Comparative genomics of Salmonella Typhi isolates in South Asian region through variant analysis

Maham Niazi[*], Zilwa Mumtaz[*], Ashiq Ali[**], Muhammad Zubair Yousaf[*]

[*] KAM School of Life Sciences, Forman Christian College University, Ferozpur Road, Lahore 54600, Pakistan.
[**]State Key Laboratory of Virology, Center for Biosafety MegaScience, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China

Abstract- To determine the genomic profile of *Salmonella* Typhi causing typhoid fever, a detailed genomic level analysis was performed on whole genome sequences to identify mutations associated with the emergence of Extensively Drug Resistant (XDR) strains by comparative analysis of isolates from Pakistan, Bangladesh and India. The Pakistani isolates exhibited a significantly higher mutation rate, along with a higher proportion of modifiers and silent mutations as compared to others. The distinct genomic characteristics of Lahore isolates emphasizes the increasing prevalence of XDR Salmonella Typhi in Asian countries and underscore the significance of analyzing and comparing its genome with relevant strains. It also suggests a potential regional difference in selective pressure causing the mutations in the genome.

Index Terms- Enterobacteriaceae, Salmonella Typhi, Typhoid Fever, Whole Genome Sequencing

I. INTRODUCTION

Typhoid fever, an illness caused by the *Salmonella* Typhi (*S*.Typhi) bacterium, is a matter of concern for many countries. This bacterium is frequently found in contaminated water and food sources. *S.* Typhi is characterized by its gram-negative, rod-shaped structure and belongs to the Enterobacteriaceae family. Its protective capsule plays a crucial role in environmental survival. Once inside the host, *S.* Typhi proliferates within host cells, shielded by its capsule, which hinders immune cells from engulfment. Human infections typically occur through the consumption of contaminated water and food, allowing the bacterium to enter the bloodstream via the intestinal tract. Symptoms in patients may include headaches, high fevers often reaching 39 to 40°C, pain, cough, constipation or diarrhea, abdominal pain, loss of appetite, and sometimes rashes. Antibiotic resistance poses a significant challenge in typhoid treatment because when a strain becomes resistant to first, second, and third-line antibiotics, it is labeled as Multi-Drug Resistant (MDR), further resistance to drugs, necessitating the use of potent or toxic medications, can lead to the classification of the strain as Extensively Drug Resistant (XDR) 1. Between 2016 and 2019, a total of 14,297 incidents were documented in Pakistan's Sindh province. Of these cases, 9,822 were identified as patients with XDR typhoid. Furthermore, from 2017 to 2021, there were 13,736 reported cases of XDR typhoid 2.

Due to inadequate access to clean water and sanitation, middle- and low-income nations, particularly Asia and Africa suffer greatly with typhoid 3. The main cause of antibiotic drug resistance is excessive antibiotic usage, often resulting from people resorting to self-medication to treat typhoid 4. Several cases have confirmed disease transmission through travel from other parts of the world to Pakistan. For instance, a boy who had traveled from Spain to Pakistan was diagnosed with XDR typhoid 5. Additionally, a pregnant woman in Denmark who had visited Pakistan was diagnosed with XDR typhoid 6. Similarly, in 2018, a child who had relocated to Canada from Pakistan was also diagnosed with XDR typhoid. The Pakistani origin of this outbreak was confirmed through whole-genome sequencing 7. MDR is characterized by resistance to ampicillin, trimethoprim-sulfamethoxazole, and chloramphenicol while XDR exhibits resistance to chloramphenicol, ampicillin, co-trimoxazole, and fluoroquinolones, as well as third-generation cephalosporins 1. Ceftriaxone became the subsequent treatment option with the development of MDR strains 8.

Pakistan has been combatting XDR typhoid since 2018, although the World Health Organization (WHO) was only altered to this in 2018 1. Presently, certain strains have undergone mutations and developing resistance within haplotype 58(H58) of *S.* Typhi.9 In the period from November 2016 to March 2017, all XDR cases belonged to the H58 haplotype. This resistance in *S.* Typhi is attributed to plasmids. Notably, females are more commonly affected than males. It is noteworthy that Azithromycin may still retain its effectiveness against XDR bacteria. However, in Nigeria, there are reports of the excessive use of azithromycin potentially leading to resistance against XDR typhoid 10,11.

