Bioscene

Bioscene
Volume- 22 Number- 04
ISSN: 1539-2422 (P) 2055-1583 (O)
www.explorebioscene.com

Whole Genome Sequencing Analysis of Streptococcus Pneumoniae

Syeda Sumaiyya Fatima [1&2], Maryam Tanzeel[1&2], Khadijah Al Khadir[2]
Ayla Sanjay[1] Boda Akhil[1] and Dr. Chand Pasha[1]*
[1]Department of Microbiology, Nizam college, Osmania University, Hyderabad, Telangana, India.
[2]Department of Microbiology, Mumtaz Degree and PG college, Malakpet, Hyderabad, India

Corresponding Author: **Dr. Chand Pasha**

Abstract: Majority of pneumococcal diseases include bronchitis, rhinitis, acute sinusitis, otitis media, conjunctivitis, meningitis are caused by *Streptococcus Pneumoniae.* The techniques like PCR, rt-PCR, qPCR are time consuming with low accuracy, they also possess the risk of showing false positive in some cases which makes Whole genome sequencing a much better alternative for the accurate analysis and its quick results. In this study, conducted the genetic relevance and clinical characteristics of *S. Pneumoniae*. 5 different strains of *S. Pneumoniae* were isolated from the Sputum and lungwash samples of RICU patients. The strain which is found to be extremely resistant is selected for the whole genome sequencing and analysis. Whole genome sequencing revealed in depth geneticcharacteristics including the coding sequences, open reading frames (ORFs), antibiotic resistance genes, pathogenicity etc. Analysis of antibiotic resistance genes highlighted multidrug resistance. In 124 contigs found 2,225,710 bases, 2245 coding sequences, 14 rRNA, 61 tRNA, 1 tm RNA, 5955 open reading frames, 274 mobile elements and 0 plasmids. The findings also include 1 intact prophage region along with 1 viral signal.The best hit ARO (Antibiotic Resistance Ontology) were reported to be vanY gene in vanM cluster, pmrA, patB,patA,RlmA(II). In Antibiotic susceptibility assay the strain was found to be resistant to majority of the tested Antibiotics and it was found to be sensitive to Augmentin, Monobactam, Vancomycin and moderately sensitive to Cefixime. The Pathogenicity of the strain was reported to be 88.3% which is pathogenic to humans.

Keywords: Streptococcus pneumoniae, whole genome sequencing, prophage regions, Antibiotic resistance, pathogenicity

1. Introduction:

Streptococcus pneumoniae (pneumococcus) is a leading cause for many upper and lower respiratory tract infections. While it is often commensal, it is an also a dominant pathogen capable of causing severe diseases, particularly in vulnerable individuals such as children, the elderly, and in immunocompromised patients. Some infections associated to S. pneumoniae carry a 40% death mortality even after treatment[1]. In clinical environments like the Respiratory Intensive Care Unit (RICU), its detection often signals severe pulmonary or systemic infections. Its ability of being adamant helps avoid immune responses, and acquire resistance. This emphasizes its status as a persistent threat to public health [2,3].Pulmonary diseases often form the core of pneumococcal pathology, with the community-acquired pneumonia (CAP) and ventilator-acquired pneumonia (VAP) and empyema being frequently observed[2]. However, the microorganism's pathogenic reach goes beyond the lungs. It can also invade the bloodstream and affect the joints, heart, and central nervous system. This increases its ability in causing abnormalities like bacteremia, septic arthritis, carditis, and meningitis[3]. While rare, pneumococcal arthritis and pericarditis are serious complications with high morbidity. Crucially, S. pneumoniae is known for its constancy which makes it not leave the host system easily, often leading to recurrent infections, especially in immune compromised individuals. Despite the worldwide use of pneumococcal conjugate vaccines (PCVs) and other polysaccharide vaccines, S. pneumoniae remains highly relevant. This is because of its part in genetic plasticity that facilitates serotype switching, antibiotic resistance, and immune evasion[4]. The emergence of multidrug-resistant (MDR) strains has additionally complicated the treatment and increased the urgency of quick and accurate diagnostics in intensive care settings[5]. Traditionally, S. pneumoniae is detected using classical microbiological methods of identification such as Gram staining, culturing on blood agar, optochin sensitivity testing and bile solubility testing by Oxgall disks. These methods while being cost-effective and easily available, result in low sensitivity after antibiotic use. Moreover, they cannot determine resistance profiles or serotype information which is critical data for modern infection control. To improve diagnostic accuracy, molecular and immunological methods have been selected. PCR and qPCR assays targeting genes like lytA(autolysin A)or cps (capsular polysaccharides) offer enhanced precision. Antigen detection methods like ELISA and latex agglutination help in the diagnosis of meningitis or pneumococcal infections, especially when culture results are negative. Matrix-assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry also provides quick species identification and high accuracy. However, these methods have many limitations S. pneumoniae being complicated by mutations and gene sharing with similar strains and species. Changes in lytA or cps might often cause false negatives, while horizontal transfer to other streptococci maylead to false positives during PCR analysis, variable lambda phage regions result in increased confusion, antigen tests vary in sensitivity across different sample types, and none of these methods offer a thorough insight into antibiotic resistance mechanisms and virulence factors.Whole Genome Sequencing (WGS) and analysis allows for precise species identification and provides a comprehensive analysis of the genomic content and its components, effectively minimizing the errors commonly associated with conventional detection methods. Hence it is aimed to do whole genome sequencing and analysis of multi drug resistant S.Pneumoniae and find its complete genetic findings.

