Summary

Objective 1: Sample collection, identification, screening of antibiotic resistance (ABR) isolates, and whole genome sequencing

- A total of 30 clinical isolates of *K. pneumoniae* were obtained from different pathology laboratories, namely Metropolis Pathology Lab and Toprani Advanced Lab System, in Gujarat. Among these isolates, 17 samples (M2, M3, M6, M10, M17B, M25, M27, M33, M34a, M35, M36, M39, M40, M44, M46, DJ, ST1) were collected in 2016-17, while the remaining 13 samples (M47, M48, M49, M50, M51, M52, M53, M54, M55, M56, M57, M58, M59) were collected in 2020-21. The isolates were cultured on MacConkey agar and incubated overnight at 37 °C, and glycerol stocks were prepared for future use and storage. Initial identification was already carried out by Vitek 2, but we performed further molecular identification using *16S rRNA* sequencing to prevent the consideration of unwanted samples for whole genome sequencing. All isolates were identified as *Klebsiella pneumoniae* by the NCBI blast as shown in **Table 3.3**.
- The drug resistance profile of all the isolates was also determined using the Vitek 2 system, and microbroth dilution technique, and then classified according to their resistance profile as shown in **Table 3.4**. Out of the isolates tested, 53% (M2, M3, M6, M17B, M48, M49, M51, M52, M53, M54, M55, M56, M57, M58, M59, ST1) were classified as extensively drug-resistant (XDR) based on their drug susceptibility profile. The remaining isolates were categorized as multidrug-resistant (MDR) (M10, M27, M34a, M39, M40, M44), pan-drug-resistant (PDR) (DJ, M47, M50), or susceptible (M25, M33, M33, M35, and M46) as shown in **Figure 3.5.**
- The analysis of isolates' drug resistance patterns across time indicated that the proportion of extensively drug-resistant (XDR) isolates, which are resistant to a greater number of antibiotics, was considerably higher in isolates from 2020 compared to isolates from 2016–17. The XDR and less susceptible isolates from 2016–17 showed sensitivity to tigecycline. Nevertheless, the XDR isolates obtained in 2020 demonstrated resistance to tigecycline, which is concerning, as seen in **Figure 3.6**.
- Whole-genome sequencing (WGS) was performed on all isolates within each resistance group. An extensive investigation was carried out on the whole genetic makeup of 30 isolates. This investigation included quantifying the number of contigs in each genome, computing the N50 values, assessing the GC content as a percentage, tallying the amount of coding sequences, and documenting the NCBI accession number for each genome. The genomes had a GC content of around 57%, suggesting their derivation

from *K. pneumoniae*. Most of the isolates had a contig count below 500, which is a significant predictor of superior genome quality, except for M3, M33, and M34a. However, because to its worse genome quality, the M3 sample, which had over 4000 contigs, was eliminated from further examinations. The number of coding sequences in the genomes ranged from 5000 to 8000, with M39 having the highest count of 7794 CDS in its genome. With the exception of M3, all genomes that have been submitted to NCBI (Bioproject number: PRJNA694019) display their accession numbers in **Table 3.6**.

Objective 2: Genotypic and phenotypic study of antibiotic resistance in *K. pneumoniae*

