA	putative arylsulfatase	upregulation with the log fold value of 0.716
aslA	putative arylsulfatase	upregulation with the log fold value of 0.147
entA	23-dihydro-23-dihydroxybenzoate dehydrogenase	upregulation with the log fold value of 7.812
entD	phosphopantetheinyl transferase component of enterobactin synthase multienzyme complex	downregulation with the log fold value -0.4143
iucB	aerobactin synthesis protein IucB	downregulation with the log fold value -0.410
iucC	aerobactin siderophore biosynthesis protein IucC	downregulation with the log fold value -0.410

csgF	curli production assembly/transport protein	downregulation with the log fold value -1.006
csgG	curli production assembly/transport protein	downregulation with the log fold value -1.006
csgD	DNA-binding transcriptional regulator	Down regulation with the log fold value of -2.742
gspL	general secretion pathway protein L	upregulation with the log fold value of 0.843
gspM	general secretion pathway protein M	upregulation with the log fold value of 0.843
fimF	FimF protein precursor	upregulation with the log fold value of 0.702
fimG	FimG protein precursor	upregulation with the log fold value of 0.702
fimA	Type-1 fimbrial protein A chain precursor	Down regulation with the log fold value of -1.359
fimE	Type 1 fimbriae Regulatory protein	Down regulation with the log fold value of -1.359
fimI	Fimbrin-like protein fimI precursor	upregulation with the log fold value of 0.885
papF	P pilus minor subunit PapF	upregulation with the log fold value of 5.066
papX	PapX protein regulates flagellum synthesis to repress motility	Down regulation with the log fold value of -1.084
papK	P pilus minor subunit PapK	upregulation with the log fold value of 1.163
papB	regulatory protein PapB	upregulation with the log fold value of 4.009
papC	usher protein PapC	upregulation with the log fold value of 0.618
chuT	periplasmic heme-binding protein ChuT	upregulation with the log fold value of 0.618
chuV	ATP-binding hydrophilic protein ChuV	upregulation with the log fold value of 0.618
chuU	heme permease protein ChuU	upregulation with the log fold value of 0.618
chuY	ChuY [Chu (VF0227)[Escherichia coli CFT073]	upregulation with the log fold value of 0.618
chuX	putative heme-binding protein ChuX	upregulation with the log fold value of 0.618
chuW	Putative oxygen independent coproporphyrinogen III oxidase	upregulation with the log fold value of 0.618
chuS	(chuS) heme oxygenase ChuS	upregulation with the log fold value of 0.6326
ybtS	(ybtS) salicylate synthase Irp9	upregulation with the log fold value of 0.2903
ybtX	(ybtX) putative signal transducer	upregulation with the log fold value of 0.2903
ybtQ	(ybtQ) inner membrane ABC-transporter YbtQ	upregulation with the log fold value of 0.2903
ybtP	(ybtP) lipoprotein inner membrane ABC-transporter	upregulation with the log fold value of 0.2903
ybtA	(ybtA) transcriptional regulator YbtA	upregulation with the log fold value of 0.2903
irp2	(irp2) yersiniabactin biosynthetic protein Irp2	upregulation with the log fold value of 0.2903
irp1	(irp1) yersiniabactin biosynthetic protein Irp1 [Yersiniabactin	upregulation with the log fold value of 0.2903
ybtU	(ybtU) yersiniabactin biosynthetic protein YbtU	upregulation with the log fold value of 0.2903
ybtT	(ybtT) yersiniabactin biosynthetic protein YbtT	upregulation with the log fold value of 0.2903
ybtE	(ybtE) yersiniabactin siderophore	up regulation with the log fold value

	biosynthetic protein	of 0.2903
fyuA	(fyuA) pesticin/yersiniabactin receptor protein	up regulation with the log fold value of 0.2903
ompA	(ompA) outer membrane protein A	Down regulation with the log fold value of -2.556
kpsD	(kpsD) KpsD [K1 capsule]	up regulation with the log fold value of 0.4573
yagV/ecpE	(yagV/ecpE) E. coli common pilus chaperone EcpE	up regulation with the log fold value of 0.3651
fdeC	(fdeC) adhesin FdeC	up regulation with the log fold value of 0.6643
ykgK/ecpR	(ykgK/ecpR) regulator protein EcpR	up regulation with the log fold value of 0.4963
yagZ/ecpA	(yagZ/ecpA) E. coli common pilus structural subunit EcpA	up regulation with the log fold value of 0.4963
yagY/ecpB	(yagY/ecpB) E. coli common pilus chaperone EcpB	up regulation with the log fold value of 0.4963
yagX/ecpC	(yagX/ecpC) E. coli common pilus usher EcpC	up regulation with the log fold value of 0.4963
entS	(entS) enterobactin exporter iron-regulated [enterobactin	up regulation with the log fold value of 0.6994
fepB	(fepB) ferrienterobactin ABC transporter periplasmic binding protein	genes that do not show significant differential expression between the two groups with log fold value of -0.0643
entC	(entC) isochorismate synthase 1	genes that do not show significant differential expression between the two groups with log fold value of -0.0643
entE	(entE) 23-dihydroxybenzoate-AMP ligase component of enterobactin synthase multienzyme complex	genes that do not show significant differential expression between the two groups with log fold value of -0.0643
entB	(entB) isochorismatase	genes that do not show significant differential expression between the two groups with log fold value of -0.0643
entA	(entA) 23-dihydro-23-dihydroxybenzoate dehydrogenase	genes that do not show significant differential expression between the two groups with log fold value of -0.0643
entF	(entF) enterobactin synthase multienzyme complex component ATP-dependent	up regulation with the log fold value of 0.1486
fepA	(fepA) ferrienterobactin outer membrane transporter	Down regulation with the log fold value of -0.2825
fes	(fes) enterobactin/ferric enterobactin esterase	Down regulation with the log fold value of -0.4253
fepC	(fepC) ferrienterobactin ABC transporter ATPase	up regulation with the log fold value of 0.1372
fepG	(fepG) iron-enterobactin ABC transporter permease	up regulation with the log fold value of 0.1372
fepD	(fepD) ferrienterobactin ABC transporter permease	up regulation with the log fold value of 0.1372

Gene Hub Analysis

Materials:

- Gene List (e.g differentially expressed genes obtained from RNA-sequencing) (DEG)
- String database (version 12.0)
- Microsoft Excel
- Cytoscape software (version 3.10.3)
- cytoHubba (version 0.1) (install within Cytoscape)

Methods:

1. Prepare gene list - Obtained a list of genes of interest from RNA-sequencing data.
2. String database analysis - Open string database website (version 12.0)
- The gene list was uploaded to the String database, and *Klebsiella pneumoniae* was selected as the organism.