To prevent this disease, it is recommended to undergo immunization and boil water before consumption. In severe cases, a course of antibiotics lasting 1 to 2 weeks, coupled with injections, may be prescribed. Improving access to vaccinations, promoting better sanitation, providing education, and ensuring clean water for all can significantly contribute to the fight against typhoid. The pattern of typhoid in India may contribute to the

emergence of ceftriaxone-resistant strains due to increasing resistance to this particular drug 12. Additionally, the significance of the Bangladesh epidemic for Pakistan is notable and can elevate the risk of disease transmission 13.

Genomic data analysis and mutation rate are crucial components to achieve better understandings of public health dynamics. WGS helps many stakeholders, including public health strategy planners and infectious disease control centers, analyze the genome. Genome sequences provide multiple advantages in endemic surveillance, efficient analysis of the genome, prediction and treatment of future mutations 14. This study analyzes *S.* Typhi XDR strain's whole genome sequencing (WGS) and compares it to the reference genome to reveal rapid mutation rates in the genome. Fig 1 shows the sample collection and antibiotic susceptibility test performance before prescribing antibiotics to reduce chances of emergence of MDR and XDR strains due to common practice of self-medication and misinformed diagnosis. "Created with BioRender.com"

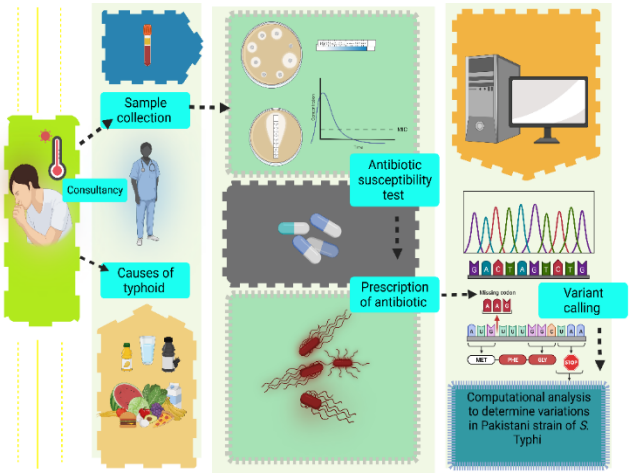

Fig 1: Prescription of right choice of antibiotic after antibiotic susceptibility test and variant calling approach to determine variations in strains

II. Materials and Methods

Data Acquisition and preparation

Genomic data sources

The Sequence Read Archive (SRA) files containing whole genome paired-end data from the S. Typhi isolates from Pakistan, Bangladesh and India were acquired for analysis. The reference genome of *S.* Typhi strain ATCC 13311 with accession no. (NZ_CP009102.1) was retrieved from NCBI database15 for comparative purposes. The input files were sourced from European Nucleotide Archive (ENA)16 with the accession numbers ERR3527964 (5 XDR isolates from India), ERR2663465 (536 antimicrobial resistant isolates from Bangladesh), and SRR10918333 (27 XDR isolates data from Lahore).

Data processing

Variant analysis was performed using the open web based platform Galaxy 17. Whole genome sequences of each variant, including both read 1 and read 2, were concatenated using Concatenate Dataset software (version 0.1.0 of Galaxy). Subsequently, SnpEff build18 (Galaxy Version 4.3+ T.galaxy4) was employed for variant effect prediction and annotation. Fastp19 (Galaxy Version 0.19.5 +galaxy1) was used for quality control, which included checking data quality, filtering, adapter trimming, and quality pruning, all performed in a single operation. The MultiQC tool20 (version 1.9 of Galaxy) was used to aggregate results into a single comprehensive report.

Alignment and Mapping

Alignment and mapping were conducted using BWA-MEM21, which aligned the sequencing reads to the reference genome. SAM or BAM files were filtered using the Samtools view command within the Samtools toolkit, considering criteria such as MAPQ (mapping quality), FLAG bits, read Group, Library, or region. Unmapped BAM files were converted to Fastq format using Samtools fastx (FASTX toolkit, Galaxy)

Post-processing and analysis

Bowtie222 (Galaxy version 2.3.4.3 +galaxy0) was employed to map reads to the reference genome, and Groups were added or replaced in input BAM or SAM files to manage and sort variant strain data. Duplicate molecules in BAM files were identified using Icates (Galaxy version 2.18.2.2).

