Using Whole Genome Sequencing to Genetically Profile and Analyze *Escherichia coli* Isolates with Varying Resistance to β-Lactam/β-lactamase Inhibitor Combinations

Hatim Amiji, B.S.[1], Benjamin Strope, B.S., M.S.[2], Pranoti Sahasrabhojane M.S.[2], Marc Gohel, B.S.[2], David Greenberg, M.D.[3], Samuel Shelburne, M.D., Ph.D.[2], William Shropshire, Ph.D.[2]

[1] Frank H. Netter MD School of Medicine at Quinnipiac University, North Haven, CT
[2] MD Anderson Cancer Center, Houston, TX
[3] University of Texas Southwestern School of Medicine, Dallas, TX

Introduction

Beta lactams are the most widely used antibiotics, accounting for 65% of all prescriptions for injectable antibiotics in the United States.[1] They are often used in conjunction with beta lactamase inhibitors to circumvent the hydrolysis of beta lactams, which is a primary mechanism for neutralizing these antibiotics.[1,2] Due to the widespread distribution of beta-lactam/beta-lactamase inhibitor combinations, there has been an increase in evolutionary pressure selecting for its resistance. Hence, we aim to use rapidly advancing Whole Genome Sequencing technology to improve the utilization of BL/BLI. In a previous study, we identified high levels of discordance for BL/BLI genotype-phenotype correlations when compared to other antibiotics when solely considering the presence/absence of antimicrobial genes.[3] Herein, we aimed to test whether adding gene amplification along with analysis of antimicrobial gene promoter mutations in addition to gene presence/absence improves genotype phenotype correlation for a variety of BL/BLI combinations.

Materials and Methods

Antimicrobial Sensitivity Testing of Bacterial Isolates

Antimicrobial susceptibility profiles of *Escherichia coli* bacteremia causing isolates were identified via Vitek2 through the MD Anderson clinical microbiology laboratory for three BL/BLI combinations: ampicillin/sulbactam (SAM), amoxicillin/clavulanic acid (AMC), and piperacillin/tazobactam (TZP). Isolates were selected for further analysis when sensitive to ceftriaxone and had various resistant patterns to the BL/BLI combinations (i.e. extended spectrum of resistance to B-lactam/B-lactamase inhibitors – ESRI).

Whole Genome Sequencing of Bacterial Isolates

gDNA was extracted using the QIAGEN DNEasy Blood & Tissue Kit. DNA was prepared for sequencing using the Illumina DNA Prep kit. Isolates were sequenced using the NovaSeq6000 platform using 150-PE reads at approximately 120X coverage depths. Fastq reads were processed and QC/QA was performed using a bespoke short-read pipeline. Assemblies were made for each isolate using the SPADES assembler.[4]

CNV Quantification:

Processed reads were inputted into a custom script named Copy Number Variant Quantification Tool (CONVICT)[5] to generate ploidy-agnostic copy numbers at the gene level. Briefly, the tool uses a read depth normalization approach whereby target gene reads are counted and normalized to the average control gene read count. Target genes were determined using KmerResistance-2.2.0.[6]

Differential CNV Analysis

Non-parametric one-way ANOVA tests were used to determine which genes differed significantly across groupings at an alpha of 0.05. Post-hoc tests were then done using pairwise Mann-Whitney U tests at an alpha of 0.05 to find groupings that differed significantly from each other by a given gene's copy number. Statistical analysis done using R (4.1.2) and scatter plots made using GGally[7] with Kendall correlation coefficient

t-SNE of Isolates Using CNVs as Inputs.

t-SNE to visualize potential group-based clustering because of its ability to find population structure . The R package Rtsne[8] was implemented. The t-SNE was parametrized by visually inspecting the output of a range of perplexity and max iteration parameters. The initial principal component parameter was determined by selecting the principal components that captured 95% of the variation in the starting dataset.