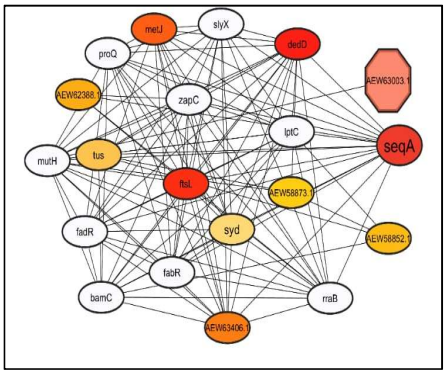
- The database was then queried using the uploaded gene list to retrieve protein-protein interaction data.
- Then downloaded the file as short tubular text output:
- TSV: tab separated values - Can be opened in Excel and Cytoscape (lists only one way edges: A-B)

3. Formatting in Microsoft Excel - The obtained gene list file from String database was formatted into a 2-column table (node 1 and node 2) using Microsoft Excel and saved as a CSV file for further analysis.
4. Cytoscape Analysis - Cytoscape software (version 3.10.3) was installed. Cytoscape visualization settings were configured as follows: network format set to BioPax_SIF, node shape set to ellipse, font size to 20, node fill color customized according to each node attributes. And border width set to 5.0.
5. cytoHubba Analysis - cytoHubba (version 0.1) was installed within Cytoscape. This software was utilized to identify gene hub nodes in a biological network. Node scores were calculated using the Maximal Clique Centrality (MCC) method, and the top 15 hub nodes were selected. The analysis was configured to display the first-stage nodes, shortest paths and expanded subnetwork. The network visualization was customized with node fill colors and arranged for clarity.

- The network was then exported as a PNG image file named 'gene hub network'. The export settings included a 500% zoom, 2925*1840 pixels, and 'show all graphics details' option, ensuring a high-quality image with detailed graphics.

Results:

Based on the results of transcriptome analysis, we found the virulence activity of the pathogenic *K. pneumonia* also the genes responsible for it. 63 genes were responsible for the virulence activity. On the basis of DEG analysis, we found the upregulated and downregulated genes. We identified significant upregulation and downregulation of genes, providing insights into metabolic adaptations, particularly in response to high glucose conditions. The top upregulated virulence gene was *yagZ/ecpA*, functioning as a *Klebsiella* common pilus structural subunit, exhibited a high (log fold value) of 13.003 compared to the control sample with no glucose concentration. The lowest downregulated genes were *csgF* and *csgG*, functioning as curli production assembly/transport proteins, with the log fold value of -1.006.



AMR – antimicrobial resistance

The KEGG database is widely used as a reference for biological interpretation of genome sequences of high-output data through processes called pathway mapping (KEGG pathway maps).

Information eroded in genome genes is converted to high-level functional information general methodology applied to microbiology genomes to infer AMR.

AMR methods

Current methods of determining AMR

↓

Phenotypic approaches

↓

Pathogen antimicrobial combinations

↓

Bacterial isolates from different species have contaminated AMR for multiple antibiotics [ex, carbapenem resistance]

↓

Identification of the AMR gene

↓

Identification of Virulence Factor Genes

↓

The virulence characteristics of a virus dictate its capacity to infect or harm the tissues of its host

↓

These are compounds that are often produced by bacteria or viruses and stored in their genomes, but they can also be ingested from the environment by means of genetic elements that are transmissible

↓

The genes responsible for the virulence activity of the organism are called as virulence factor genes

Genes	Product	Gene expression	RESISTANCE
bacA	The bacA gene product (BacA) recycles undecaprenyl pyrophosphate during cell wall biosynthesis which confers resistance to bacitracin.	upregulation with the log fold value of 8.973	Peptide
Klebsiella_pneumoniae_KpnF	KpnF subunit of KpnEF resembles EbrAB from E. coli	upregulation with the log fold value of 7.662	aminoglycoside; cephalosporin; macrolide;peptide;rifamycin;tetracycline
Klebsiella_pneumoniae_KpnE	Mutation in KpnEF resulted in increased susceptibility to cefepime ceftriaxon	upregulation with the log fold value of 7.662	aminoglycoside; cephalosporin; macrolide;peptide;rifamycin; tetracycline
	colistin erythromycin rifampin tetracycline and streptomycin		
FosA5	It is susceptible to amikacin tetracycline and imipenem and resistant to sulphonamide cephalosporins gentamicin ciprofloxacin chloramphenicol and streptomycin	upregulation with the log fold value of 7.485	fosfomycin
marA	the global activator protein MarA which besides inducing MDR efflux pump AcrAB also down-regulates synthesis of the porin OmpF.	upregulation with the log fold value of 9.294	ephamycin;fluoroquinolone; glycylcycline;monobactam;penam; penem;phenicol;rifamycin;tetracycline; t riclosan
CTX-M-15	CTX-M-15 is a beta-lactamase found in the Enterobacteriaceae family	downregulation with the log fold value -2.134	Cephalosporin
mdtB	MdtB is a transporter that forms a heteromultimer complex with MdtC to form a multidrug transporter. MdtBC is part of the MdtABC-TolC efflux complex.	upregulation with the log fold value of 8.384	Aminocoumarin
mdtA	MdtA is the membrane fusion protein of the multidrug efflux complex mdtABC.	downregulation with the log fold value -0.075	aminocoumarin
baeR	BaeR is a response regulator that promotes the expression of MdtABC and AcrD efflux complexes.	upregulation with the log fold value of 7.531	aminocoumarin;aminoglycoside
acrB	Protein subunit of AcrA-AcrB-TolC multidrug efflux complex.	downregulation with the log fold value -0.751	cephalosporin;fluoroquinolone; glycylcycline;penam;phenicol;rifamycin ;tetracycline;triclosan
emrB	emrB is a translocase in the emrB -TolC efflux protein in E. coli.	downregulation with the log fold value -0.507	Fluoroquinolone
emrA	EmrA is a membrane fusion protein providing an efflux pathway with EmrB and TolC between the inner and outer membranes of E. coli a Gram-negative bacterium.	downregulation with the log fold value -0.507	Fluoroquinolone
_KpnH	KpnH consists of ~511 residues resembles EmrB of E. coli and is probably a translocase in the KpnGH-TolC efflux protein in K.	upregulation with the log fold value of 11.041	aminoglycoside;carbapenem ;cephalosporin;fluoroquinolone ;macrolide;penam;penem;peptide
KpnG	KpnG consists of ~390 residues and resembles EmrA of E. coli.	upregulation with the log fold value of 11.041	aminoglycoside;carbapenem;cephalosporin ;fluoroquinolone;macrolide;penam;penem; Peptide
emrR	EmrR is a negative regulator for the EmrAB-TolC multidrug efflux pump in E. coli.	downregulation with the log fold value -0.219	fluoroquinolone
	Mutations lead to EmrAB-TolC overexpression.		
AAC(6')-Ib-cr	AAC(6')-Ib-cr is an aminoglycoside acetyltransferase encoded by plasmids transposons integrons in Enterobacteriaceae.	downregulation with the log fold value -4.017	aminoglycoside;fluoroquinolone
OKP-B-3	OKP-B-3 is a beta-lactamase found in Klebsiella pneumoniae	upregulation with the log fold value of 5.677	cephalosporin;penam
cpxA	CpxA is a membrane-localized sensor kinase that is activated by envelope stress.	downregulation with the log fold value -1.460	aminocoumarin;aminoglycoside
AAC(3)-IIe	AAC(3)-IIe is a plasmid-encoded aminoglycoside acetyltransferase in E. coli	downregulation with the log fold value -1.874	Aminoglycoside
ANT(2'')-Ia	Plasmid or integron-encoded nucleotidylylation of 2-deoxystreptamine aminoglycosides at the hydroxyl group at position 2" in P. aeruginosa K. pneumoniae Morganella morganii E. coli	upregulation with the log fold value of 8.7620	Aminoglycoside
OXA-2	OXA-2 is a beta-lactamase found in the Enterobacteriaceae family.	upregulation with the log fold value of 7.214	carbapenem;cephalosporin;penam
sul1	Sul1 is a	upregulation	sulfonamide

