Research Article

Comparative Genome Analysis Reveals Insights into New Agr-operon Defects in *Staphylococcus aureus*

Waqas Ahmad1* , Erman Salih İstifli²

¹ Dr. Ikram-ul-Haq Institute of Industrial Biotechnology, Department of Medical Laboratory Technology, Government College University, Lahore, Pakistan ² Faculty of Science and Literature, Department of Biology, Çukurova University, Adana 01330, Turkey

*Corresponding author:

Dr. Waqas Ahmad (e-mail: drwaqasahmadrajpoot@gmail.com)

Date of Receiving: 22/01/2024 Date of Acceptance 16/04/2024 Date of Publishing 10/06/2024 **ABSTRACT** Among the emerging pathogens, *Staphylococcus aureus* has ability to secrete virulence factors solely depends upon the Quorum Sensing (QS) system especially Accessory gene regulating (agr) operon. If agr-operon is disrupted, they are no longer virulent to establish the disease. This study aims to understand the disruptions in the agr-operon. A total of 460 complete genomes of *S. aureus* were downloaded from NCBI and analyzed using ANI and insilico MLST. For downstream analyses, 459 genomes were selected and pangenome was estimated along with core genome phylogeny and recombination identification. It was revealed that these genomes were mainly isolated from humans ($n=308$) and animals ($n=54$) while isolation history of 98 genomes was unknown. There was also significant difference in distribution of genomes association with clonal complexes. Pan-genome studies revealed higher percentage of accessory genes as compared to core and unique genes. On the other hand, agr-defects were present in 30 (6.5%) strains of the pangenome. Except agrD, other agr genes i.e. agrA, agrB and agrC were found defective. AgrB defects have not been previously reported. Frame shift mutations or small indels could be major causes of these defects. This could be a possible outcome of genome reduction, facilitating prolonged infection in chronic cases, antibiotic-resistant strains, elderly individuals, and immuno-compromised cases.

KEYWORDS Staphylococcus, Frameshifts, Genome, Insertion sequence, Genome reduction

Introduction

Staphylococcus aureus, especially Methicillin Resistant *S. aureus* (MRSA), is one of the emerging pathogens associated with nosocomial and community-acquired infections. The ability of this pathogen to secrete virulence factors solely depends upon the Quorum Sensing (QS) system (Vinodhini and Kavitha, 2024). The Accessory gene regulating (agr) operon, LuxS, and RAP-TRAP are the prevailing QS mechanisms in *S. aureus* (Schlievert *et al*, 2023). Briefly, agr system is primarily comprised of agr-operon that in turn composed of 4 genes such as agrB, agrD, agrC and agrA. The histidine kinase agrC senses the auto-inducing protein signals (AIP) produced by agrD and transported by agrB channels (Osbourn and Field, 2009; Omer *et al*, 2024). Phosphorilization of agrC activates the agrA that transcribes the RNAIII and RNAII. RNAIII transcription results in the expression of multiple virulence genes. On the other hand, expression of agrA inhibits the biofilm production by upregulating the proteases and down-regulating the surface proteins production and decreased dispersion of mature biofilm (Inagaki *et al*, 2024). As agr-operon is associated with regulating bacterial pathogenesis and virulence, agroperon is a clinically important target (Inagaki *et al*, 2024). If agr-operon is disrupted, these strains are no longer virulent to establish the disease.

Agr-defective strains are increasingly gaining importance among staphylococcus aureus isolates. Genetically engineered mutants of *Staphylococcus aureus* with defective agr have substantially reduced virulence (Wright *et al*, 2005). Several studies reported the benefits of agr dysfunction to bacteria such as survival advantage under vancomycin selection pressure and association with persistent bacteremia (Fowler *et al*, 2004; Tsuji *et al*, 2007). Moreover, it is also important in the suppression of biofilm production. Such agrdefective strains demonstrate less virulence with high biofilm production and are extremely important in device-related infections.

Genetic variations due to gene disruptions caused by insertion sequences (IS) or integrated prophages affect gene expressions (Osbourn and Field, 2009; Greenrod *et al*, 2023). Additionally, small insertions and deletions can introduce

To cite this article: Ahmad, W., and E. S. İstifli. (2024). *Comparative Genome Analysis Reveals Insights into New Agr-operon Defects in Staphylococcus aureus*. Journal of Epidemiology and Infection Biology 1(1):24-30.

frameshifts and stop codons which will also affect the gene function, microevolution, and host adaptation (Tenaillon *et al*, 2016). Recent studies have mentioned that agr-disrupted strains are often associated with chronic infections which shows the prolonged survival and persistence of bacteria in the host (Bezar *et al*, 2019). Previously, presence of insertion sequence (Is256) and mutational defects in agrC and agrA are reported respectively. But agrB and agrD are never studied for this kind of disruptions. The study of such mutations is important for understanding the mechanisms of protein sequence evolution.