- To study the genomics of *K. pneumoniae* antibiotic resistance, we retrieved an additional 187 drug-resistant genomes from India from the PATRIC database, along with our isolates.
- The highest isolation rate of *K. pneumoniae* isolates in India is found in Tamil Nadu (61.24%), followed by Gujarat, West Bengal, New Delhi, and Assam with 10.52, 10.04, 5.26%, and 4.78% respectively as shown in **Figure 4.2.** Bloodstream infections were the most common, followed by respiratory and urinary diseases. Nosocomial, sepsis, wound, CNS, and implant artificial isolates were less than 1% as depicted in figure 4.3. Isolates were sorted by sample origin, human (92.28%) and environmental (6.69%). The study found 50 sequence types in India *K. pneumoniae* genomes, with ST231 dominating 33.49% of samples. The most prevalent sequence type was ST231 in Tamil Nadu, Gujarat, West Bengal, and Uttar Pradesh. A total of 36 distinct beta-lactamase genes were identified, including carbapenemase genes, extended spectrum beta-lactamases, broad-spectrum beta-lactamases, and 12 other beta-lactamases as depicted in **Figure 4.7**.
- The investigation of the genomic locations and distribution of carbapenemase genes across STs was also done. The most circulating carbapenemase genes in Indian isolates were *bla*OXA-48-like (including *bla*OXA-48, 232, and 181). Of which, *bla*OXA-232 was the most common and detected in 46.41% of genomes, primarily in ST231, with ST2096 and ST14 following closely behind. We also found four types of *bla*NDM: 1, 4, 5, and 7. *bla*NDM-5 was most prevalent (18.18%), followed by *bla*NDM-1 (5.74%). We detected ST147, a prevalent ST in India, in 28.94% of *bla*NDM-5 samples. We mostly found the *bla*NDM-1 gene in the ST14 and ST11 strains. The *bla*NDM-1 gene

was noticed in the ST231, ST147, ST273, ST624, and ST2816 strains. Only four ST147 and one ST101 genomes have the *bla*KPC-2 gene. In India, *bla*NDM and *bla*KPC were less common than *bla*OXA-232. In 11.48% of samples, ST43, ST147, and ST16 had the *bla*OXA-181 gene. Only two ST101 genomes have the *bla*OXA-48 gene. A few genomes revealed five different combinations of the dual carabapenemase. We found *bla*NDM-5 and *bla*OXA-232 in ST147, ST437, ST2096, ST395, and ST14. ST14. ST16 and ST147 have *bla*NDM-5 + *bla*OXA-181. ST101 has *bla*NDM-5 and *bla*OXA-48. Mostly ST14 harbored both carabepenemase genes *bla*NDM-1 and *bla*OXA-232. Lastly, ST14, ST11, and ST42 possess *bla*NDM-1 and *bla*OXA-181.

- In the case of the genomic locations of carbapenemase genes, the majority of the *bla*OXA-232 genes were detected on plasmids. The C18 and SBS3 genomes contained the *bla*OXA-232 gene. Two genomes (B35725 and SBS12) carried *bla*OXA-181 on the chromosome, whereas the others had it on plasmids. All genomes, except B35725 and SBS12, had *bla*NDM-5 genes on plasmids. Furthermore, very few genomes contain both *bla*OXA-48 and *bla*KPC on plasmids.
- The most common ESBL was *bla*CTX-M-15, with frequency of 80.38%. *bla*CTX-M-15 was found in most ST231, ST147, ST2096, ST14, ST16, ST43, ST395, ST11, ST15, ST23, ST45, ST48, ST307, and ST437 genomes. And *bla*CTX-M-15 gene was detected on both plasmids and chromosomes. Along with *bla*CTX-M-15, *bla*OXA-1 was the most frequent ESBL in genomes. No *bla*OXA-1 gene was discovered in the most prevalent ST231 strain. Most often seen is ST2096, followed by ST14, ST147, and ST395. *bla*OXA-1 was also found in a few genomes of ST16, ST11, ST15, ST35, ST48, and ST307. The *bla*OXA-1 gene was discovered on the chromosome in one ST2096, ST48, ST11, and three ST14 genomes. For the remaining genomes, the gene was on plasmids among STs.
- This study also looked at the distribution of broad-spectrum *β*-lactamases and found that *bla*TEM-1 was the most common BSBL, being found in 74.16% of cases. ST231 had the highest prevalence at 42.58%, followed by ST147, ST2096, ST14, ST16, ST43, and ST395. Across STs, 41.14% of genomes had both *bla*OXA-232 and *bla*TEM-1 genes, while 65.55% had both *bla*CTX-M-15 and *bla*TEM-1 genes. Most genomes had *bla*TEM-1 genes on their plasmids. We identified two other BSBLs, *bla*SHV-1 and *bla*SHV-11, as the most prevalent among the genomes studied. ST231 (72.15%) had the highest prevalence of *bla*SHV-1, followed by ST16 (5.06%), ST101 (3.79%), ST515 (3.79%), and ST48 (2.53%). We did not detect *bla*SHV-1 in the genomes of

other commonly found sequence types (STs). And we identified all the *bla*SHV-1 genes on chromosomes, with the exception of one genome each from ST16 and ST101. The ST147 strain predominantly carried the *bla*SHV-11 gene (39.62%), with ST43 (13.20%) and ST395 (7.54%) following closely behind. Additional common sequence types (STs) included ST11, ST23, and ST437. The most widespread ST231, along with ST2096, ST14, and ST16, did not possess the *bla*SHV-11 gene, except for one instance each in the genomes of ST14 and ST16.