AMR Genes are the genes, which confer antibiotic resistance, typically have minimal impact on the overall fitness of the host cell. For instance, a large number of these genes do not change the target site; instead, they actively kill the antibiotic. Additionally, certain resistance genes are actually made to express themselves by the antibiotic, which means that maintaining the gene costs the organism even less. Based on the results of AMR analysis, we identified 40 genes that were responsible for the antimicrobial activity. The gene which showed the highest upregulation was *mphK* with log fold value of 11.102. This gene was found to be resistant to the antibiotic macrolide. The gene that showed the lowest level of downregulation was *mdtA*, with the log fold value of -0.075. This gene was recognised as resistant to aminocoumarin

The KEGG pathway map is the most intuitive database display results in analysis results of the transcriptome. In transcriptome analysis, 1000 or 10,000 genes are involved. To classify genes and to analyse genes with the same function together.

These are the KEGG results with annotation

Trinity	KO
TRINITY_DN204324_c0_g6_i2	K00845
TRINITY_DN162161_c0_g1_i2	K01810
TRINITY_DN194794_c0_g1_i1	K00850
TRINITY_DN197706_c8_g2_i2	K00330
TRINITY_DN221532_c0_g1_i1	K03885
TRINITY_DN130575_c0_g1_i1	K00239
TRINITY_DN205059_c0_g3_i1	K01126
TRINITY_DN158954_c0_g1_i1	K05939
TRINITY_DN165877_c0_g3_i1	K04019
TRINITY_DN200630_c2_g4_i2	K01515
TRINITY_DN200261_c6_g1_i1	K08312
TRINITY_DN182015_c0_g2_i1	K01839
TRINITY_DN189421_c0_g1_i1	K00812
TRINITY_DN121180_c0_g1_i1	K00278
TRINITY_DN205107_c2_g5_i1	K01779
TRINITY_DN204530_c0_g1_i1	K01579
TRINITY_DN132092_c0_g1_i1	K00823
TRINITY_DN206306_c0_g1_i1	K15372
TRINITY_DN171795_c0_g1_i1	K01207
TRINITY_DN203606_c1_g3_i3	K07406
TRINITY_DN171015_c0_g1_i2	K04487
TRINITY_DN203049_c3_g5_i2	K03151
TRINITY_DN204186_c4_g2_i3	K05349
TRINITY_DN193517_c0_g1_i1	K01031
TRINITY_DN201778_c2_g1_i2	K02878
TRINITY_DN205215_c18_g2_i2	K02819

This gene classification can be achieved through gene annotation. Transcriptome analysis obtained by referencing information in the genome. For the unreferenced transcriptome, it is obtained by comparing with specific data bases KEGG database, expressed genes.

KEGG analysis in biological research, gene function, biological pathways, molecular mechanisms, study of genes with functions, pathways, and databases.

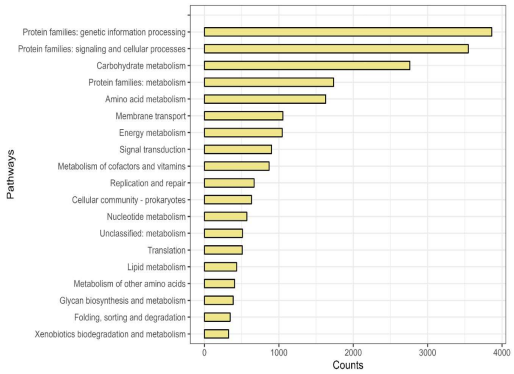
Pathway Analysis

Pathway
Glycolysis / Gluconeogenesis [PATH:ko00010]
Glycolysis / Gluconeogenesis [PATH:ko00010]
Glycolysis / Gluconeogenesis [PATH:ko00010]
Oxidative phosphorylation [PATH:ko00190]
Oxidative phosphorylation [PATH:ko00190]
Oxidative phosphorylation [PATH:ko00190]
Glycerophospholipid metabolism [PATH:ko00564]
Glycerophospholipid metabolism [PATH:ko00564]
Glycerophospholipid metabolism [PATH:ko00564]
Purine metabolism [PATH:ko00230]
Purine metabolism [PATH:ko00230]
Purine metabolism [PATH:ko00230]
Alanine, aspartate and glutamate metabolism [PATH:ko00250]
Alanine, aspartate and glutamate metabolism [PATH:ko00250]
Alanine, aspartate and glutamate metabolism [PATH:ko00250]
beta-Alanine metabolism [PATH:ko00410]
beta-Alanine metabolism [PATH:ko00410]
beta-Alanine metabolism [PATH:ko00410]
Glycosaminoglycan degradation [PATH:ko00531]
Glycosphingolipid biosynthesis - globo and isoglobo series [PATH:ko00603
Thiamine metabolism [PATH:ko00730]
Thiamine metabolism [PATH:ko00730]
Biosynthesis of various plant secondary metabolites [PATH:ko00999]
Benzoate degradation [PATH:ko00362]
Ribosome [PATH:ko03010]
Phosphotransferase system (PTS) [PATH:ko02060]