2. Materials and Methods:

2.1. Sampling, Isolation and Characterization of *S.pneumoniae*:

Samples: Sputum and Lung wash samples were obtained from RICU patients of NIMS hospitals. A written consent was taken from patient/relative. Samples were spread on Blood Agar plates. Five *S. pneumoniae* strains were isolated from the collected samples. Among these, two of the strains were isolated from sputum and three were from the lung wash samples. These bacterial isolates were further sub-cultured onto Blood agar plates at 37°C. Morphological characteristics were recorded on Blood agar and microscopic morphology after gram staining and Biochemical identification like Catalase test, oxidase test, Haemolysis on blood agar. Antibiotic susceptibility testing was done against various antibiotics by Kirby Bauer method. Antibiotic discs were procured from Himedia. A solution of Augmentin antibiotic was prepared by mixing 100μg of Amoxicillin & Potassium Clavulanate. The strain which is found to be extremely resistant is processed for whole genome sequencing

2.2 DNA isolation, library preparation and Next Generation Sequencing:

Isolated bacteria which is multidrug resistance was inoculated in Luria-Bertanibroth containing 1% plasmaand was incubated in the orbital shaking incubator for 18hrs at 37°C and 100 RPM. This was followed by centrifugation of the bacterial culture at 5000 RPM for ten minutes for pelleting of cells. DNA was extracted using QIAampBiOsticBacteremia DNA kit (QIAGEN Germany) by manufacturer's instructions. The extracted DNA was then placed at -20°C till the time it was required for library preparation. Library preparation was performed by using the Nextera XT DNA Library Preparation kit (Illumina, USA). DNA was amplified and disrupted by using transposons present into the Nextera XT Kit .Individual adapters were used for every sample to label appropriately. PCR reaction was performed with the program of 12 cycles of amplification of DNA fragments that were tagged with primers and indices for the generation of dual-indexed sequencing of pooled libraries. After sample normalization, pooling was made and next 300-base paired-end reads sequencing was carried on Illumina (Novaseq 6000), 150PE instrument. Starting from preparation, followed by the sequencing, all the procedures has been done according to the manufactures' guidelines.

2.3Genomeassembly, Annotation,Completeness, ORF (Open Reading Frames) and GC content and GC skew:

2.3.1 Genome assembly

The genome assembly was done by using Megahit v1. 2. 9. The total number of contigs generated were 213,124 contigs greater than 200 base pairs included in the final assembly. Of these 77 contigs were larger than 1000 bp, 29 exceeded 10,000 bp and 18 were longer than 100,000 bp. No contigs exceeded 1 million bp. The largest contig in the assembly measured 282,442 bp.

Assembly statistics total genome size, number of contigs and N50 value were assessed using QUAST v5.0.2[6].

2.3.2 Genome annotation

PROKKA tool was used to generate genes annotation of the prokaryotic genomes and predict the coding regions within these genomes (Prokka Software 1.14.6, Version (Proksee)1.1.1[7].

2.3.3 Genome Completeness:

To know the amount of contamination,present in the genome and the amount of completed genome, Genome completeness was evaluated by BUSCO v5. 3. 2[8]with bacteria_odb10 as the reference. Circular genome was constructed by Proksee[9].

2.3.4 ORF (Open Reading Frames)

Proksee 1.0.0 bioinformatics tool was used for identification of open reading frames in microbial genomic sequence[9].

2.3.5 GC content and GC skew:

GC Content and GC skew i.e the relative amount of G nucleotides over C nucleotides on the leading and lagging strands was analysed with the help of toolkit Proksee 1.0.2 [9].

2.4. Identification:

2.4.1 MLST Species identification:

Species identification was undertaken through Ribosomal Multi locus Sequence Typing (RMLST) [10].

2.4.2 Phylogenetic analysis:

Phylogenetic analysis was conducted to determine the evolutionary relationship of organism. It was carried out by the FASTA sequences that were aligned using Muscle and the Neighbour Joining phylogenetic tree was bulit with bootstrap values of 1000 using MEGA7.0 [11,12].

2.5Antibiotic Resistance and Pathogenicity:

2.5.1 Antibiotic resistance genes:

CARD RGI 6.0.2 Bioinformatical tool of Proksee 1.2.0 was used to identify the antibiotic resistance genes[13].

2.5.2 Pathogenicity

To understand the pathogenicity of the organism at genetic level, the PathogenFinder 1.1 web-based tool was applied[14].

2.6Virus identification and Viral signal detection, and Mobile elements or transposons:

2.6.1 Virus identification

PHASTER tool was used to detect the prophage points in genomic[15].

2.6.2 Viral signal detection

The viral signals were detected by using the Software VirSorter 2.2.4 of Proksee 1.1.1.

2.6.3 Plasmid finder

The presence of Plasmids in the genome were predicted using PlasmidFinder-2.0 Server [16].