- In this study, a few less important beta-lactamases were also found, other than those typically observed. Among them, *bla*ampH was found to be the most widespread and was detected in around 85% of genomes. Interestingly, all *bla*ampH genes were located on chromosomes. The second most prevalent *bla*SHV-28 gene was detected in ST2096, ST14, ST15, and ST307. Also, multiple variants of these *bla*SHV, *bla*CTX-M, *bla*TEM, and *bla*CMY were also detected, along with *bla*OXA-9, *bla*DHA-1, and *bla*LAP-2 in a small number of genomes.
- Indian *K. pneumoniae* genomes had a wide range of plasmid combinations, some of which were linked to specific sequence types (STs) as shown in **Figure 4.10**. Most ST231 genomes contained Col440I and ColKP3, while ST2096 and ST23 primarily linked ColRNAI. The Col(BS512) plasmid was found in high abundance in ST147 strains, while ColpVC had a strong association with ST43. The genomes of ST16, ST23, and ST101 contained Col440II. IncF plasmids were linked with multiple STs, with ST231 genomes exhibiting a high abundance of Inc-type plasmids such as IncFIA, IncFIB(pQil), IncFII(K), and IncF(pAMA1167-NDM-5). ST14 genomes only included IncFII(K) and IncFIB(K). IncR had a substantial correlation with ST147, while ST147 genomes showed the presence of IncFIB(pKPHS1), IncFII(pKPX1), and ColpVC plasmids.
- To examine the relationship between genotype and phenotype in carbapenem resistance, the genomes of chosen isolates (M40, M52, M39, J20, M53, M49, and M17B) were studied as mentioned in **Table 4.2**. Subsequently, the MIC of four different carbapenem drugs was determined. Furthermore, the drugs were then examined for their binding affinity with the most commonly found carbapenemase genes. The study found that different strains of bacteria yielding single carbapenemase had different MIC values for ertapenem, meropenem, and dual carbapenemase producers. M52 had a MIC of 128 µg/mL for ertapenem, M39 had a MIC of 16 µg/mL, J20 had a MIC of >512 µg/mL, and M53 had a MIC of >512 µg/mL for meropenem.

Dual carbapenemase producers had MIC values of 128 µg/mL for ertapenem and >512 µg/mL for meropenem as shown in **Figure 4.11**. Meropenem's docking scores with NDM-1, NDM-5, OXA-181, and OXA-232 were -6.5, -6.2, -6.9, and -7.5 kcal/mol, respectively, indicating hydrogen bonding, van der waals interaction, and other interactions. OXA-232's lower binding energy (-7.5) indicates a reduced binding affinity for meropenem drug. Ertapenem's binding affinity with NDM-1, NDM-5, OXA-181, and OXA-232 was -7.5, -8.0, -7.6, and -9.1 kcal/mol, with ertapenem having a lower affinity than meropenem for all four proteins as shown in **Figure 4.12 & 4.13.**