Description
glk; glucokinase [EC:2.7.1.2]
GPI, pgi; glucose-6-phosphate isomerase [EC:5.3.1.9]
pfkA, PFK; 6-phosphofructokinase 1 [EC:2.7.1.11]
nuoA; NADH-quinone oxidoreductase subunit A [EC:7.1.1.2]
ndh; NADH:quinone reductase (non-electrogenic) [EC:1.6.5.9]
sdhA, frdA; succinate dehydrogenase flavoprotein subunit [EC:1.3.5.1]
E3.1.4.46, glpQ, ugpQ; glycerophosphoryl diester phosphodiesterase [EC:3.1.4.46
aas; acyl-[acyl-carrier-protein]-phospholipid O-acyltransferase / long-chain-fatty
eutA; ethanolamine utilization protein EutA
nudF; ADP-ribose diphosphatase [EC:3.6.1.13 3.6.1.-]
nudE; ADP compounds hydrolase [EC:3.6.1.-]
deoB; phosphopentomutase [EC:5.4.2.7]
aspB; aspartate aminotransferase [EC:2.6.1.1]
nadB; L-aspartate oxidase [EC:1.4.3.16]
racD; aspartate racemase [EC:5.1.1.13]
panD; aspartate 1-decarboxylase [EC:4.1.1.11]
puuE; 4-aminobutyrate aminotransferase [EC:2.6.1.19]
toa; taurine---2-oxoglutarate transaminase [EC:2.6.1.55]
nagZ; beta-N-acetylhexosaminidase [EC:3.2.1.52]
melA; alpha-galactosidase [EC:3.2.1.22]
iscS, NFS1; cysteine desulfurase [EC:2.8.1.7]
thiI; tRNA uracil 4-sulfurtransferase [EC:2.8.1.4]
bglX; beta-glucosidase [EC:3.2.1.21]
pcaI; 3-oxoadipate CoA-transferase, alpha subunit [EC:2.8.3.6]
RP-L16, MRPL16, rplP; large subunit ribosomal protein L16
treB, treP; trehalose PTS system EIIBC or EIIBCA component [EC:2.7.1.201]

Pathway Function	Major Group
Carbohydrate metabolism	Metabolism
Carbohydrate metabolism	Metabolism
Carbohydrate metabolism	Metabolism
Energy metabolism	Metabolism
Energy metabolism	Metabolism
Energy metabolism	Metabolism
Lipid metabolism	Metabolism
Lipid metabolism	Metabolism
Lipid metabolism	Metabolism
Nucleotide metabolism	Metabolism
Nucleotide metabolism	Metabolism
Nucleotide metabolism	Metabolism
Amino acid metabolism	Metabolism
Amino acid metabolism	Metabolism
Amino acid metabolism	Metabolism
Metabolism of other amino acids	Metabolism
Metabolism of other amino acids	Metabolism
Metabolism of other amino acids	Metabolism
Glycan biosynthesis and metabolism	Metabolism
Glycan biosynthesis and metabolism	Metabolism
Metabolism of cofactors and vitamins	Metabolism
Metabolism of cofactors and vitamins	Metabolism
Biosynthesis of other secondary metabolites	Metabolism
Xenobiotics biodegradation and metabolism	Metabolism
Translation	Genetic Information Processing
Membrane transport	Environmental Information Proc

Pathway assignment and mapping of the CDS to the biological pathways were performed using KEGG automatic annotation server (KAAS)

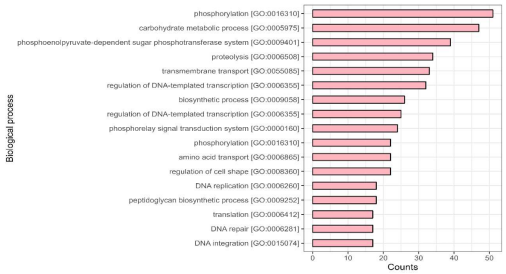


Gene ontology molecular function annotation

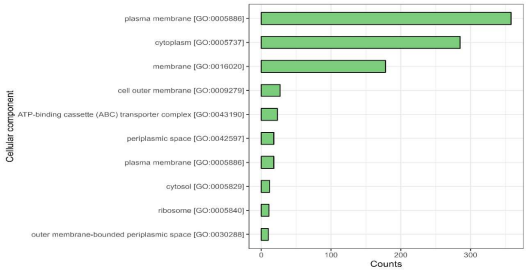
Gene Ontology (molecular function)
ATP binding [GO:0005524]; ATP hydrolysis activity [GO:0016887]; DNA binding [GO:0003677]; DNA helicase activity [GO:0003678]
nucleotide binding [GO:0000166]
ATP binding [GO:0005524]; glutamine synthetase activity [GO:0004356]
DNA binding [GO:0003677]; type II site-specific deoxyribonuclease activity [GO:0009036]
hydrolase activity, acting on ester bonds [GO:0016788]
siderophore uptake transmembrane transporter activity [GO:0015344]; signaling receptor activity [GO:0038023]
long-chain fatty acid-CoA ligase activity [GO:0004467]
hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]; metal ion binding [GO:0046872]; oxidoreductase activity, actir
4 iron, 4 sulfur cluster binding [GO:0051539]; aconitate hydratase activity [GO:0003994]; iron-responsive element binding [GO:003(
myo-inositol:proton symporter activity [GO:0005366]
mRNA binding [GO:0003729]; rRNA binding [GO:0019843]; structural constituent of ribosome [GO:0003735]
inositol 2-dehydrogenase activity [GO:0050112]; nucleotide binding [GO:0000166]
ribose-5-phosphate isomerase activity [GO:0004751]
GTP binding [GO:0005525]; GTPase activity [GO:0003924]; guanosine tetraphosphate binding [GO:0097216]; translation elongation
protein-N(PI)-phosphohistidine-sugar phosphotransferase activity [GO:0008982]
2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase activity [GO:0043719]; FAD binding [GO:0071949]; pentachlorophe
succinate-semialdehyde dehydrogenase (NAD+) activity [GO:0004777]
kinase activity [GO:0016301]; protein-N(PI)-phosphohistidine-sugar phosphotransferase activity [GO:0008982]; protein-phosphocy
allantoinase activity [GO:0004038]; cobalt ion binding [GO:0050897]; zinc ion binding [GO:0008270]

Gene ontology

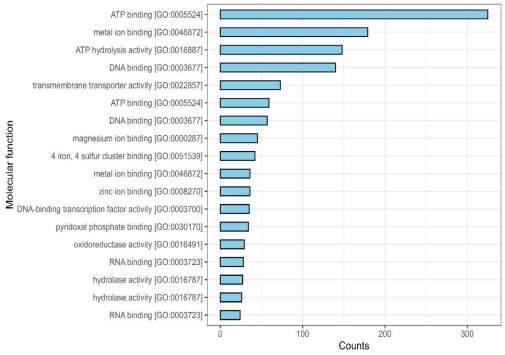
GO assignments were used to classify the functions of the transcripts. The GO mapping provides an ontology of defined terms representing gene product properties, which are grouped into three main domains: Biological Process, Molecular Function and Cellular Component.