Variant Calling and Annotation

Somatic single nucleotide variants (SNVs) and indels were called via local assembly of haplotypes using Galaxy (version 4.1.7.0+galaxy02). Vcf Allelic Primitives was used to split gaps and mismatches into multiple lines, as specified by the previous tool. SnpEff annotate variants (Galaxy version 4.3+ T.galaxy1) annotated and predicted variant effects, including changes in amino acids and their effects. SnpSift Extract Fields extracted and selected columns from a VCF dataset, originally generated by the previous SnpEff tool. Finally, the extract tables of each variant were concatenated (tail to head) using Concatenate Dataset software (version 0.1.0 of Galaxy). The complete workflow of variant calling using the Genome Analysis Tool Kit (GATK4) pipeline is illustrated in Figure 2.

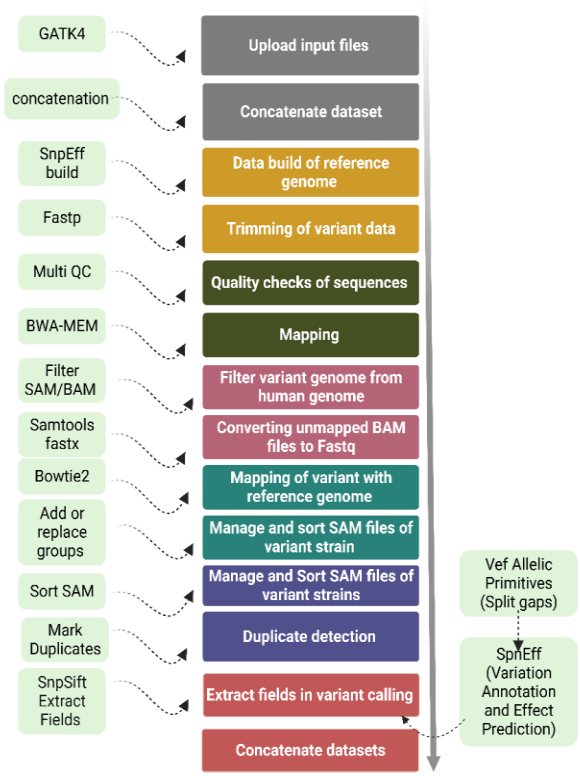

Fig 2. Workflow of variant calling by using GATK4 pipeline. "Created with BioRender.com"

III. Results

Upon variant calling of S. Typhi isolates from Bangladesh, India and Pakistan against the reference strain, substantial variations were observed, including insertions, deletions, and SNPs. While various types of variants were identified, SNPs were the most prevalent in Lahore isolates. Variants were categorized by their type, impact, and genomic regions, revealing diverse mutations, including missense mutations, deletions, stop/gain, upstream and downstream variants. The majority of variants were located in the upstream and downstream regions. The count of each variant is shown in Table 1.

Table 1. Type and count of genetic variants (Single Nucleotide Polymorphism, Multiple Nucleotide Polymorphism, Insertion, Deletion, Mixed, Inversion, Duplication, Breakend and Interval) detected in sequencing data of *S.* Typhi from Lahore, Bangladesh and India.

The analysis revealed the highest count of SNPs from Lahore followed by India and Bangladesh.

Type	Lahore	Bangladesh	India
SNP	60,589	52,646	53,700
MNP	0	0	0
INS	740	576	702
DEL	858	651	752
MIX	0	0	0
INV	0	0	0
DUP	0	0	0
BRE	0	0	0
INTERVAL	0	0	0
TOTAL	62,187	53,873	55,154

The genetic variants types are categorized into various regions in Table 2 including downstream, exon, gene, intergenic, splice_site_region, transcript and upstream regions, and the count of each variant type is provided from Lahore, Bangladesh and India respectively.

Table 2. Distribution of genetic variant types in various regions of S. Typhi isolates from three locations.

Type	Count		
	Lahore	Bangladesh	India
DOWNSTREAM	296,586	258,737	264,344
EXON	53,468	46,243	47,162
GENE	391	312	626
INTERGENIC	619	541	574
SPLICE_SITE_REGION	57	52	53
TRANSCRIPT	57,622	49,929	51,059
UPSTREAM	292,785	253,005	258,427

Highest count of downstream and upstream variants was obtained from Lahore isolates of S. Typhi followed by India and Bangladesh. Number of effects by impact and functional class

When we analyzed the effect of variants in S. Typhi isolates from Lahore and compared them to isolates from Bangladesh and India, it was observed that Lahore isolates exhibited a significant number of modifiers, while India and Bangladesh isolates showed a comparatively lower count.