2.6.4 Mobile elements

Mobile elements were analysed by Software mobileOG-db (beatrix-1.6) of Proksee 1.1.3 since they play crucial role in the spread of antibiotic resistance[17].

3. Results:

3.1. Isolation, Morphology and Biochemical Identifications:

3.1.1 Isolation:

The five bacteria from RICU Patients sputum and lung wash samples were isolated on Blood Agar and identified as *S.pneumoniae*. It is observed that the colonies were mucoid or slimy appearance on blood agar.

3.1.2 Morphology and Biochemical characterization:

Gram staining showed purple colour (Gram positive) cocci shape in chain under the microscope. Biochemical identification showed catalase test negative, oxidase test negativeand haemolysis on blood agar by alpha haemolysis showing greenish discolouration due to partial haemolysis

By above biochemical tests isolates were identified as *S.* pneumoniae.

3.1.3 Antibiotic Sensitivity Assay:

The strain was found to be Sensitive to Augmentin, Cefexime and Monobactum and Resistant for all other antibiotics tested. (table 1)

Table 1: The Antibiotic–sensitivity profile of *Streptococcus pneumoniae* as Zone of inhibition (mm) with 100µg concentration

Antibiotic tested	Zone of inhibition(mm)
Ampicillin	6 ± 0.18
Cephalosporin	24 ± 0.16
Macrolide	7 ± 0.21
Tetracycline	5 ± 0.15
Monobactam	25 ± 0.75
Carbapenem	7 ± 0.28
Vancomycin	5 ± 0.78
Nitroimidazole	9 ± 0.32
Macrolide	9 ± 0.27
Rifamycin	8 ± 0.24
Fluroquinolone	7 ± 0.28
Elfamycin	8 ± 0.27
Ceftazidime	5 ± 0.20
Cefexime	27 ± 0.20
Norfloxacin	6 ± 0.18
Levofloxacin	8 ± 0.24
Choramphenicol	7 ± 0.21
Streptomycin	5 ± 0.20
Augmentin amoxicillin + k clavulanate	32 ±0.81
Kanamycin	4 ± 0.16

Pencillin-G	5 ± 0.20
Gentamycin	7 ± 0.21

3.2. Genome Annotation, Completeness, Open reading frames and GC content and GC skew:

3.2.1 Genomeannotation performed using **Prokka**. The annotation statistics are provided below.

1. Number of contigs: 124
2. Number of bases: 2,225,710
3. Number of Coding sequences (CDS): 2245
4. Number of misc RNA: 0
5. Number of rRNA: 14
6. Number of tRNA: 61
7. Number of tmRNA: 1

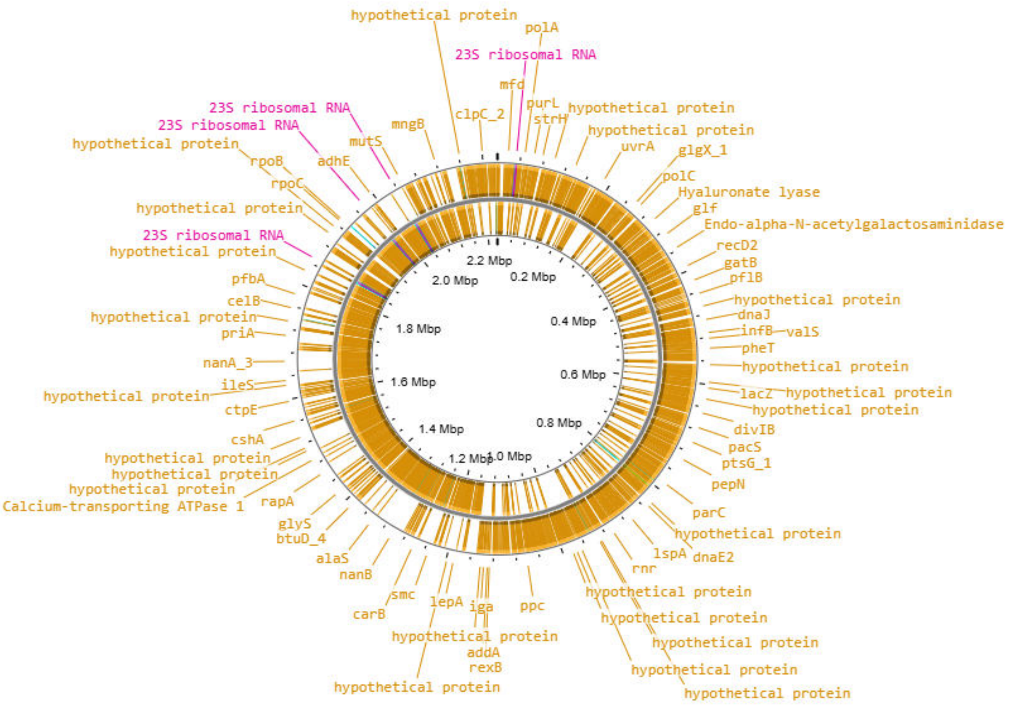

Figure 1: Genome Annotation map of Streptococcus pneumoniae

3.2.2 Genome Completeness:

Completeness Assessment Results:

Number of core genes detected

Complete: 121 (97.58%)