Biological process



Cellular component



Molecular function

Gene Ontology (cellular component)
cytosol [GO:0005829]; primosome complex [GO:1990077]
cytoplasm [GO:0005737]; membrane [GO:0016020]
cell outer membrane [GO:0009279]
membrane [GO:0016020]
cytosol [GO:0005829]
apical plasma membrane [GO:0016324]
cytosolic small ribosomal subunit [GO:0022627]
cytosol [GO:0005829]
plasma membrane [GO:0005886]
ATP-binding cassette (ABC) transporter complex [GO:0043190]
cytosol [GO:0005829]
membrane [GO:0016020]
plasma membrane [GO:0005886]
cytoplasm [GO:0005737]

Gene ontology cellular component annotation

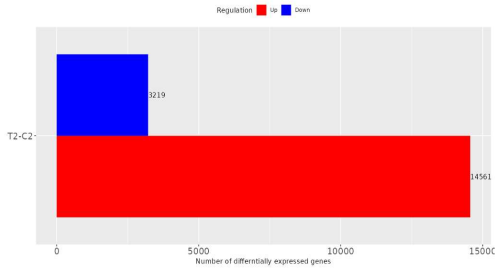
High glucose levels are known to promote the virulence of certain bacteria, including *K. pneumoniae.* This study aimed to identify genes and pathways involved in this process. Under high glucose conditions, *Klebsiella pneumoniae's* transcriptome study revealed an intricate response that includes both virulence and antibiotic resistance. The formation of antibiotic-resistant phenotypes may be aided by high glucose environments, as seen by the upregulation of genes linked to antimicrobial resistance, such as beta-lactamases and efflux pumps. This is concerning because it may decrease the number of treatments available for *K. pneumoniae* infections, especially for those who have diabetes or other disorders that are associated with elevated blood glucose levels.

In a bioinformatics lab, data quality control (QC) for pathogenic. *Klebsiella pneumoniae* samples contain an array of steps that ensure the accuracy and reliability of the data generated via sequencing or other methods of analysis.

Pathogenic *Klebsiella pneumoniae* Samples: C2, T2

Biological QC: Perform additional checks about the biology of *Klebsiella pneumoniae* or the objectives of the investigation. For instance, utilise specialised databases and tools (e.g., ResFinder, CARD) to validate the presence of known virulence factors or genes associated with antibiotic resistance.

Comparisons - Comparison of C2 vs T2



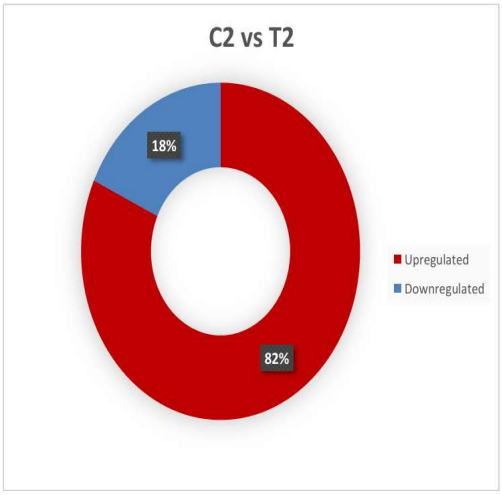
Number of differentially expressed genes

Note: DEGs have been filtered using a minimum fold change threshold of 2 and an adjusted p-value threshold of 0.05.

A low p-value (typically < 0.05) suggests that the observed change in expression is statistically significant and not due to random variation.

A high p-value suggests that the observed change might be due to chance. Typically, such a file includes several columns, among which logFC and p-value are key metrics. Here's a detailed explanation of these terms and how they are interpreted

The organism column depicts the result based on blast result against bacterial database so our particular transcript matched with the gene present different bacterial species.



Pictorial representation of C2 vs T2 (Based on DEG)

Fold change is calculated by dividing the expression level of a gene in one condition by its expression level in another condition. The logarithm (usually base 2, but sometimes base 10 or natural log) of this fold change is then taken.

For example, if a gene's expression is twice as high in the treated condition as in the control, the fold change is 2, and the log2 fold change (logFC) would be log2(2) = 1.

Positive logFC values indicate upregulation (higher expression in the treated condition).

Negative logFC values indicate downregulation (lower expression in the treated condition). A logFC of 1 means the expression is doubled, -1 means the expression is halved.

The p-value indicates the statistical significance of the observed changes in gene expression.

Calculation: It is derived from a statistical test (e.g., t-test, Wald test, etc.) that compares the expression levels between the two conditions.

Interpretation:

A low p-value (typically < 0.05) suggests that the observed change in expression is statistically significant and not due to random variation. A high p-value suggests that the observed change might be due to chance.

DEG results for the upregulated genes

	A	B	C	D
1	Trinity ID	Uniprot ID	log2FoldChange	padj
2	TRINITY_DN199298_C5_G1_I1	A0A6M4MJJ0	14.85495499	9.69E-23
3	TRINITY_DN204944_C0_G2_I11	A0A3N2RUZ2	12.64444488	9.62E-17
4	TRINITY_DN205211_C1_G2_I1	A0A378GJ24	12.49763315	2.34E-16
5	TRINITY_DN205199_C12_G6_I1	A0A6P1V6L2	12.09822167	2.48E-15
6	TRINITY_DN200848_C0_G1_I1	#N/A	12.03909958	3.53E-15
7	TRINITY_DN204812_C1_G2_I1	A0A6M4MDA7	11.86135021	9.88E-15
8	TRINITY_DN205080_C14_G1_I1	#N/A	11.76781262	1.69E-14
9	TRINITY_DN201950_C7_G1_I3	#N/A	11.54725749	5.87E-14
10	TRINITY_DN204055_C40_G1_I1	#N/A	11.54585376	1.86E-14
11	TRINITY_DN205182_C8_G2_I4	A0A3N2EVQ6	11.43388486	3.45E-14
12	TRINITY_DN198809_C0_G1_I1	#N/A	11.41465034	1.24E-13
13	TRINITY_DN205074_C3_G9_I1	A0A6L6Y6L2	11.31164665	2.23E-13
14	TRINITY_DN203801_C1_G3_I2	A0AAE3UW90	11.29435398	2.44E-13
15	TRINITY_DN205161_C10_G1_I2	#N/A	11.27363319	2.73E-13
16	TRINITY_DN204944_C0_G1_I3	A0A0H3H6S9	11.26272965	2.93E-13
17	TRINITY_DN119606_C0_G1_I1	#N/A	11.23249394	3.44E-13
18	TRINITY_DN205057_C1_G1_I11	A0A6L6Y3V2	11.15897918	5.17E-13
19	TRINITY_DN202742_C1_G2_I2	#N/A	11.15004096	5.43E-13
20	TRINITY_DN204465_C0_G1_I1	#N/A	11.14104703	5.71E-13
21	TRINITY_DN205160_C0_G7_I3	A0AAE2NEJ0	11.13620416	5.86E-13
22	TRINITY_DN205189_C11_G1_I3	A0A6L6Y7S1	11.12647116	6.15E-13
23	TRINITY_DN205179_C1_G4_I3	A0A0A5SFT1	11.07140499	8.32E-13
24	TRINITY_DN205035_C1_G1_I4	A0A514EWT4	11.05303113	9.27E-13
25	TRINITY_DN203867_C0_G5_I4	#N/A	11.0427177	9.77E-13