Table 3. Allocation of genetic variant effects categorized by their impact and functional class for S. Typhi isolates from Lahore, along with comparative data from isolates in Bangladesh and India.

Type	Count		
	Lahore	Bangladesh	India
HIGH	1,208	825	1,211
LOW	38,913	36,141	36,728
MODERATE	12,878	8,929	9,098
MODIFIER	648,529	562,924	575,208

MISSENSE	12,799	8,844	8,990
NONSENSE	183	123	131
SILENT	38,909	36,135	36,723

When other types of genetic variants were compared, the isolates from Lahore had likely high count of mutations in exons, gene, intergenic, transcript, low moderate and silent mutations. S. Typhi isolates from all three localities had non-sense mutation below 0.5%.

IV. Discussion

The evolution of bacterial pathogens within a host can give rise to variants within the same species specific to that host. Recognizing and studying those closely related variants across various host species is essential for both public health and research on how pathogens adapt to their hosts. Nevertheless, this area of research received little attention at the strain level until the introduction of WGS 23,24. Our study utilized the whole genome sequencing data of S. Typhi isolates from three different localities of South Asia and analyzed the genetic diversity among them utilizing the Galaxy platform. Previously, a study was conducted using the same strategy to analyze genetic variants in 787 *S.* Typhi strains collected from diverse bird populations across 18 countries 25. Another study compared *S.* Typhi isolates from the two different era for determination of genotypes, determinants of antimicrobial resistance and plasmid content of isolates using WGS and phylogenetic screening methods 26.

Determination of rare variants are not only used to explain heterogeneity of a certain gene but can also contribute to tell the severity of the disease 27,28. Within the realm of genetic variants, the potentially elevated count of SNPs, modifiers, upstream and down-stream mutations were found in Pakistani isolates of *S.* Typhi in comparison with Bangladesh and India. Earlier, a comprehensive investigation of *S.* Typhi's genome uncovered that the genomes exhibit strong clonal nature, showing limited genetic diversity resulting from SNPs, recombination events and acquisition of genes through horizontal gene transfer 29. Similarly, another study classified the core functional gene clusters with SNPs and revealed that a significant proportion of these genes were associated with metabolic functions and how the *S.* Typhi genome has adopted a strategy to preserve its genome size by regulating presence of both functional and non-functional pseudogenes 30. The variability observed in the genome with high SNP count may support the notion that restoration of functions might be taking place through mutations. The potential drivers molding the genome may be selection pressures and a dynamic evolutionary process in Lahore region. Similarly, the low occurrence of non-sense mutations also suggests that the essential genes are largely preserved in *S.* Typhi isolates from all three regions.

The presence of modifiers and mutations may be driven by multiple evolutionary forces such as antibiotic resistance, host adaptation or immune system evasion apart from environmental variables 31,32. The findings suggest that the genome of *S.* Typhi is subjected to ongoing evolution, with different regions experiencing varying rates and types of genetic changes. The changes can influence pathogen's virulence, antibiotic resistance and overall adaptation to its local environment and host populations 33. Furthermore, these mutations may facilitate rapid bacterial growth and infection in more individuals that necessitates monitoring, and eradication of harmful strains evolved as a result of these mutations. Widespread transmissions of these strains may result from the spread of these strains through human travel 34. Limitations of the study include comparison of *S.* Typhi from only three South Asian regions, the data can be increased to have a better understanding of the actual picture of genomic variations.

V. Conclusion

The Lahore isolate stands out with its significantly higher mutation rate, suggesting potential regional differences in selective pressures. A higher proportion of SNPs could provide evidence for the possibility that functional restoration occurs through mutations. Likewise, the infrequent incidence of non-sense mutations implies that vital genes are predominantly conserved among *S.* Typhi isolates from all three regions. However, high count of modifiers and silent mutations in Lahore isolates suggest that it has accumulated mutations that, while not strongly affecting protein function, may still play a role in adaptation. Further investigations into the genetic determinants of the pathogen-specific traits, as well as the impact of these variations on bacterial physiology and host-pathogen interactions, are essential for advancing our understanding of S. Typhi evolution and for guiding public health efforts in the fight against typhoid fever.

