ORIGINAL ARTICLE

Variant Analysis of XDR Salmonella Typhi Strains Using Global Alignment Tool Kit in South Asian Region

Maham Niazi[1], Zilwa Mumtaz[1], Maqsood Ahmed[2], Saeed Ahmad[3], Ashiq Ali[4], Muhammad Zubair Yousaf[1*]

ABSTRACT

Objective: To concisely compare genomic profiles of XDR *Salmonella* Typhi isolates from Lahore with antimicrobial-resistant *Salmonella* Typhi isolates from other developing nations.

Study Design: Comparative analysis of whole genome sequences.

Place and Duration of Study: The study was conducted from December 2020 to September 2021 at Kauser Abdulla Malik (KAM) School Life Sciences, Forman Christian College University, Lahore, Pakistan.

Methods: The Galaxy pipeline was run to obtain detailed information at the nucleotide level regarding mutations that lead to the emergence of XDR strains. Whole genome sequences were analyzed to compare the genomes of selected three developing nations.

Results: The Pakistani isolates had a significantly higher mutation rate, higher proportion of modifiers, and silent mutations as compared to isolates of Bangladesh and India.

Conclusion: Cases of *Salmonella* Typhi XDR are rapidly rising in Asian countries such as Pakistan, Bangladesh, and India emphasizing the need to analyze and compare its genome with relevant strains. Our study highlights the unique profile of the Lahore (Pakistan) isolate with the highest mutation rate suggesting the potential regional differences in selective pressure. Further spotlights the necessity to elucidate the functional consequences of the identified mutations in *S*. Typhi isolates.

Keywords: *Enterobacteriaceae, Salmonella Typhi, Typhoid Fever, Whole Genome Sequencing.*

How to cite this: *Niaz M, Mumtaz Z, Ahmed M, Ahmad S, Ali A, Yousaf MZ. Variant Analysis of XDR Salmonella Typhi Strains Using Global Alignment Tool Kit in South Asian Region. Life and Science. 2024; 5(2): 244-250. doi: http://doi.org/10.37185/LnS.1.1.432*

Introduction

Typhoid fever, an illness caused by the *Salmonella* Typhi (*S.* Typhi) bacterium, is a matter of concern for many countries. The 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, and high fevers often reaching 39 to 40°C along with 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

[1]*KAM School of Life Sciences*
Forman Christian College University, Lahore, Pakistan
[2]*Department of Biochemistry*
Post Graduate Medical Institute Lahore, Pakistan
[3]*Department of Biological Sciences*
National University of Medical Sciences (NUMS), Rawalpindi, Pakistan
[4]*State Key Laboratory of Virology*
Center for Biosafety Mega Science, Wuhan Institute of Virology
Chinese Academy of Sciences, Wuhan, China

Correspondence:
Dr. Muhammad Zubair Yousaf
Associate Professor, KAM School of Life Sciences
Forman Christian College University, Lahore, Pakistan
E-mail: mzubairyousaf@fccollege.edu.pk

Funding Source: NIL; Conflict of Interest: NIL
Received: June 20, 2023; Revised: Nov 17, 2023
Accepted: Jan 05, 2024

as Multi-Drug Resistant (MDR), further resistance to drugs, necessitating the use of potent medications, can lead to the classification of the strain as Extensively Drug Resistant (XDR).[1] 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.[2] Additionally, a pregnant woman in Denmark who had visited Pakistan was diagnosed with XDR typhoid.[3] Similarly, in 2018, a child who had relocated to Canada from Pakistan was also diagnosed with XDR typhoid. Pakistani origin of this pathogen was confirmed through whole-genome sequencing.[4] Due to inadequate access to clean water and sanitation, middle- and low-income nations, particularly in Asia and Africa suffer greatly from typhoid.[5] The main cause of antibiotic drug resistance is excessive antibiotic usage, often resulting from people resorting to self-medication to treat typhoid.[6] 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.[7] Pakistan has been combatting XDR typhoid since 2018, although the World Health Organization (WHO) was only alerted to this in 2018.[1] Presently, certain strains have undergone mutations and developing resistance within haplotype 58(H58) of *S*. Typhi.[8] 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.[9,10] 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.[11] Additionally, the significance of the Bangladesh epidemic for Pakistan is notable and can elevate the risk of disease transmission.[12] Genomic data analysis and mutation rate are crucial components to achieve a better understanding of pathogen's public health dynamics. WGS helps many stakeholders, including public health strategy planners and infectious disease control centers to analyze the genome. Genome sequences provide multiple advantages in endemic surveillance, efficient analysis of the genome, and prediction and treatment of future mutations.[13] This study analyzes the *S*. Typhi XDR strain's whole genome sequencing (WGS) and compares it to the reference genome to reveal mutation rates in the genome.