Complete + Partial: 123 (99.19%)

Number of missing core genes: 1 (0.81%)

Average Number of orthologs per core genes: 1.00

% of detected core genes that have more than 1 ortholog: 0.00
Scores in BUSCO format: C:97.6%[S:97.6%,D:0.0%],F:1.6%,M:0.8%,n:124
=99.19%

Length Statistics and Composition:

Number of sequences: 124
Total length (nt): 2225710
Longest sequence (nt): 282,442
Shortest sequence (nt): 4810
Mean sequence length (nt): 22410
Median sequence length (nt): 22250
N50 sequence length (nt): 22570

T:30.12
G:19.97
C:19.57
N:0.00
Other:0.00

Number of gaps (>=5 N's): 0
GC-content (%): 39.54
Number of sequences containing non-ACGTN (nt): 0

3.2.3Open Reading Frames:

The total ORF's identified in the nucleotide sequence of *S. pneumoniae*is 5955 (Figure 2)

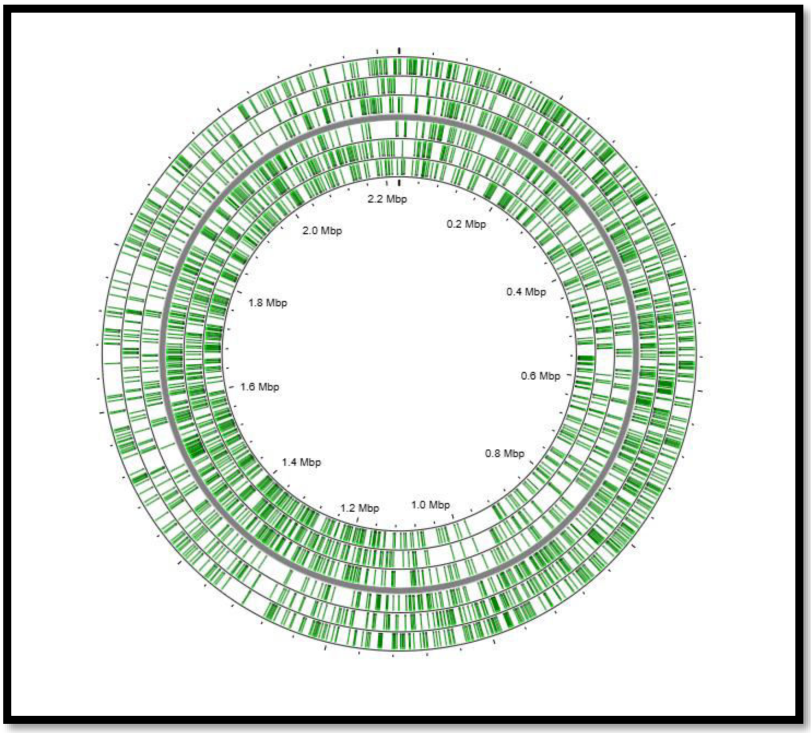

Figure 2: Genomic map containing ORF of S. pneumoniae.

3.2.4 GC Content and GC skew:

The overall GC content of the genome was determined to be 39.5 %(Figure 3) & (Table 2).

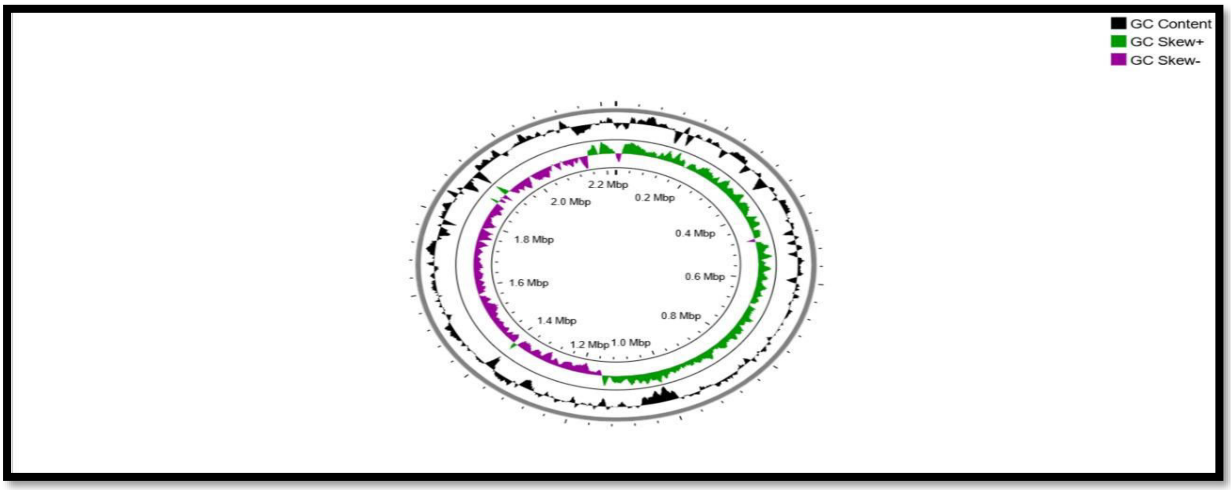

Figure 3: Circular map containing GC content and GC skew constructed by Proksee.