Gene Annotation

	Protein names	Gene Names	Organism	Length
2	Replicative DNA helicase (EC 3.6.4.12)	dnaB CA267_011575	Alteromonas pelagimontana	462
3	Gfo/Idh/MocA family oxidoreductase	EB837_19135	Kluyvera ascorbata	431
4	Glutamine synthetase (EC 6.3.1.2)	glnA EB837_13580 KATP_44760	Kluyvera ascorbata	469
5	Restriction endonuclease	GW952_31205	Klebsiella michiganensis	404
6	Rhamnogalacturonan acetylesterase	CA267_010155	Alteromonas pelagimontana	268
7	Iron complex outermembrane receptor protein	EDF82_1229	Raoultella sp. BIGb0399	720
8	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3) (L	fadD GO007_16600	Raoultella sp. 10-1	561
9	GntP family permease	QUS25_04950	Klebsiella aerogenes (Enterobacter aero	463
10	Alpha-galactosidase	KOX_16445	Klebsiella michiganensis (strain ATCC 87	435
11	Aconitate hydratase (Aconitase) (EC 4.2.1.3)	acnA GO007_06530	Raoultella sp. 10-1	890
12	PRD domain-containing protein	HBL31_04825	Klebsiella quasipneumoniae	278
13	MFS transporter	GO007_14790	Raoultella sp. 10-1	475
14	Small ribosomal subunit protein uS3	rpsC CKG00_11750 CYG68_19455 N0392_19445	Morganella morganii (Proteus morganii)	233
15	Inositol 2-dehydrogenase (EC 1.1.1.18) (Myo-in	iolG electrica_04533	Klebsiella electrica	337
16	Uncharacterized protein	NUKP37_52170 NUKP37_52180	Klebsiella variicola	63
17	Ribose-5-phosphate isomerase B (EC 5.3.1.6)	KOX_08520	Klebsiella michiganensis (strain ATCC 87	149
18	Elongation factor G (EF-G)	fusA CA267_011850	Alteromonas pelagimontana	700
19	Permease IIC component	celB EB837_06305	Kluyvera ascorbata	445
20	Metal ABC transporter permease	CYG68_05835 OSC06_10975	Morganella morganii (Proteus morganii)	284
21	Pentachlorophenol 4-monooxygenase (EC 1.14.	pcpB electrica_00701	Klebsiella electrica	501
22	ABC transporter ATP-binding protein	EGY08_24975	Klebsiella sp. FDAARGOS_511	278
23	Succinate-semialdehyde dehydrogenase (EC 1.2	gabD AO703_17205	[Enterobacter] lignolyticus	482
24	DMT family transporter	HV064_20440 HV104_18240 KOSB73_270002	Klebsiella grimontii	298
25	Negative regulator of SacY activity	sacX SB6408_02377	Klebsiella spallanzanii	456

Upregulated genes and their protein names

Virulence factors

	A	B	C	D	E	F	G	H	I	J	K	L
1	SEQUENCE	START	END	STRAND	GENE	COVERAG	COVERAG	GAPS	%COVERA	%IDENTIT	DATABASE	ACCESSIO
2	TRINITY_DN105774_c0_g1_i1	143	448	+	pilH	1-306/36(	========	0/0	83.61	84.31	vfdb	NP_24910
3	TRINITY_DN200082_c0_g1_i2	1	1152	+	fes	27-1178/1	========	0/0	95.76	98.96	vfdb	NP_75260
4	TRINITY_DN200798_c1_g1_i1	3	521	+	bslA/yuaB	1-519/54(	========	0/0	95.05	99.42	vfdb	NP_39098
5	TRINITY_DN201233_c1_g1_i5	1232	1976	-	yagV/ecpI	10-754/75	========	0/0	98.54	97.99	vfdb	NP_28600
6	TRINITY_DN201233_c1_g5_i1	502	2675	-	yagX/ecpC	353-2526/	========	0/0	86.06	97.52	vfdb	NP_28600
7	TRINITY_DN201929_c4_g3_i1	526	1377	-	entB	1-858/858	========	02-Jun	99.3	82.05	vfdb	NP_75261
8	TRINITY_DN202118_c0_g2_i3	242	663	-	gspM	1-422/423	========	0/0	99.76	92.42	vfdb	YP_40460
9	TRINITY_DN202118_c0_g2_i3	779	1638	-	gspL	1-861/861	========	01-Jan	99.88	94.66	vfdb	YP_40460
10	TRINITY_DN202399_c0_g1_i2	35	586	-	papX	1-552/552	========	0/0	100	99.09	vfdb	NP_75545
11	TRINITY_DN202399_c0_g1_i6	92	620	-	papX	1-529/552	========	0/0	95.83	99.81	vfdb	NP_75545
12	TRINITY_DN204013_c0_g6_i3	2	644	-	entD	1-643/771	========	0/0	83.4	97.51	vfdb	NP_75259
13	TRINITY_DN204455_c1_g10_i2	1173	2664	-	kpsD	186-1677/	========	0/0	88.97	99.13	vfdb	AAA21682
14	TRINITY_DN204535_c1_g3_i4	1	453	+	chuX	43-495/45	========	0/0	91.52	100	vfdb	NP_75617
15	TRINITY_DN204535_c1_g3_i4	453	1076	+	chuY	1-624/624	========	0/0	100	98.24	vfdb	NP_75617
16	TRINITY_DN204535_c1_g3_i4	1125	2117	+	chuU	1-993/993	========	0/0	100	99.5	vfdb	NP_75617
17	TRINITY_DN204535_c1_g3_i4	2084	2884	+	chuV	1-801/801	========	0/0	100	99.75	vfdb	NP_75618
18	TRINITY_DN204535_c1_g6_i2	621	1649	+	chuS	1-1029/1(	========	0/0	100	99.03	vfdb	NP_75616
19	TRINITY_DN204535_c1_g7_i3	346	1528	-	chuW	142-1325/	========	01-Jan	88.42	98.73	vfdb	NP_75617
20	TRINITY_DN204572_c1_g3_i2	250	789	-	papK	1-540/54(	========	0/0	100	100	vfdb	NP_75546
21	TRINITY_DN204572_c1_g4_i4	146	379	-	papI	1-234/234	========	0/0	100	95.73	vfdb	NP_75546
22	TRINITY_DN204611_c2_g1_i1	283	879	+	fimE	1-597/597	========	0/0	100	98.83	vfdb	NP_75724
23	TRINITY_DN204611_c2_g1_i1	1302	1906	+	fimA	1-606/60(	========	01-Jan	99.83	90.76	vfdb	NP_75724
24	TRINITY_DN204611_c2_g1_i1	1971	2510	+	fimI	1-540/54(	========	0/0	100	98.33	vfdb	NP_75724
25	TRINITY_DN204611_c2_g1_i1	2547	3272	+	fimC	1-726/726	========	0/0	100	98.76	vfdb	NP_75724
26	TRINITY_DN204611_c2_g1_i1	3414	4093	-	fimC	1-680/726	========	0/0	93.66	98.53	vfdb	NP_75724
27	TRINITY_DN204611_c2_g1_i1	4130	4669	-	fimI	1-540/54(	========	0/0	100	98.33	vfdb	NP_75724
28	TRINITY_DN204611_c2_g1_i1	4734	5338	-	fimA	1-606/60(	========	01-Jan	99.83	90.76	vfdb	NP_75724
29	TRINITY_DN204611_c2_g1_i1	5761	6357	-	fimE	1-597/597	========	0/0	100	98.83	vfdb	NP_75724
30	TRINITY_DN204611_c2_g1_i5	283	879	+	fimE	1-597/597	========	0/0	100	98.83	vfdb	NP_75724
31	TRINITY_DN204611_c2_g1_i5	1302	1906	+	fimA	1-606/60(	========	01-Jan	99.83	90.76	vfdb	NP_75724
32	TRINITY_DN204611_c2_g1_i5	1971	2510	+	fimI	1-540/54(	========	0/0	100	98.33	vfdb	NP_75724
33	TRINITY_DN204611_c2_g1_i5	2547	3272	+	fimC	1-726/726	========	0/0	100	98.76	vfdb	NP_75724