Phylogenetic Analysis of Isolates

Genome assemblies were first annotated using Prokka.[9] Annotated .gff files were then used to produce core genome alignment file using Panaroo.[10] We used the MAFT[11] aligner, with strict mode and maintaining genes at a 99% core threshold. The core genome alignment file then used to generate a phylogeny using IQtree2[12] with maximum likelihood model finding, bootstrapping, and branch testing. Bootstrapped consensus tree midpoint rooted with R package Phytools[13] and visualized with R package Ape[14]

blaTEM-1B Promoter Curation

Abricate[15] was used to find contig locations of bla_{TEM} genes within isolates using assemblies as inputs. In silico PCR[16] was used in combination with Abricate output to find promoter regions of isolates containing a hit. Promoters were then aligned to the consensus P3 promoter[17] and was manually sorted through using SnapGene to classify each promoter sequence. To promoters that did not align well with the consensus P3, we aligned the IS26 promoter to find the presence of hybrid promoters and or promoter sequences that were not easily discernible by any known motifs.

Results

Epidemiology of Isolates

- After antimicrobial sensitivity testing, isolates were grouped into Group1, Group 2, Group 3, and Group 4 to correspond to the BL/BLI groupings.
- Group 1 SAM/AMC/TZP susceptible
- Group 2 SAM resistant, AMC/TZP susceptible
- Group 3 SAM/AMC resistant, TZP susceptible
- Group 4 SAM/AMC/TZP resistant.

Group level β-lactamase gene CNVs Differentiate BL/BLI Groupings

- The 109 isolates contained a total of 13 different β-lactamase genes
- Non-parametric one-way ANOVA showed the only significantly different β-lactamase gene CNVs between groups were bla_{OXA-1} ($p < 0.05$) and bla_{TEM-1B} ($p < 0.05$), Table 1.
- Post-hoc tests show that bla_{OXA-1} presence predominates in Group 4 isolates ($p < 0.05$) in addition to its relatively high CNV and that bla_{TEM-1B} differentiates Group 3 from Group 1 ($p < 0.05$), Table 1.
- Kendall rank correlation found that bla_{OXA-1} CNV was significantly correlated with TZP resistance ($R = 0.48$, $p < 0.05$) and bla_{TEM-1B} CNV was significantly positively correlated with AMC ($R = 0.29$, $p < 0.05$) and SAM ($R = 0.38$, $p < 0.05$) resistance.

t-SNE Unsupervised Clustering of CNV Covariates Separates BL/BLI Groupings

- Statistically significantly different gene CNVs found previously were used to generate a t-SNE plot and compared to a t-SNE generated using only β-lactamase gene CNV covariates whereby each dataset produced similar plots. Thus, we further showed that bla_{OXA-1} and bla_{TEM-1B} alone can elucidate population structure.

BL/BLI Phenotypic groupings hold across core element derived population structure

- Multilocus Sequence Typing (MLST) groups cluster around suspected ESRI mediating genotypes.
- Isolates with MSLT 648 are both in Group 4 and are bla_{OXA-1} positive
- Isolates with MLST 1193 are more commonly Group 2 with low bla_{TEM-1B} CNVs
- MSLT 131 contains a heterogenous group of ESRI phenotypes, which is not entirely explained by non-core elements

bla_{TEM-1B} promoters weakly explain BL/BLI Phenotypic groupings

- Canonical P3 promoter in 61 isolates
- Presence of a P3 IS26 hybrid promoter in 7 isolates
- 5/7 of these isolates had a hybrid promoter with a spacer region of greater than 20-bp, of which 4 were in Group 1 and 1 was in Group 3 Isolates with MSLT 648 are both in Group 4 and are bla_{OXA-1} positive
- 2/7 of the isolates had a hybrid promoter with a spacer region of less than 20-bp, both belonging to Group 3

Table 1. Post-hoc Comparisons of CNVs Between Groups with Reported Mean Difference of CNVs