Materials and Methods

Data Collection

All the *Staphylococcus aureus* genomes were retrieved from NCBI database (Awan *et al*, 2021) comprising of 460 complete genomes and chromosomes (Supplementary Table 1). All these genomes were included in whole-genome Average Nucleotide Identity (ANI) computation; performed by FASTANI (website and package). It utilizes Mashmap to calculate the ortholog sequence mappings and alignment identity. Furthermore, to perform the similar computations based on genome-to-genome genetic distance, Genome-to-Genome Distance Calculator (GGDC) was utilized (Meier-Kolthoff *et al*, 2013; Meier-Kolthoff *et al*, 2022).

Metadata isolation

Traits such as isolation host, country of origin, and antibiotic resistance (MRSA/MSSA) related to *Staphylococcus aureus* were downloaded from NCBI and PATRIC. As the data in public databases is not complete, in-silico MLST was performed to find out the MLST groups and clonal complex groups of this dataset. This metadata was used to associate patterns of gene presence and absence with phenotypes exhibited by groups of taxa.

Genome annotation

To get the stable design in the dataset, Prokka was utilized to predict the genes and annotate all the genomes. Prokka utilizes the Prodigal in gene ORF prediction that further annotated by databases from UNIPROT and ISfinder. MinCED package utilized by Prokka predicted the CRISPR sequences in the genomes (https://github.com/ctSkennerton/minced).

Pan-genome inference

The pan-genome of *Staphylococcus aureus* was inferred with Roary. The Prokka annotated genomes in gff format were used as input to Roary. These genomes were grouped by CD-Hit and gene presence-absence binary matrix was produced. Additionally, Roary also produced multi-fasta core gene alignment and first 4000 accessory genes-based tree using PRANK and FastTree 2 (version 2.1.9) respectively. Roary categorized the pan-genome into core, soft core, shell, and cloud. For further simple categorization, Epi Gene package in R was utilized that categorizes the binary matrix into core, accessory and unique genes along with the identification of respective groups/clusters and genomes.

Core genome analysis and Recombination detection

Recombination detection is an important step in prediction of accurate phylogeny. The core genome produced in previous step was used as input in gubbins package that detected the recombination sequences in the core genome alignment. The snp-sites package was utilized that filtered the SNPs. This package identified both the substitution and insertions/deletions. Furthermore, FastTree was used to build a new tree while all visualizations of phylogenetic trees were produced with FigTree version 1.4.2 and iTOL version 3.5.

Agr-operon and agr-defects identification

Agr-operon genes (Agr A, B, C, D) were initially retrieved from NCBI gene database. Local databases consisting of these genes and genomes were setup. An in-house script was developed to run the local BLAST using USEARCH and the local databases. Frequency of the Agr genes in the included genomes was determined along with the existence of disrupted genes.

Results

ANI and Annotation

FastANI package revealed that only one genome (GCF_008619595.1) was not related to *S. aureus* which was removed and remaining 459 genomes were continued in downstream analysis. Among all the remaining genomes, genome-to-genome genetic distance also confirmed the relation of all the genomes to the same species (data not shown). These genomes were mainly isolated from humans $(n=308)$ and animals $(n=54)$ while isolation history of 98 genomes was unknown. Annotation of all the remaining genomes (n=459) estimated the median number of genes (n=2954) and proteins (n=2770). After annotation, in-silico MLST estimation revealed variety of MLST groups. To better understand those groups, genomes based on the MLST were assigned the clonal complex (CC) as mentioned in literature previously. A total of 10 CC groups were determined with different proportions among the study genomes. Clonal complex 8 being the highest (n=160) while the clonal complex 45 was found to be lowest $(n=6)$ (Fig.1a). Many untypeable and unclassified CC groups were also determined which shows genomes that have large variations and are difficult to classify.

Pangenome

Afterwards, pangenome building revealed that the number of core genes (99% - 100% strains) and soft-core genes (95% - 99% strains) were 2093 and 30, respectively. While number of shell genes (15% - 95% strains) and cloud genes (0% - 15% strains) were revealed to be 761 and 3275, respectively (Fig. 1b). A total of 6159 gene clusters were found to be present in this pangenome. It was determined that low number of core or soft-core genes as compared to accessory or cloud genes show that there is a proportion of genes that is conserved and a mechanism is possibly executed to control the introduction of mutation.