References

[1] Akram J, Khan AS, Khan HA, Gilani SA, Akram SJ, Ahmad FJ, et al. Extensively drug-resistant (XDR) typhoid: Evolution, prevention, and its management. BioMed Research International. 2020;2020:1–7.

[2] Federal Disease Surveillance & Response Unit Field Epidemiology & Disease Surveillance Division Islamabad (NIH). 18th ed. WEEKLY FIELD EPIDEMIOLOGY REPORT 2019.

[3] Mogasale V, Maskery B, Ochiai RL, Lee JS, Mogasale VV, Ramani E, et al. Burden of typhoid fever in low-income and middle-income countries: A systematic, literature-based update with risk-factor adjustment. The Lancet Global Health. 2014;2(10).

[4] Saeed M, Rasool MH, Rasheed F, Saqalein M, Nisar MA, Imran AA, et al. Extended-spectrum beta-lactamases producing extensively drug-resistant

salmonella typhi in Punjab, Pakistan. The Journal of Infection in Developing Countries. 2020;14(02):169–76.
[5] López-Segura N, Corberó-Rivali C, Maldonado-Fernández MC, Calpe-Fraile S, Peyra-Ros J, Martínez-Roig A. Imported extensively drug resistant typhoid fever in a child travelling to Spain from Pakistan. Journal of Travel Medicine. 2019;26(8).
[6] Engsbro AL, Riis Jespersen HS, Goldschmidt MI, Mollerup S, Worning P, Pedersen MS, et al. Ceftriaxone-resistant salmonella enterica serotype typhi in a pregnant traveller returning from Karachi, Pakistan to Denmark, 2019. Eurosurveillance. 2019;24(21).
[7] Wong W, Rawahi HA, Patel S, Yau Y, Eshaghi A, Zittermann S, et al. The first Canadian pediatric case of extensively drug-resistant salmonella typhi originating from an outbreak in Pakistan and its implication for empiric antimicrobial choices. IDCases. 2019;15.
[8] Levine MM, Simon R. The gathering storm: Is untreatable typhoid fever on the way? mBio. 2018;9(2).
[9] Klemm EJ, Shakoor S, Page AJ, Qamar FN, Judge K, Saeed DK, et al. Emergence of an extensively drug-resistant salmonella enterica serovar Typhi clone harboring a promiscuous plasmid encoding resistance to fluoroquinolones and third-generation cephalosporins. mBio. 2018;9(1).
[10] Rasheed MK, Hasan SS, Babar Z-U-D, Ahmed SI. Extensively drug-resistant typhoid fever in Pakistan. The Lancet Infectious Diseases. 2019;19(3):242–3.
[11] Ahmad S, Tsagkaris C, Aborode AT, Ul Haque MT, Khan SI, Khawaja UA, et al. A skeleton in the closet: The implications of covid-19 on XDR strain of typhoid in Pakistan. Public Health in Practice. 2021;2:100084.
[12] Jacob JJ, Pragasam AK, Vasudevan K, Veeraraghavan B, Kang G, John J, et al. Salmonella typhi acquires diverse plasmids from other Enterobacteriaceae to develop cephalosporin resistance. Genomics. 2021;113(4):2171–6. https://doi.org/10.1016/j.ygeno.2021.05.003
[13] Lima NC, Tanmoy AM, Westeel E, de Almeida LG, Rajoharison A, Islam M, et al. Analysis of isolates from Bangladesh highlights multiple ways to carry resistance genes in salmonella typhi. BMC Genomics. 2019;20(1). https://doi.org/10.1186/s12864-019-5916-6
[14] Argimón S, Yeats CA, Goater RJ, Abudahab K, Taylor B, Underwood A, et al. A global resource for genomic predictions of antimicrobial resistance and surveillance of salmonella typhi at pathogenwatch. Nature Communications. 2021;12(1).
[15] http://www.ncbi.nlm.nih.gov/refseq/
[16] Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, Cheng Y, Cleland I, Faruque N, Goodgame N, Gibson R, Hoad G. The European nucleotide archive. Nucleic acids research. 2010 Oct 22;39(suppl_1):D28-31.
[17] http://usegalaxy.org
[18] Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. fly. 2012 Apr 1;6(2):80-92.
[19] https://github.com/OpenGene/fastp
[20] http://dx.doi.org/10.1093/bioinformatics/btw354
[21] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997. 2013 Mar 16.
[22] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012 Apr;9(4):357-9.
[23] Arnold BJ, Huang IT, Hanage WP. Horizontal gene transfer and adaptive evolution in bacteria. Nature Reviews Microbiology. 2022 Apr;20(4):206-18.
[24] Ebert D, Fields PD. Host–parasite co-evolution and its genomic signature. Nature Reviews Genetics. 2020 Dec;21(12):754-68.
[25] Fu Y, M'ikanatha NM, Dudley EG. Population structure and genetic diversity of Salmonella Typhimurium in avian hosts. bioRxiv. 2022:2022-11.
[26] Maes M, Sikorski MJ, Carey ME, Higginson EE, Dyson ZA, Fernandez A, et al. Whole genome sequence analysis of Salmonella Typhi provides evidence of phylogenetic linkage between cases of typhoid fever in Santiago, Chile in the 1980s and 2010–2016. Bourret TJ, editor. PLoS Negl Trop Dis. 2022 Jun 29;16(6):e0010178. 10.1371/journal.pntd.0010178
[27] Chattaway MA, Gentle A, Nair S, Tingley L, Day M, Mohamed I, Jenkins C, Godbole G. Phylogenomics and antimicrobial resistance of Salmonella Typhi and Paratyphi A, B and C in England, 2016–2019. Microbial Genomics. 2021;7(8).
[28] Hantaweepant C, Suktitipat B, Pithukpakorn M, Chinthammitr Y, Limwongse C, Tansiri N, et al. Whole exome sequencing and rare variant association study to identify genetic modifiers, KLF1 mutations, and a novel double mutation in Thai patients with hemoglobin E/beta-thalassemia. Hematology. 2023 Dec 31;28(1):2187155.
[29] Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I, Rance R, Baker S, Maskell DJ, Wain J, Dolecek C. High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. Nature genetics. 2008 Aug;40(8):987-93.
[30] Baddam R, Kumar N, Shaik S, Lankapalli AK, Ahmed N. Genome dynamics and evolution of Salmonella Typhi strains from the typhoid-endemic zones. Scientific reports. 2014 Dec 12;4(1):7457.
[31] Kariuki S, Mbae C, Van Puyvelde S, Onsare R, Kavai S, Wairimu C, Ngetich R, Clemens J, Dougan G. High relatedness of invasive multi-drug resistant non-typhoidal Salmonella genotypes among patients and asymptomatic carriers in endemic informal settlements in Kenya. PLOS Neglected Tropical Diseases. 2020 Aug 3;14(8):e0008440.
[32] Ali A, Amjad N, Javed F, Abbas ZU, Muzammil S, Ahmad MZ, Oranab S, Umar M, Sajid M. Presence and Antibiotic Resistance of MDR Salmonella Isolates Recovered from Zea mays L. Farms Located near the Poultry Farms in Faisalabad-Pakistan. Advancements in Life Sciences. 2023 Jan 21;9(4):516-20.