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 the NCBI database,[14] for comparative purposes. The input files were sourced from the European Nucleotide Archive (ENA),[15] 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.[16] 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 build,[17] (Galaxy Version 4.3+ T.galaxy4) was employed for variant effect prediction and annotation. Fastp,[18] (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 tool,[19] (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-MEM,[20] 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

Bowtie,[21] (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 were used to split gaps and mismatches into multiple lines, as specified by the previous tool. SnpEff eff 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 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.1. "Created with BioRender.com"

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

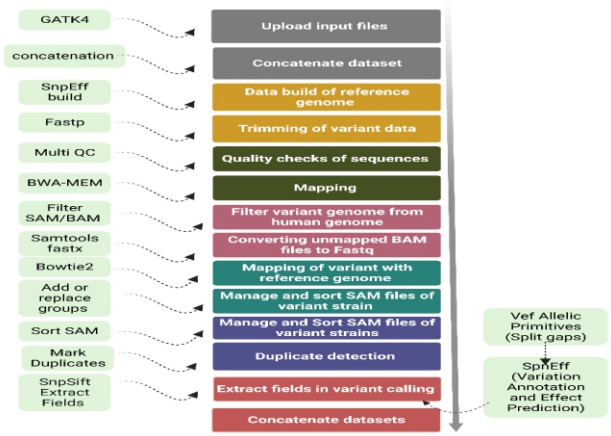

Fig.1: Workflow of variant calling by using the GATK4 pipline

missense mutations, deletions, stop/gain, and 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.

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.

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).

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%.

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

Table-1: Genetic Variant Analysis of *Salmonella* Typhi Sequencing data

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

Table-2: Distribution of Genetic Variant Types in *Salmonella* Typhi Isolates

Type	Count (n)		
	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

Table-3: Allocation of Genetic Variant Effects in *Salmonella* Typhi Isolates

Type	Count (n)		
	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

pathogens adapt to their hosts. Nevertheless, this area of research received little attention at the strain level until the introduction of WGS.[22,23] 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.[24] Another study compared *S.* Typhi isolates from two different eras for the determination of genotypes, determinants of antimicrobial resistance and plasmid content of isolates using WGS and phylogenetic screening methods.[25] Determination of rare variants is not only used to explain the heterogeneity of a certain gene but can also contribute to telling the severity of the disease.[26] Within the realm of genetic variants, the potentially elevated count of SNPs, modifiers, and upstream and downstream 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 a strong clonal nature, showing limited genetic diversity resulting from SNPs, recombination events, and acquisition of genes through horizontal gene transfer.[27] 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 the presence of both functional and non-functional pseudogenes. The variability observed in the genome with a high SNP count may support the notion that the 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 the 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.[28] 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 the pathogen's virulence, antibiotic resistance, and overall adaptation to its local environment and host populations.[29] Furthermore, these mutations may facilitate rapid bacterial growth and infection in more individuals that necessitates monitoring, and eradication of harmful strains evolved in a result of these mutations. Widespread transmissions of these strains may result from the spread of these strains through human travel.[30] Limitations of the study include a 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.