Fig. 1: Comparison of clonal complexes distribution and pan-genome estimation. a. Distribution of Clonal complexes of *S. aureus* genomes after annotation and in-silico MLST analysis. **b.** Distribution of core genes (all genomes), accessory genes (not present in all genomes) and unique genes (only in one genome). Another estimation of pangenome revealed that around 1217 core clusters of genes were present in this pangenomes. While dispensable or accessory genome which includes shell gene clusters contained 10,038 genes. On the other hand, unique genes were found to be 6,781 genes.

Another distribution of genes in pangenome was estimated showing the similar relationship between core genes and accessory genes. On the other hand, imbalance between core-genes and shell/cloud genes or the unique genes and accessory genes was observed. The number of unique genes (only in one organism) is relatively low but the accessory genome is high showing that there is higher number of genes shared or there may be a common source of origin

Core genome analysis

To identify the recombination and diversity among the genomes, core genome analyses was performed (Fig. 2). Core genome analysis revealed that most animal origin and human origin genomes were found clustered separately except in few clades. Mostly animal-origin genomes were found associated with CC-1, unclassified and untypeable clonal complexes while human origin genomes were distributed among other clonal complexes.

Fig. 2: Core genome phylogeny and tree visualization in relation to clonal complexes and origin.

Agr-operon and defects

All the genomes possessed the agr operon (at least one CDS from all the specified agr operon genes) showing that this operon is conserved in the studied sequences. BLAST revealed the various defects in the agr-operons in few genomes (n=28) (Fig. 3). Disruptions such as frameshift, non-synonymous mutations, and presence of insertion sequences (IS-256 and IS-1181) were found in agrC, agrA and agrB genes. AgrC genes were disrupted in 20 genomes while agrA genes were found disrupted in 8 genomes. Only one genome was found with disruptions in agrB (FDA209P, GCA_001548295.1) which has never been reported earlier as per literature review. There was no single organism possessing both agrA and agrC truncations. No truncations or multicopy genes were found in agrD genes. High variability was observed in the starting region of agrC aligned sequences. Truncations in the mid and ending parts were observed. Loss of functionality was observed by finding the binding sites of the agrC and agrA (Table 1).

Fig. 3: Variability and disrupted regions among the sequences of agrC, agrA and agrB genes as per literature review and finding of current study. In agrC, there are parts of protein i.e. (i) Transmembrane Domain (TM) (ii) dimerization and histidine phosphotransfer (DHp) (iii) Catalytic binding activity (CA). Majority of disruptions found in agrC are in TM as compared to DHp and CA. Of 1200 bp sequence, 13 disruptions were found in <600 bp while 8 disruption were found in >600bp. In agrA, protein sequence composed of 800 bp and two parts i.e. Response regulator domain and DNA binding domain. Majority of the disruptions were found in DNA binding domain. In agrB, IS1181 transposase family insertion sequence was found which disrupted this protein. In agrD, no disruption was determined possibly due to small sequence.

Table 1: List of all the genomes in the pangenome possessing the disruption in agr operon. Origin and antibiotic resistance status have been shown in this table as well along with type of disruptions

Association with isolates and Antibiotic resistance:

Majority of agr-defective genomes were isolated from human samples (n=27) while only one genome was of animal origin. On the other hand, most of the agr defective genomes were also revealed to be MRSA (n=18) while 10 genomes were found to be MSSA.

Discussion

Pan-genome studies are important in comparing and studying the hundreds of whole genomes (Guimarães *et al*, 2015). After such comparisons, the identification of the gene clusters of interest is relatively easy and more rational. Based on the core-genome analysis, population structure is estimated and can be related to undermine the specific genes that are needed to study. In this study, the number of unique genes (only in one organism) is relatively low but the accessory genome is high showing that there is higher number of gene sharing or there may be common source of origin. Previously, comparison of core genes and accessory genes have already been reported in other bacterial pangenomes (Page *et al*, 2015; Awan *et al*, 2018). Such studies also show agreement to results in current study.

Agr-defects is important in understanding the protein mechanisms that let the agr operon to disrupt. Agr-operon disruption is mainly dependent on agrA and agrC genes which are important in promoter activations and transcription (Green *et al*, 2023). Current study has determined the presence of another insertion sequence in AgrB gene which was never reported earlier. Mechanisms behind the introduction of this IS and disruption of this gene could reveal many other mechanisms other than chronic infections (Hu and Ng, 2012; Greenrod *et al*, 2023). Chronic infections can introduce such genetic sequences variations but how *S. aureus* controls this time-to-time need remains unknown (Anani *et al*, 2020; Hachani *et al*, 2023).