Table 2: Nucleotide base count and percentage

Base	Count
Total	2225710
A	675328
C	670331
G	444574
T	435477
% GC content	39.5

3.3 Identification

3.3.1 MLST Species identification:

The species was identified by Ribosomal Multilocus Sequence Typing (rMLST). rMLST results assessed that sample belongs to *Streptococcus pneumoniae* (Table 3).

Table 3: Species identification done by r MLST species identification.

Rank	Taxon	Support	Taxonomy
Species	*Streptococcus pneumoniae*	100%	*Pseudomonadota> Gammaproteobacteria> Enterobacterales> Enterobacteriaceae> Streptococcus> Streptococcus pneumoniae*

3.3.2 Phylogenetic analysis:

Phylogenetic analysis for the reference genome (Figure 4) showed that isolated strain K1*Streptococcus pneumoniae 21* is highly relavant to complete genome of *S. pneumoniae*.

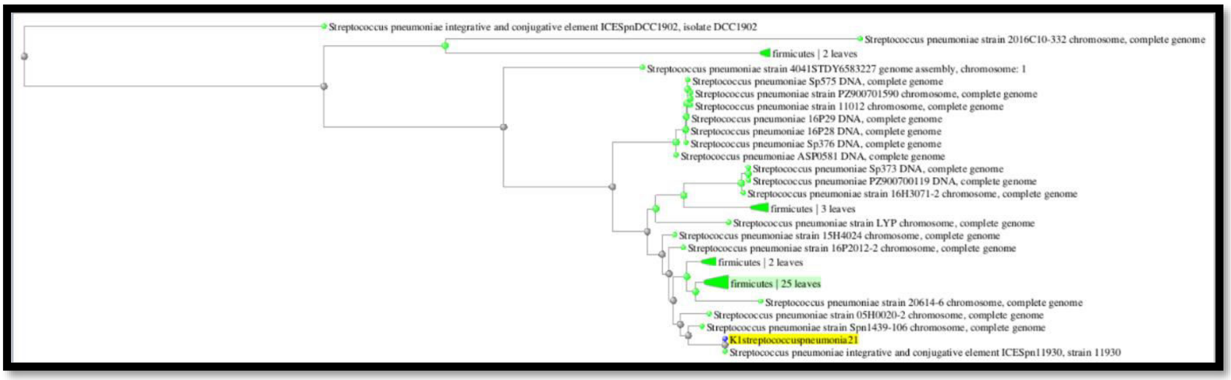
Streptococcus pneumoniae integrative and conjugative element ICESpnDCC1902, isolate DCC1902
Streptococcus pneumoniae strain 2016C10-332 chromosome, complete genome
firmicutes | 2 leaves
Streptococcus pneumoniae strain 4041STDY6583227 genome assembly, chromosome: 1
Streptococcus pneumoniae Sp575 DNA, complete genome
Streptococcus pneumoniae strain PZ900701590 chromosome, complete genome
Streptococcus pneumoniae strain 11012 chromosome, complete genome
Streptococcus pneumoniae 16P29 DNA, complete genome
Streptococcus pneumoniae 16P28 DNA, complete genome
Streptococcus pneumoniae Sp376 DNA, complete genome
Streptococcus pneumoniae ASP0581 DNA, complete genome
Streptococcus pneumoniae Sp373 DNA, complete genome
Streptococcus pneumoniae PZ900700119 DNA, complete genome
Streptococcus pneumoniae strain 16H3071-2 chromosome, complete genome
firmicutes | 3 leaves
Streptococcus pneumoniae strain LYP chromosome, complete genome
Streptococcus pneumoniae strain 15H4024 chromosome, complete genome
Streptococcus pneumoniae strain 16P2012-2 chromosome, complete genome
firmicutes | 2 leaves
firmicutes | 25 leaves
Streptococcus pneumoniae strain 20614-6 chromosome, complete genome
Streptococcus pneumoniae strain 05H0020-2 chromosome, complete genome
Streptococcus pneumoniae strain Spn1439-106 chromosome, complete genome
K1streptococcuspneumonia21
Streptococcus pneumoniae integrative and conjugative element ICESpn11930, strain 11930

Figure 4: Phylogenetic analysis of *Streptococcus pneumoniae*

3.4Antibiotic Resistance and Pathogenicity:

3.4.1 Antibiotic resistance gene analysis

The genes for antibiotic resistance were predicted from genome analysis by CARD database, which identified multiple resistance determinants within the genome. These genes were associated with various antibiotic classes, including Vancomycin, Teicoplanin, Ciprofloxacin, Norfloxacin, Tylosin, Mycinamicin(Figure 5). The results obtained from CARD analysis were matched with that of disc diffusion tests.