- The comprehensive analysis of antibiotic resistance revealed that the XDR and PDR isolates were primarily associated with two specific sequence types, namely ST231 and ST147. Further, isolates from these groups also included several genes, including *bla*OXA-48-like, *bla*NDM-type, *bla*CTX-M-1-like, and others that confer resistance to almost all classes of antibiotics, except for colistin.
- In this study, we also identified additional resistance genes like *fos* in all isolates, we exclusively found the genes *aadA2* for aminoglycoside resistance, *mph* & *erm(B)* for macrolide resistance, *arr-2* for rifampin, and *catA1* for phenicol in XDR and PDR isolates as shown in **Figure 4.14**. All colistin-resistant isolates (M47 and M50, excluding M48) showed alterations in pmrB as shown in **Figure 4.15**, while tigecycline-resistant isolates showed numerous mutations in the ramR amino acid sequences as shown in **Figure 4.17**.
- Nevertheless, in this study, we also attempted an alternative method to address the obstacles associated with AMR and biofilm (one of the major challenges responsible for both increased drug resistance and virulence) by using the bacteria's inherent genetic elements called prophages as described in **Figure 4.18**.
- In the genome analysis, we found *n* = 25 "Klebsi " phages that belonged to eight unique "Klebsi" type phages. Of which, PHAGE_Klebsi_phiKO2_NC_005857 and PHAGE_Klebsi_ST147_VIM1phi7.1_NC_049451 were the most prevalent. We found 80% of the 25 "Klebsi" phages intact in ST147, ST5438, ST5217, ST2096, ST1715, ST1087, ST280, ST42, ST16, ST15, and ST11. The ST231 strain lacks "Klebsi" phages, whereas all four ST147 strains, as well as the ST42 and ST16 strains, have two fully functional phages. The study also identified 53 different types of “other than Klebsi” (OTK) phages, with the most common being PHAGE_Salmon_118970_sal3_NC_031940. Other frequently appearing phages included PHAGE_Escher_500465_1_NC_049342,

PHAGE_Escher_RCS47_NC_042128, PHAGE_Entero_c_1_NC_019706, and PHAGE_Entero_mEp237_NC_01970. The majority of intact and questionable prophages from both types were found on chromosomes. However, most of the prophages discovered on plasmids were incomplete and belonged to the "OTK" prophages. A limited number of intact and questionable prophages from both categories were identified on plasmids as shown in **Table 8.1** (See in Appendices).

- The biofilm was categorized using a crystal violet assay and categorized into weak, moderate, and strong groups. We found five weak, ten moderate, seven strong, and five non-biofilm-forming isolates as shown in **Figure 4.19** and **Figure 4.20**. Further, we optimized the concentration of MMC in order to assess the viability of a few selected isolates (M6, M10, M36, DJ, M52, M53, M55, and M57). Following the growth pattern, which showed nearly similar O.D. when exposed to 1µg/ml and 3µg/ml of MMC. However, significant cell death was observed in all isolates, except for M10, when supplemented with 5 µg/ml of MMC as depicted in **Figure 4.21**. Therefore, we chose concentrations of 1 ug/ml and 3 ug/ml for prophage induction as shown in **Figure 4.22 & 4.23.**

Objective 3: Genotypic and phenotypic study of virulence factors in *K. pneumoniae.*

- The ultimate aim of the final objective was to determine the predominant K-types and O-types, as well as the associations with sequence types and virulence, using phenotypic and genomic investigations of virulence. In order to investigate the pathogenicity, we obtained *K. pneumoniae* genomes from the PATRIC database, specifically focusing on Indian genomes. However, the number of genomes (n = 351) used in this research was much more than the number used in the previous study on antimicrobial resistance as shown in **Figure 5.5**.

- The genomes of *K. pneumoniae* contained 93 different sequence types, with ST231 being the most frequent type (20.91%). There were 43 different K-types and 13 O-types, with K51 being the highest (31.9%) and O1 being the most circulating (60.96%). We detected a total of 44 virulence genes, with the most circulating genes in cluster related to biofilm/type-3 fimbriae (*mrkABCDFHIJ*) in the range of (87.46%–93.33%). We also found other genes *fyuA*, *irp1*, *irp2*, and *ybtAEPQSTUX* related to yersiniabactin biosynthesis (55.3%), *kfuABC* related to ferrus ion uptake (44.16%–60.68%), *iutA* and *iucABCD* related to aerobactin transport and synthesis respectively (27.07%–74.6%).