The results, combined with observations from another study, indicate that this conclusion is supported by previous studies showing loss of operon function. Frameshift indels or nonsense variants can influence splicing patterns, leading to shorter proteins with a possibly maintained partial function(s) (Lalonde *et al*, 2017; Vihinen, 2021). Kapahnke and colleagues determined that CRISPR-Cas9-mediated frameshift indels can randomly modulate the splicing pattern of nearby exons, resulting in shorter proteins, or proteins with new amino acid sequences (Kapahnke *et al*, 2016).

In conclusion, Agr-defects are present in 30 (6.5%) strains of the SA pangenome. agrA, agrB and agrC are the genes found defective. Interestingly, agrB defects have never been reported. Frame shift mutations or small indels are the major cause of this defects. This could be a possible outcome of genome reduction to prolong the infection. This mechanism could show higher occurrence in chronic cases, antibiotic-resistance cases, in advanced age and immunocompromised cases. These defects can lead to even loss of this conserved operon of gram-positive bacteria.

Supplementary Materials

Supplementary table 1: All the 460 genomes with the NCBI provided identifier numbers, strain names that were included in this study.

Declaration of Competing Interest

The authors declare that they have no competing or conflict of interests.

Author Contributions:

Waqas Ahmed: Conceptualization, Methodology, formal analysis, Writing—original draft preparation. **Erman Salih İstifli**: Conceptualization, Methodology, formal analysis, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

References

1. Anani, H., R. Zgheib, I. Hasni, D. Raoult, and P.-E. Fournier. (2020). *Interest of bacterial pangenome analyses in clinical microbiology*. Microbial Pathogenesis 149:104275. doi: <https://doi.org/10.1016/j.micpath.2020.104275>

- 2. Awan, F., M. M. Ali, Y. Dong, Y. Yu, Z. Zeng, and Y. Liu. (2021). *In Silico Analysis of Potential Outer Membrane Beta-Barrel Proteins in Aeromonas hydrophila Pangenome*. International journal of peptide research and therapeutics 27(4):2381-2389. doi: 10.1007/s10989-021-10259-z
- 3. Awan, F., Y. Dong, J. Liu, N. Wang, M. H. Mushtaq, C. Lu, and Y. Liu. (2018). *Comparative genome analysis provides deep insights into Aeromonas hydrophila taxonomy and virulence-related factors*. BMC Genomics 19(1):712. doi: 10.1186/s12864-018- 5100-4
- 4. Bezar, I. F., A. A. Mashruwala, J. M. Boyd, and A. M. Stock. (2019). *Drug-like Fragments Inhibit agr-Mediated Virulence Expression in Staphylococcus aureus*. Scientific Reports 9(1):6786. doi: 10.1038/s41598-019-42853-z
- 5. Fowler, V. G., Jr., G. Sakoulas, L. M. McIntyre, V. G. Meka, R. D. Arbeit, C. H. Cabell, M. E. Stryjewski, G. M. Eliopoulos, L. B. Reller, G. R. Corey, T. Jones, N. Lucindo, M. R. Yeaman, and A. S. Bayer. (2004). *Persistent bacteremia due to methicillin-resistant Staphylococcus aureus infection is associated with agr dysfunction and low-level in vitro resistance to thrombin-induced platelet microbicidal protein*. The Journal of infectious diseases 190(6):1140-1149. doi: 10.1086/423145
- 6. Green, M. J., E. J. Murray, P. Williams, A. M. Ghaemmaghami, J. W. Aylott, and P. M. Williams. (2023). *Modelled-Microgravity Reduces Virulence Factor Production in Staphylococcus aureus through Downregulation of agr-Dependent Quorum Sensing*. International Journal of Molecular Sciences 24(21)doi: 10.3390/ijms242115997
- 7. Greenrod, S. T. E., M. Stoycheva, J. Elphinstone, and V. P. Friman. (2023). *Influence of insertion sequences on population structure of phytopathogenic bacteria in the Ralstonia solanacearum species complex*. Microbiology 169(7)doi: 10.1099/mic.0.001364
- 8. Guimarães, L. C., J. Florczak-Wyspianska, L. B. de Jesus, M. V. Viana, A. Silva, R. T. Ramos, C. Soares Sde, and C. Soares Sde. (2015). *Inside the Pangenome - Methods and Software Overview*. Curr Genomics 16(4):245-252. doi: 10.2174/1389202916666150423002311
- 9. Hachani, A., S. G. Giulieri, R. Guerillot, C. J. Walsh, M. Herisse, Y. M. Soe, S. L. Baines, D. R. Thomas, S. D. Cheung, A. S. Hayes, E. Cho, H. J. Newton, S. Pidot, R. C. Massey, B. P. Howden, and T. P. Stinear. (2023). *A high-throughput cytotoxicity screening platform reveals agr-independent mutations in bacteraemia-associated Staphylococcus aureus that promote intracellular persistence*. Elife 12doi: 10.7554/eLife.84778
- 10. Hu, J., and P. C. Ng. (2012). *Predicting the effects of frameshifting indels*. Genome Biology 13(2):R9. doi: 10.1186/gb-2012-13-2-r9