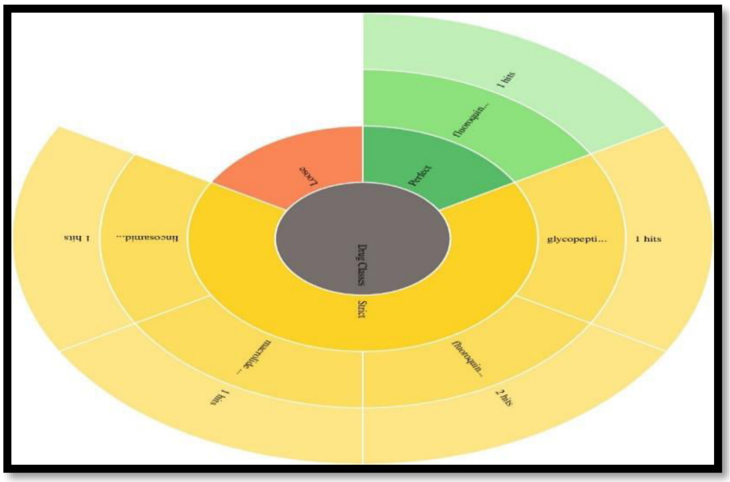

Figure 5: Representation of antibiotic resistance genes by CARD.

3.4.2 Pathogenicity

The probability of being the human pathogen is found to be 88.3% with the minimum threshold value of 100.0

Probability of being a human pathogen	0.883
Input proteome coverage (%)	5.12
Matched pathogenic families	115
Matched not pathogenic families	0

Extra Chromosomal DNA Analysis:

3.5Virus identification, Viral signal detection, Plasmid finder and mobile elements:

3.5.1 Virus identification

Totally 1 intact prophage region were detected using PHASTER tool.

3.5.2 Viral signal detection

One viral signal was predicted using Vir Sorter 2.2.4 tool. Low viral load, indicates no/small amount of viral genetic material present in the genome when compared to the host's nucleic acids

3.5.3 Plasmid finder

Plasmid hits were not found which indicates the absence of plasmids in the genomic sequence.

3.5.4 Mobile elements

A total of **274 mobile orthologous groups (OGs)** were identified through analysis using the mobileOG-db database. Functional annotation revealed the presence of **21 gene associated with phage activity**. Furthermore, **32 mobile genetic elements** involved in gene transfer processes were detected. The classification of mobile elements yielded the following distribution: **156 elements** were associated with **integration and excision mechanisms**, **43 elements** were related to **replication, recombination, and repair**, while **22 elements** were linked to **stability, transfer, and defence functions**. These findings highlight the diversity and complexity of mobile genetic elements present in the *E. coli* under study.

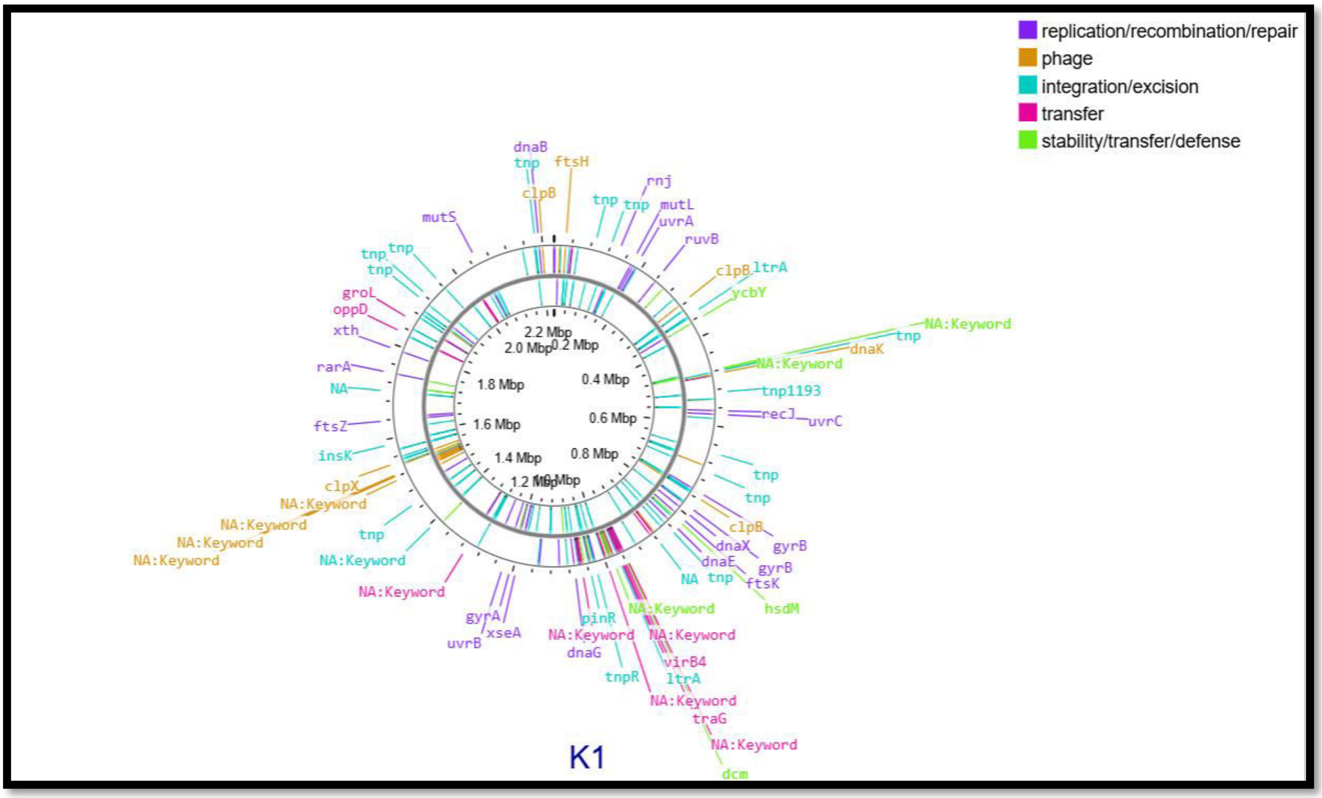

Figure 6: Circular genomic map of *S. Pneumoniae* illustrating Mobile elements.