Virulence factors annotation

PRODUCT	RESISTANCE
(pilH) twitching motility protein PilH [Type IV pili (VF0082)] [Pseudomonas aeruginosa PAO1]	
(fes) enterobactin/ferric enterobactin esterase [enterobactin (IA019)] [Escherichia coli CFT073]	
(bslA/yuaB) hydrophobin BslA [BslA (VF0411)] [Bacillus subtilis subsp. subtilis str. 168]	
(yagV/ecpE) E. coli common pilus chaperone EcpE [ECP (VF0404)] [Escherichia coli O157:H7 str. EDL933]	
(yagX/ecpC) E. coli common pilus usher EcpC [ECP (VF0404)] [Escherichia coli O157:H7 str. EDL933]	
(entB) isochorismatase [Enterobactin (VF0228)] [Escherichia coli CFT073]	
(gspM) general secretion pathway protein M [T2SS (VF0333)] [Shigella dysenteriae Sd197]	
(gspL) general secretion pathway protein L [T2SS (VF0333)] [Shigella dysenteriae Sd197]	
(papX) PapX protein regulates flagellum synthesis to repress motility [P fimbriae (CVF425)] [Escherichia coli CFT073]	
(papX) PapX protein regulates flagellum synthesis to repress motility [P fimbriae (CVF425)] [Escherichia coli CFT073]	
(entD) phosphopantetheinyl transferase component of enterobactin synthase multienzyme complex [Enterobactin	
(kpsD) KpsD [K1 capsule (VF0239)] [Escherichia coli O18:K1:H7 str. RS218]	
(chuX) putative heme-binding protein ChuX [Chu (VF0227)] [Escherichia coli CFT073]	
(chuY) ChuY [Chu (VF0227)] [Escherichia coli CFT073]	
(chuU) heme permease protein ChuU [Chu (VF0227)] [Escherichia coli CFT073]	
(chuV) ATP-binding hydrophilic protein ChuV [Chu (VF0227)] [Escherichia coli CFT073]	
(chuS) heme oxygenase ChuS [Chu (VF0227)] [Escherichia coli CFT073]	
(chuW) Putative oxygen independent coproporphyrinogen III oxidase [Chu (VF0227)] [Escherichia coli CFT073]	
(papK) P pilus minor subunit PapK [P fimbriae (VF0220)] [Escherichia coli CFT073]	
(papI) regulatory protein PapI [P fimbriae (VF0220)] [Escherichia coli CFT073]	
(fimE) Type 1 fimbriae Regulatory protein fimE [Type 1 fimbriae (VF0221)] [Escherichia coli CFT073]	
(fimA) Type-1 fimbrial protein A chain precursor [Type 1 fimbriae (VF0221)] [Escherichia coli CFT073]	
(fimI) Fimbrin-like protein fimI precursor [Type 1 fimbriae (VF0221)] [Escherichia coli CFT073]	
(fimC) Chaperone protein fimC precursor [Type 1 fimbriae (VF0221)] [Escherichia coli CFT073]	
(fimC) Chaperone protein fimC precursor [Type 1 fimbriae (VF0221)] [Escherichia coli CFT073]	
(fimI) Fimbrin-like protein fimI precursor [Type 1 fimbriae (VF0221)] [Escherichia coli CFT073]	
(fimA) Type-1 fimbrial protein A chain precursor [Type 1 fimbriae (VF0221)] [Escherichia coli CFT073]	

The PRODUCT column describes the functional product of the gene identified. This could be a protein or enzyme that the gene encodes.

The RESISTANCE column specifies the type of resistance conferred by the gene, often found when screening against databases related to antimicrobial resistance (AMR).

Antimicrobial Resistance Genes

GENE
ErmB
AAC(3)-IIe
tolC
mdtH
baeR
baeS
emrR
emrB
CTX-M-15
CRP
AAC(6')-Ib-cr
gadW
Escherichia_coli_ampC
kdpE
Escherichia_coli_acrA
msbA
mdtG
mdtA
H-NS
TEM-116
eptA
mdtE
cpxA
CRP
bacA
emrR
emrA
Klebsiella_pneumoniae_KpnH

The results of DEG analysis and AMR gene analysis helped us identify the genes responsible for virulence activity and the bacteria's resistance to various antibiotics.

PRODUCT	RESISTANCE
ErmB conf	lincosamide;macrolide;streptogramin
AAC(3)-IIe	aminoglycoside
TolC is a p	aminocoumarin;aminoglycoside;carbapenem;cephalosporin;cephamycin;fluoroquin
Multidrug	fluoroquinolone
BaeR is a r	aminocoumarin;aminoglycoside
BaeS is a s	aminocoumarin;aminoglycoside
EmrR is a r	fluoroquinolone
emrB is a t	fluoroquinolone
CTX-M-15	cephalosporin
CRP is a gl	fluoroquinolone;macrolide;penam
AAC(6')-Ib-	aminoglycoside;fluoroquinolone
GadW is ar	fluoroquinolone;macrolide;penam
A class C a	cephalosporin;penam
kdpE is a tr	aminoglycoside
AcrA is a s	cephalosporin;fluoroquinolone;glycylcycline;penam;phenicol;rifamycin;tetracycline;
MsbA is a	nitroimidazole
The MdtG	fosfomycin
MdtA is th	aminocoumarin
H-NS is a h	cephalosporin;cephamycin;fluoroquinolone;macrolide;penam;tetracycline
TEM-116 i	cephalosporin;monobactam;penam;penem
PmrC medi	peptide
MdtE is th	fluoroquinolone;macrolide;penam
CpxA is a n	aminocoumarin;aminoglycoside
CRP is a gl	fluoroquinolone;macrolide;penam
The bacA g	peptide
EmrR is a r	fluoroquinolone
EmrA is a r	fluoroquinolone
KpnH cons	aminoglycoside:carbapenem:cephalosporin:fluoroquinolone:macrolide:penam:pene

The PRODUCT column describes the functional product of the gene identified. The RESISTANCE column indicates the antimicrobial resistance.