- Further, in the study of the correlation of major prevalent STs with K-type, O-type, and virulence genes among *K. pneumoniae* genomes, a total of 11 frequent ST types (ST231, ST14, ST147, ST2096, ST11, ST395, ST43, ST16, ST23, and ST196), 9 K-types (K51, K64, K2, K10, K24, K30, K81, K1, and K46), and 5 O-types (O1, O2afg, O3/O3a, O2a, and OL101) were assembled to demonstrate their relationship to one another, as seen in **Figure 5.6**.
- The genomes of ST231 (*n* = 105) had a high prevalence of K51-type (98.09%) and O1-type (96.19%) isolates, in that K51 + O1 was the most predominant. Most isolates of ST231 had a *mrkABCDFHIJ* gene cluster, with *mrkH* being the most frequent gene. The ferric uptake operon *kfuABC*, of which *kfuA* and *kfuB* were dominant, was present in both ST231(K51 + O1) and ST231(K51 + O2afg) isolates. The presence of the yersiniabactin biosynthetic genes *fyuA, irp1* and *irp2*, and *ybtAEPQSTUX* ranged from 79.04% to 94.28%. Only *ybtA* and *ybtT* were found in duplicate, with the rest remaining in single copies. Gene *ybtA* was the most frequent gene, while *irp1* and *irp2* were the least abundant. No salmochelin gene cluster *iroBCDN* was detected in any ST231 genome. The *iutA* gene responsible for aerobactin transport was detected in all genomes, but only around half contained the *iucABCD* genes necessary for aerobactin synthesis. Aerobactin synthesis genes *iucABCD* were primarily found in genomes with a dual copy of the *iutA* gene.
- Further, in ST14 (*n* = 37), three different K-type (K2, K51, and K64) and O-type (O1, O2a, and O3/O3a) combinations were detected, with K2 + O1 having the greatest combination. All ST14 genomes had all *mrkABCDFHIJ* gene clusters, except for one that lacked the *mrkA* gene. *mrkC* duplication was found in all ST14 genomes, except for two. All ST14 genomes had the ferric uptake gene *kfuA*, with frequencies of 83.33% and 72.22%, respectively. Ybt synthesis genes *fyuA, irp1* and *irp2*, and *ybtAEPQSTUX* were observed in all ST14 (K2 + O3/O3a) and ST14 (K64 + O1) genomes, except irp2 (75%), ybtQ (50%), and *ybtU* (75%), in ST14 (K64 + O1). Almost all *ybtA* (96.96%) and *ybtT* (90.90%) were found in duplicate.
- The ST2096 genomes, except for a specific combination (K64 + O2a), belonged to the K64-type and O1-type. The gene *rmpA* was absent from all genomes, but *rmpA2* was found in 47.61% of them. The distribution of mrk genes showed diversity, with three genes (*mrkA, mrkC,* and *mrkJ*) found in every genome. The *kfuABC* genes were predominantly present in the ST2096 genomes. Ybt Synthesis Genes *fyuA* and *ybtASTX* were circulating in all genomes, while *irp1, irp2*, and *ybtEPQU* were found in the range

of 66.67% and 95.23%. The *iutA* gene was present in all ST2096 isolates except for one, and only two copies possessed the *iucABCD* genes.

- The ST147 genome had two primary K-type and O-type pairings: K64 + O2a and K10 + O3/O3a. All ST147 genomes lacked the *rmpA* and *rmpA2* genes. Type 3 fimbriae (*mrk*) genes were abundant in all genomes, with *mrkBCFI* found in every genome. We detected most *mrkC* in duplicates, while *mrkH* was less common in K10 + O3/O3a. No ST147 genomes contained *kfuABC* genes. In the ST147 (K10 + O3/O3a) combination, only 5 genomes had ybt genes, compared to 84.61% to 100% in the ST147 (K64 + O2a). All ST147 genomes carried the *iutA* gene, but none contained the *iucABCD* genes.
- Only a small number of less frequent sequence types (ST11, ST395, ST43, ST16, ST23, ST196, and ST101) were detected, with each having 10 or less occurrences. The primary combinations of K-types and O-types seen among these STs were as follows: K24 + O2a, K64 + O1, K30 + O1, K81 + OL101, K1 + O1, K46 + O3/O3a, and K17 + O1, respectively.