[33] Willemse SW, Van Es MA. Susceptibility and disease modifier genes in amyotrophic lateral sclerosis: from genetic associations to therapeutic implications. Current Opinion in Neurology [Internet]. 2023 Jun 15 [cited 2023 Jul 12];Publish Ahead of Print. Available from: https://journals.lww.com/10.1097/WCO.0000000000001178

[34] Ingle DJ, Nair S, Hartman H, Ashton PM, Dyson ZA, Day M, Freedman J, Chattaway MA, Holt KE, Dallman TJ. Informal genomic surveillance of regional distribution of Salmonella Typhi genotypes and antimicrobial resistance via returning travellers. PLoS neglected tropical diseases. 2019 Sep 12;13(9):e0007620.

AUTHORS

First Author –Maham Niazi, M.Phil, KAM School of Life Sciences, Forman Christian College University, Ferozpur Road, 54600 Lahore, Pakistan.

Second Author – Zilwa Mumtaz, PhD Scholar, KAM School of Life Sciences, Forman Christian College University, Ferozpur Road, 54600 Lahore, Pakistan.

Third Author – Ashiq Ali, State Key Laboratory of Virology, Center for Biosafety MegaScience, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China

Correspondence Author – Dr. Muhammad Zubair Yousaf, KAM School of Life Sciences, Forman Christian College University, Ferozpur Road, 54600 Lahore, Pakistan.