The expression pattern of beta-lactam genes in resistant isolates was analysed in the test sample. The crucial step in the RT-PCR assay is the variance in the percentage of transcribed rRNA. The resistant isolates showed a difference in the level of expression of all beta-lactamase genes. Genes of ESBL classes were found to be highly resistant in isolates. The discovery of virulence factors, such as toxins, adhesins, and invasins, emphasises *K. pneumoniae's* capacity to cause serious illness in conditions with elevated glucose levels. The upregulation of these genes helps the bacteria invade and colonise host tissues more effectively, resulting in more serious and invasive infections.

The downregulation of genes linked to breakdown processes, like the fatty acid degradation and tricarboxylic acid cycle, implies that *K. pneumoniae* might be changing its metabolic priorities in response to the high glucose environment. The bacteria may be able to maximise energy production and survival in conditions with a lot of glucose because of this metabolic reprogramming.

The results of this investigation have significant implications for our comprehension of the transcriptome analysis of *K. pneumoniae* and the creation of efficient treatment plans.

The discovery of antibiotic resistance genes and virulence factors highlights the necessity for innovative antimicrobial drugs and tailored therapies that can interfere with these systems. Furthermore, the metabolic reprogramming seen in this work might present chances for the creation of cutting-edge antibiotic treatments that specifically target the metabolic weaknesses of *K. pneumoniae* few isolates where rRNA expression is comparable to control strains.

Discussion & Conclusion:

The observed decline in the growth of *Klebsiella pneumonia* with increasing glucose concentrations is a phenomenon that can be attributed to metabolic control mechanisms and the toxic by-products of glucose metabolism [6,7]. *Klebsiella*, when exposed to high concentrations of glucose in minimal medium, experiences a metabolic imbalance that leads to the secretion of acetate [8,9]. This acetate accumulation is detrimental as it can inhibit growth and reduce the overall viability and productivity of the bacteria [10]. Acetate is a known inhibitor of cellular processes, and its accumulation is a result of overflow metabolism, where excess carbon flux through glycolysis exceeds the capacity of downstream pathways [11].

The role of transcriptome, proteome, and metabolism analysis using KEGG pathway underscores the pivotal role in unravelling the complex biological processes [12].

The KEGG analysis facilitates a deeper knowledge of high-throughput sequencing data, scientific discovery and innovation [13].

Based on the KEGG pathway analysis results of the pathogenic *K. pneumonia*, we discovered the number of genes responsible for various metabolic features [14,15]. For example, 2768 genes were responsible for carbohydrate metabolism, and 1629 genes were responsible for amino acid metabolism.

References

1. Morphological, genomic and transcriptomic responses of Klebsiella pneumoniae to the last- line antibiotic colistin. Amy K. Cain, Christine J. Boinett, Lars Barquist, Janina Dordel, Maria Fookes, Matthew Mayho, Matthew J. Ellington, David Goulding, Derek Pickard, Ryan R. Wick, Kathryn E. Holt, Julian Parkhill & Nicholas R. Thomson
2. Klebsiella pneumoniae liver abscess in diabetic patients: association of glycemic control with the clinical characteristics. Yi-Tsung Lin 1, Fu-Der Wang, Ping-Feng Wu, Chang-Phone Fung

3. Screening and isolation of Klebsiella spp. From COPD patients for biochemical and genetic analysis unveiling its pathogenic relevance. Safeena Khanum, Talluri Rameshwari KR and Sumana K
4. Klebsiella pneumoniae OmpR facilitates lung infection through transcriptional regulation of key virulence factors. Authors: Axel B. Janssen, Vincent de Bakker, Rieza Aprianto, Vincent Trebosc, Christian Kemmer, Michel Pieren, Jan-Willem Veening
5. Upregulation of transcripts for metabolism in diverse environments is a shared response associated with survival and adaptation of Klebsiella pneumoniae in response to temperature extremes. S. Tripathy, R. Sen, S. K. Padhi, S. Mohanty & N. K. Maiti
6. Pulmonary Surfactant Promotes Virulence Gene Expression and Biofilm Formation in Klebsiella pneumoniae. Authors: Graham G. Willsey, Sebastian Ventrone, Kristin C. Schutz, Aaron M. Wallace, John W. Ribis, Benjamin T. Suratt, Matthew J. Wargo
7. Identification of the BolA Protein Reveals a Novel Virulence Factor in K. pneumoniae That Contributes to Survival in Host. Authors: Feiyang Zhang, Xiangjin Yan, Jiawei Bai, Li Xiang, Manlin Ding, Qin Li, Biying Zhang, Qinghua Liang, Yingshun Zhou
8. Effects of Space Environment on Genome, Transcriptome, and Proteome of Klebsiella pneumoniae. Author links open overlay panel: Yinghua Guo A, Jia Li AB, Jinwen Liu C, Tong Wang C, Yinhu Li C, Yanting Yuan C, Jiao Zhao C, De Chang A, Xiangqun Fang A, Tianzhi Li A, Junfeng Wang A, Wenkui Dai C, Chengxiang Fang D, Changting Liu A
9. Epidemiology and Virulence of Klebsiella pneumoniae Steven Clegg, Caitlin N. Murphy
10. Molecular Pathogenesis of Klebsiella Pneumoniae. Bei Li, Yuling Zhao, Changting Liu, Zhenhong Chen & Dongsheng Zhou
11. Overview of Transcriptomic Research on Type 2 Diabetes: Challenges and Perspectives. Ziravard N. Tonyan, Yulia A. Nasykhova, Maria M. Danilova, Yury A. Barbitoff, Anton I. Changalidi, Anastasiia A. Mikhailova and Andrey S. Glotov
12. Whole transcriptome RNA-seq reveals key regulatory factors involved in type 2 diabetes pathology in peripheral fat of Asian Indians. Aditya Saxena, Nitish Mathur, Pradeep Tiwari & Sandeep Kumar Mathur
13. Genome-Wide Transcriptome Analysis in Type 2 Diabetes Patients Treated by Sitagliptin Rui Ma, Xiao-long Deng, Qi-qi-ge Aleteng, Lei Li, Jun Zhu
14. Transcriptome Sequencing Analysis of Peripheral Blood of Type 2 Diabetes Mellitus Patients With Thirst and Fatigue. Bohan Lv, Xueli Bao, Ping Li, Juan Lian, Yanxiang Wu, Tian An, Jing Zhang, Xiuyan Yang, Tingye Wang, Jiajian Zhu
15. Upregulation of transcripts for metabolism in diverse environments is a shared response associated with survival and adaptation of Klebsiella pneumoniae in response to temperature extremes. S. Tripathy, R. Sen, S. K. Padhi, S. Mohanty & N. K. Maiti