Data availability: The Whole genome sequence of isolated *S. Pneumoniae* was submitted to NCBI database and the Accession number is SUB15436170

Discussion:

The whole genome analysis of *Streptococcus pneumoniae* isolated from sputum samples of patients provides valuable information of the genetic makeup, antibiotic resistance genes, mobile elements, pathogenic nature, virulence factor etc. Kapatai et al.,(2016)performed whole genome analysis of strain *Streptococcus pneumoniae* acquired from Statens Serum Institute. The study performed by Mironov KO et al., (2020) used *S. pneumoniae* strains isolated during the PEHASusmulticenter studies in 2015–2018. Sequencing was performed using Illumina protocols and equipment. The confirmation of isolated S. pneumoniae through both cultural, biochemical methods and whole genome analysisexplain the accuracy and reliability of the identification process. The distinct morphological characteristics observed on blood Agar, along with results in biochemical tests such asCatalase test, oxidase test, Hemolysis on blood agar, confirms the isolates as S. pneumoniae. Furthermore, molecular techniques such as rMLST further confirm *Streptococcus pneumoniae*.TheSPAdes, SeroBA, PneumoCaT software were used for data processing, as well as BIGSdb software, where in the present study Prokka, BUSCO v5.3.2, Virsorter 1.1.1, Phigaro, CARD RGI 6.0.2, Pathogen finder 1.1, Plasmid finder 2.0, Proksee 1.0.2 were used for the genome analysis. 124 contigs, 2,225,710 bases, 2245 coding sequences, 14 rRNA, 61 tRNA and 1 tmRNA were found in the genome sequence whereas, 2,160,837 base pairs containing 2236 predicted coding regions which are closely related in number to the present organism's genomic base pairs and coding regions was observed by Tettelin et al., (2001). Slager et al., (2018) used the genomic map to identify several new small RNAs (sRNAs), RNA switches (including 16 previously misidentified as sRNAs) and antisense RNAs. In total they annotated 89 new protein encoding genes, 34 sRNAs and 165 pseudogenes. Whereas additionally reported 56 to 71 ncRNA, along with 2 to 5 rRNA and 43 to 49 tRNA genesby Khoeri et al., (2024). As open reading frames (ORFs) play a key role in gene prediction, 5955 ORFs were detected in the present sequence when compared to the study conducted by Dopazo et al., (2001)reported a total of 2046 putative open reading frames which were longer than 100 amino acids and14 ORFs, 13 of which were adjacent to the genes encoding histidine kinase were detected by Throup et al., (2000). In the present study the percentage of guanine and cytosine bases in the sequence was 39.5% whereas40% of G+C content was reported in the study conducted by Hoskins et al., (2001) and G+C content of 39 to 40%as in the same line by Khoeri et al., (2024). In the present study no pathogen islands were detected and the results were same line as carried out by Lau et al., (2001). Antibiotic resistance genes against Penicillin which was found at 30%, Ceftriaxone resistance was found in 19%, Co-trimoxazole 45%, chloramphenicol 13%, Erythromycin and Tetracycline 17%.Skull et al., (1999) whereas the present analysis detected the antibiotic resistance genes which were resistant against Vancomycin, Teicoplanin, Ciprofloxacin, Norfloxacin, Tylosin, Mycinamicin.The results obtained from CARD analysis were matched with that of disc diffusion tests. The present study has reported vanY gene in vanM cluster, pmrA, patB, patA, RlmA(II) genes.Pce (cbpE), pavA, lmb, srtA, slrA, plr(gapA), nanA, eno, piaA, piuA, psaA, cppA, htrA (degP), tig(ropA),andply genes were detected by Yan et al., (2021). Hence whole genome will provide completegenetic insight of *S.* Pneumoniae.

Conclusion:

Whole genome sequencing (WGS) provided a complete picture of an organism's genetic makeup, offering higher accuracy and deeper insights than early techniques that relied only on partial gene sequencing or culture-based methods which are time consuming and with low accuracy

References:

1. Houlihan E, Mc Loughlin R, Waldron R. Streptococcus pneumoniae purulent pericarditis secondary to influenza A infection and pneumococcal pneumonia in an immune competent woman. BMJ Case Reports CP. 2021 Mar 1;14(3):e240763.