Conclusion

The Lahore isolates stand out with their 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, the high count of modifiers and silent mutations in Lahore isolates suggest that accumulated mutations, 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: 6432580. doi: 10.1155/2020/6432580
2. 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: taz066. doi: 10.1093/jtm/taz066
3. Engsbro AL, 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: 1900289. doi: 10.2807/1560-7917.ES.2019.24.21.1900289

4. Wong W, Al Rawahi H, 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. ID Cases. 2019; 15: e00492. doi: 10.1016/j.idcr.2019.e00492
5. 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: e570-80. doi: 10.1016/S2214-109X(14)70301-8
6. 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: 169-76. doi: 10.3855/jidc.12049
7. Levine MM, Simon R. The gathering storm: Is untreatable typhoid fever on the way? mBio. 2018; 9: e00482-18. doi: org/10.1128/mbio.00482-18
8. 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: 10-128. doi :10.1128/mbio.00105-18
9. Rasheed MK, Hasan SS, Babar ZUD, Ahmed SI. Extensively drug-resistant typhoid fever in Pakistan. The Lancet Infectious Diseases. 2019; 19: 242-3. doi: 10.1016/S1473-3099(19)30051-9
10. 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. doi: 10.1016/j.puhip.2021.100084
11. 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: 2171-6. doi: 10.1016/j.ygeno.2021.05.003
12. 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: 530. doi: 10.1186/s12864-019-5916-6
13. 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: 2879. doi: 10.1038/s41467-021-23091-2
14. Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, Cheng Y, et al. The European nucleotide archive. Nucleic acids research. 2010; 39: D28-31. doi: 10.1093/nar/gkq967
15. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome research. 2005; 15: 1451-5. doi: 10.1101/gr.4086505
16. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. 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; 6: 80-92. doi: 10.4161/fly.19695
17. Chen S, Zhou Y, Chen Y, Gu J. fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; 34: i884-90. doi: 10.1093/bioinformatics/bty560
18. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016; 32: 3047-8. doi: 10.1093/bioinformatics/btw354
19. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint arXiv:1303.3997. 2013: 1-3. doi: 10.48550/arXiv.1303.3997
20. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9: 357-9. doi: 10.1038/nmeth.1923
21. Arnold BJ, Huang IT, Hanage WP. Horizontal gene transfers and adaptive evolution in bacteria. Nature Reviews Microbiology. 2022; 20: 206-18. doi: 10.1038/s41579-021-00650-4
22. Ebert D, Fields PD. Host–parasite co-evolution and its genomic signature. Nature Reviews Genetics. 2020; 21: 754-68. doi: 10.1038/s41576-020-0269-1
23. Fu Y, M'ikanatha NM, Dudley EG. Population structure and genetic diversity of Salmonella Typhimurium in avian hosts. bioRxiv. 2022: 2022-11. doi: 10.1101/2022.11.17.516949
24. 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. PLoS neglected tropical diseases. 2022; 16: e0010178. doi: 10.1371/journal.pntd.0010178

25. Chattaway MA, Gentle A, Nair S, Tingley L, Day M, Mohamed I, et al. Phylogenomics and antimicrobial resistance of Salmonella Typhi and Paratyphi A, B and C in England, 2016-2019. Microbial Genomics. 2021; 7: 000633. doi: 10.1099/mgen.0.000633
26. 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; 28: 2187155. doi: 10.1080/16078454.2023.2187155
27. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I, et al. High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. Nature genetics. 2008; 40: 987-93. doi: 10.1038/ng.195
28. Kariuki S, Mbae C, Van Puyvelde S, Onsare R, Kavai S, Wairimu C, et al. 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; 14: e0008440. doi: 10.1371/journal.pntd.0008440
29. Willemse SW, van Es MA. Susceptibility and disease modifier genes in amyotrophic lateral sclerosis: from genetic associations to therapeutic implications. Current Opinion in Neurology. 2023; 36: 365-70. doi: 10.1097/WCO.0000000000001178
30. Ingle DJ, Nair S, Hartman H, Ashton PM, Dyson ZA, Day M, et al. Informal genomic surveillance of regional distribution of Salmonella Typhi genotypes and antimicrobial resistance via returning travellers. PLoS neglected tropical diseases. 2019; 13: e0007620. doi: 10.1371/journal.pntd.0007620

Authors Contribution

MN: Data collection, data analysis, results and interpretation
ZM: Data analysis, results and interpretation, manuscript writing, and proofreading
MA: Data analysis, results and interpretation, manuscript writing, and proofreading
SA: Data analysis, results and interpretation, manuscript writing, and proofreading
AA: Data analysis, results and interpretation, manuscript writing, and proofreading
MZY: Idea conception, study designing, manuscript writing, and proofreading