MEPIBIO

- 11. Inagaki, R., A. Koshiba, E. Nasuno, and N. Kato. (2024). *Eliminating extracellular autoinducing peptide signals inhibits the Staphylococcus aureus quorum sensing agr system*. Biochemical and Biophysical Research Communications 711:149912. doi: 10.1016/j.bbrc.2024.149912
- 12. Kapahnke, M., A. Banning, and R. Tikkanen. (2016). *Random Splicing of Several Exons Caused by a Single Base Change in the Target Exon of CRISPR/Cas9 Mediated Gene Knockout*. Cells 5(4)doi: 10.3390/cells5040045
- 13. Lalonde, S., O. A. Stone, S. Lessard, A. Lavertu, J. Desjardins, M. Beaudoin, M. Rivas, D. Y. R. Stainier, and G. Lettre. (2017). *Frameshift indels introduced by genome editing can lead to in-frame exon skipping*. PLOS ONE $12(6):e0178700.$ doi: 10.1371/journal.pone.0178700
- 14. Meier-Kolthoff, J. P., A. F. Auch, H. P. Klenk, and M. Göker. (2013). *Genome sequence-based species delimitation with confidence intervals and improved distance functions*. BMC Bioinformatics 14:60. doi: 10.1186/1471-2105-14-60
- 15. Meier-Kolthoff, J. P., J. S. Carbasse, R. L. Peinado-Olarte, and M. Göker. (2022). *TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes*. Nucleic acids research 50(D1):D801-d807. doi: 10.1093/nar/gkab902
- 16. Omer, S. A., A. R. Ganjo, S. H. Haji, and S. B. Smail. (2024). *Accessory Gene Regulator (agr) group polymorphisms in methicillin-resistant Staphylococcus aureus and its association with biofilm formation*. Cellular and Molecular Biology 70(4):1-7. doi: 10.14715/cmb/2024.70.4.1
- 17. Osbourn, A. E., and B. Field. (2009). *Operons*. Cellular and molecular life sciences 66(23):3755-3775. doi: 10.1007/s00018-009-0114-3
- 18. Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. G. Holden, M. Fookes, D. Falush, J. A. Keane, and J. Parkhill. (2015). *Roary: rapid largescale prokaryote pan genome analysis*. Bioinformatics 31(22):3691-3693. doi: 10.1093/bioinformatics/btv421 %J Bioinformatics
- 19. Schlievert, P., S. Kilgore, and D. Leung. (2023). *Agr Regulation of Streptococcal Pyrogenic Exotoxin A in Staphylococcus aureus*. MicroPublication Biology 2023doi: 10.17912/micropub.biology.000795
- 20. Tenaillon, O., J. E. Barrick, N. Ribeck, D. E. Deatherage, J. L. Blanchard, A. Dasgupta, G. C. Wu, S. Wielgoss, S. Cruveiller, C. Médigue, D. Schneider, and R. E. Lenski. (2016). *Tempo and mode of genome evolution in a 50,000-generation experiment*. Nature 536(7615):165-170. doi: 10.1038/nature18959
- 21. Tsuji, B. T., M. J. Rybak, K. L. Lau, and G. Sakoulas. (2007). *Evaluation of accessory gene regulator (agr) group and function in the proclivity towards vancomycin intermediate resistance in Staphylococcus aureus*. Antimicrobial agents and chemotherapy 51(3):1089-1091. doi: 10.1128/aac.00671-06
- 22. Vihinen, M. (2021). *Functional effects of protein variants*. Biochimie 180:104-120. doi: <https://doi.org/10.1016/j.biochi.2020.10.009>
- 23. Vinodhini, V., and M. Kavitha. (2024). *Deciphering agr quorum sensing in Staphylococcus aureus:*

insights and therapeutic prospects. Molecular Biology Reports 51(1):155. doi: 10.1007/s11033-023-08930-3

24. Wright, J. S., 3rd, R. Jin, and R. P. Novick. (2005). *Transient interference with staphylococcal quorum sensing blocks abscess formation*. Proceedings of the National Academy of Sciences of the United States of America 102(5):1691-1696. doi: 10.1073/pnas.0407661102