2. O'brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, Lee E, Mulholland K, Levine OS, Cherian T. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. The Lancet. 2009 Sep 12;374(9693):893-902.
3. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. The Lancet. 2011 Dec 3;378(9807):1962-73.
4. van der Linden M, Winkel N, Küntzel S, Farkas A, Perniciaro SR, Reinert RR, Imöhl M. Epidemiology of Streptococcus pneumonia e Serogroup 6 Isolates from IPD in Children and Adults in Germany. PloS one. 2013 Apr 9;8(4):e60848.
5. Chan WT, Espinosa M. The Streptococcus pneumoniae pezAT toxin–antitoxin system reduces β-lactam resistance and genetic competence. Frontiers in Microbiology. 2016 Aug 25;7:1322.
6. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013 Apr 15;29(8):1072-5.
7. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068-9.
8. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular biology and evolution. 2018 Mar 1;35(3):543-8.
9. Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen CY, Graham M, Van Domselaar G, Stothard P. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic acids research. 2023 Jul 5;51(W1):W484-92.
10. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, Wimalarathna H, Harrison OB, Sheppard SK, Cody AJ, Maiden MC. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology. 2012 Apr;158(4):1005-15.
11. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research. 2004 Mar 1;32(5):1792-7.
12. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology and evolution. 2016 Jul 1; 33(7):1870-4.
13. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome. 2021 Feb 1;9(1):37.
14. Cosentino S, Voldby Larsen M, Møller Aarestrup F, Lund O. Pathogen Finder-distinguishing friend from foe using bacterial whole genome sequence data. PloS one. 2013 Oct 28; 8 (10):e77302.

15. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic acids research. 2016 May 3;44 (W1):W16-21.
16. Carattoli A, Hasman H. Plasmid Finder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). InHorizontal gene transfer: methods and protocols 2019 Oct 5 (pp. 285-294). New York, NY: Springer US.
17. Brown CL, Mullet J, Hindi F, Stoll JE, Gupta S, Choi M, Keenum I, Vikesland P, Pruden A, Zhang L. mobileOG-db: a manually curated database of protein families mediating the life cycle of bacterial mobile genetic elements. Applied and environmental microbiology. 2022 Sep 22;88(18):e00991-22.
18. Kapatai G, Sheppard CL, Al-Shahib A, Litt DJ, Underwood AP, Harrison TG, Fry NK. Whole genome sequencing of Streptococcus pneumoniae: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. PeerJ. 2016 Sep 14;4:e2477.
19. Mironov KO, Korchagin VI, Mikhailova YV, Yanushevich YG, Shelenkov AA, Chagaryan AN, Ivanchik NV, Kozlov RS, Akimkin VG. Characterization of Streptococcus pneumoniae strains causing invasive infections using whole-genome sequencing. Journal of microbiology, epidemiology and immunobiology. 2020 May 6;97(2):113-8.
20. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, Heidelberg J, DeBoy RT, Haft DH, Dodson RJ, Durkin AS. Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. Science. 2001 Jul 20;293(5529):498-506.
21. Slager J, Aprianto R, Veening JW. Deep genome annotation of the opportunistic human pathogen Streptococcus pneumoniae D39. Nucleic acids research. 2018 Nov 2;46(19):9971-89.
22. Khoeri MM, Maladan Y, Salsabila K, Alimsardjono L, Vermasari N, Puspitasari I, Yunita R, Tafroji W, Sarassari R, Sari RF, Balqis SA. Whole genome sequencing data of Streptococcus pneumoniae isolated from Indonesian population. Data in Brief. 2024 Apr 1;53:110251.
23. Dopazo J, Mendoza A, Herrero J, Caldara F, Humbert Y, Friedli L, Guerrier M, Grand-Schenk E, Gandin C, de Francesco M, Polissi A. Annotated draft genomic sequence from a Streptococcus pneumoniae type 19F clinical isolate. Microbial drug resistance. 2001 Jun 1;7(2):99-125.
24. Throup JP, Koretke KK, Bryant AP, Ingraham KA, Chalker AF, Ge Y, Marra A, Wallis NG, Brown JR, Holmes DJ, Rosenberg M. A genomic analysis of two-component signal transduction in Streptococcus pneumoniae. Molecular microbiology. 2000 Feb;35(3):566-76.
25. Hoskins J, Alborn Jr WE, Arnold J, Blaszczak LC, Burgett S, DeHoff BS, Estrem ST, Fritz L, Fu DJ, Fuller W, Geringer C. Genome of the bacterium

Streptococcus pneumoniae strain R6. Journal of bacteriology. 2001 Oct 1;183(19):5709-17.

26. Lau GW, Haataja S, Lonetto M, Kensit SE, Marra A, Bryant AP, McDevitt D, Morrison DA, Holden DW. A functional genomic analysis of type 3 Streptococcus pneumoniae virulence. Molecular microbiology. 2001 May;40 (3):555-71.
27. Skull SA, Shelby-James T, Morris PS, Perez GO, Yonovitz A, Krause V, Roberts LA, Leach AJ. Streptococcus pneumoniae antibiotic resistance in Northern Territory children in day care. Journal of paediatrics and child health. 1999 Oct 5;35(5):466-71.
28. Yan Z, Cui Y, Huang X, Lei S, Zhou W, Tong W, Chen W, Shen M, Wu K, Jiang Y. Molecular characterization based on whole-genome sequencing of Streptococcus pneumoniae in children living in Southwest China during 2017-2019. Frontiers in Cellular and Infection Microbiology. 2021 Nov 2;11:726740