  With the exception of ST23 genomes, all rest did not possess both the *rmpA* and *rmpA2* genes. However, just the *rmpA* gene was found in a few genomes of ST11 and ST43. Approximately half of the ST43 and ST196 strains did not possess the complete gene cluster responsible for type 3 fimbriae. Despite the lower quantity of ST23 genomes, they had the highest number of virulence genes, which might be a cause for concern in the near future. Fortunately, the genomes of ST16 and ST196 have a smaller number of pathogenicity genes.
- Following the genome analysis, we determined the phenotypic assays for lab isolates. The string test revealed only isolate M58 as positive; however, we considered an additional isolate, M20, a known string positive isolate from the lab, for replication in this study. The chosen isolates were stained using Maneval's staining method, as shown in **Table 5.1**, and 100 cell's capsule sizes were measured using Image J software. M48 and M49 from the K64 subtype were similar in size, but DJ had a smaller capsule. M50 and M51 have bigger capsulars than M57. String positive strain M58 had the second-smallest capsule as shown in **Figure 5.8 & 5.9**.
- Further, we also examined the mucoviscosity and susceptibility to phagocytosis, followed by a serum killing assay. M17B had the lowest mucoviscosity, while M58 had the highest. DJ and M49 that belonged to the K64 group had a greater degree of mucoviscosity than M48, whereas in the K51 group, isolates M51 and M57 exhibited

higher mucoviscosity than M50 as shown in **Figure 5.10.** In the phagocytosis assay, M51 and M49 were most susceptible to phagocytosis, while DJ had the lowest susceptibility, as detected via measuring the fluorescence intensity of neutrophils as depicted in **Figure 5.11**. Further, in the serum killing experiment, M17B, M49, M57, and M58 were all serum-sensitive, with M48 and DJ being intermediately sensitive. Serum killing resistance was identified in M50 and M51, both the isolates were belonged the O1 + K51-type and prevalent in Indian *K. pneumoniae* genomes as shown in **Figure 5.12**.

- In qualitative estimation of siderophores, all laboratory isolates were tested positive, however differences in % siderophore production was observed, Isolate M57 had the highest siderophore production, with over 20%, followed by M53, M2, M54, and M59. Other isolates, including M6, M10, M25, M27, M34a, M35, M39, M46, M50, M56, and ST1, had low siderophore production as shown in **Figure 5.15**. Further, effect of iron, DIP (iron chelator) and antibiotics on siderophore production were studied in selected isolates. All isolates showed siderophore formation, with levels reaching up to 10%. At all three doses, all isolates showed robust growth compared to the control in the presence of iron, however siderophore synthesis was not detected. We observed a consistent pattern in isolates M10 and M48, where an increase in DIP concentration led to a rise in siderophore synthesis. Isolate M51 and M52 had a similar pattern but had a total production of siderophores that was either equivalent to or less than the control. Isolates M33 and M50 reached their peak siderophore generation when they increased the DIP concentration to 400 uM as shown in **Figure 5.16**.
- The study examined the growth of isolates in the presence of antibiotics (ampicillin, ciprofloxacin, and colistin) at various concentrations. In which highly drug-resistant isolates M48, M51, and M52 showed excellent growth even with ampicillin, while low drug resistant isolates M10, M33, and M51 (which is exception) showed limited growth at all doses. Interestingly, siderophore production was not observed in isolates M10, M48, and M52. In isolate M33, the synthesis of siderophore decreased as the concentration of ampicillin increased. In isolate M50, siderophore production was detected in a significant amount at a concentration of 40 ug/ml, more than the control. In M51, the proportion of siderophore rose as the concentration of ampicillin increased, while the generation of siderophore was lower than the control. The rising concentrations of the ciprofloxacin drug directly correlated with siderophore production in all isolates. M33 showed up to 50% siderophore production at a concentration of 10

ug/ml of ciprofloxacin. Colistin also decreased growth in isolates, except for M48 and M50. At lower concentrations, M10 and M33 showed increased siderophore production compared to the control but decreased at higher drug doses. M50 showed a contrasting trend, with siderophore synthesis rising in tandem with higher doses of colistin drug as shown in **Figure 5.17**.