Comparison	*aac6.Ib*	*aadA5*	*blaOXA-1*	*blaTEM-1B*	*catA1*	*catB3*	*qacE*	*tetB*
Group 4 vs Group 1	3.45*	2.38*	3.45*	3.56	0.42	2.25*	2.16*	0.75*
Group 4 vs Group 2	3.45*	2.51*	3.45*	3.11	0.36	2.25*	2.14	0.5
Group 4 vs Group 3	3.41*	2.18	3.42*	1.07	0.33	2.25*	1.9	0.69*
Group 3 vs Group 1	0.03	0.2	0.03	2.49*	0.09	0	0.26	0.05
Group 3 vs Group 2	0.03	0.33	0.03	2.04	0.02	0	0.23	0.21
Group 2 vs Group 1	0	0.13	0	0.46	0.06	0	0.03	0.26

Table 1. Comparisons done between absolute difference of CNVs first using a one-way ANOVA followed by post hoc tests with pairwise Mann-Whitney U tests with an alpha of 0.05. Group 4 is defined primarily by bla_{OXA-1} gene presence, while Group 3 is differentiated from Group 1 by bla_{TEM-1B} relative CNVs. Group 2 isolates do not differentiate from 3 or 1.

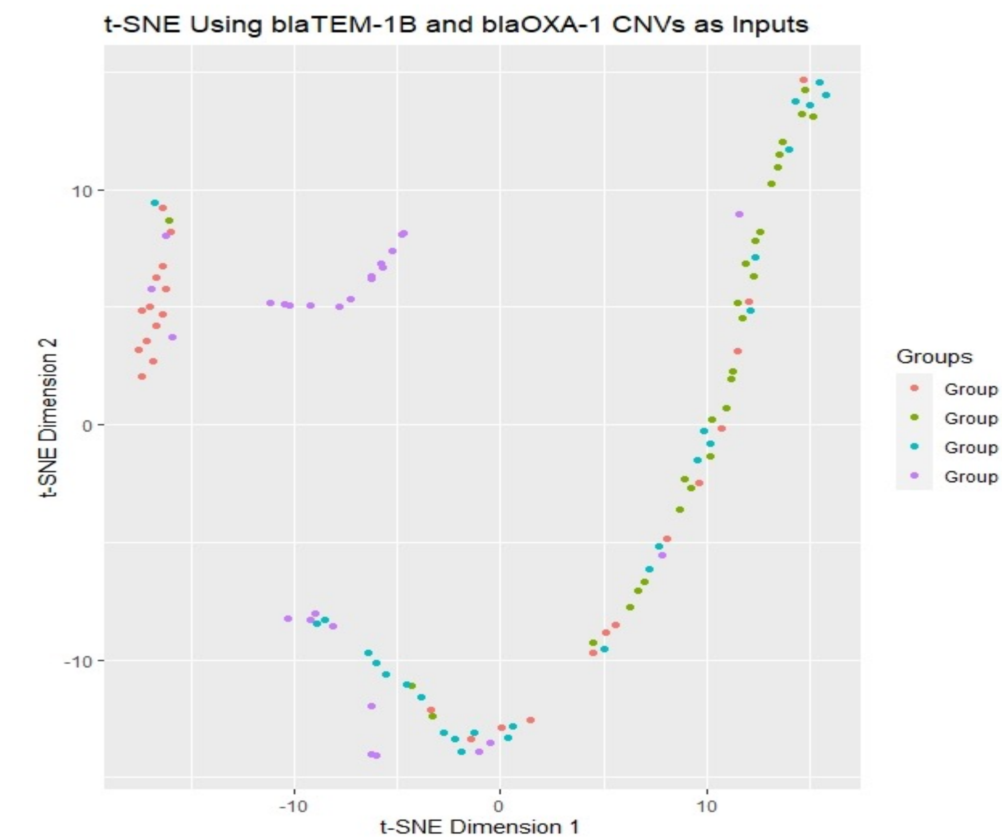

Figure 1. t stochastic neighbor embedding (t-SNE) plot with 150 iterations and a perplexity of 7 using bla_{OXA-1} and bla_{TEM-1B} genes as covariates separates groups based on phenotype.

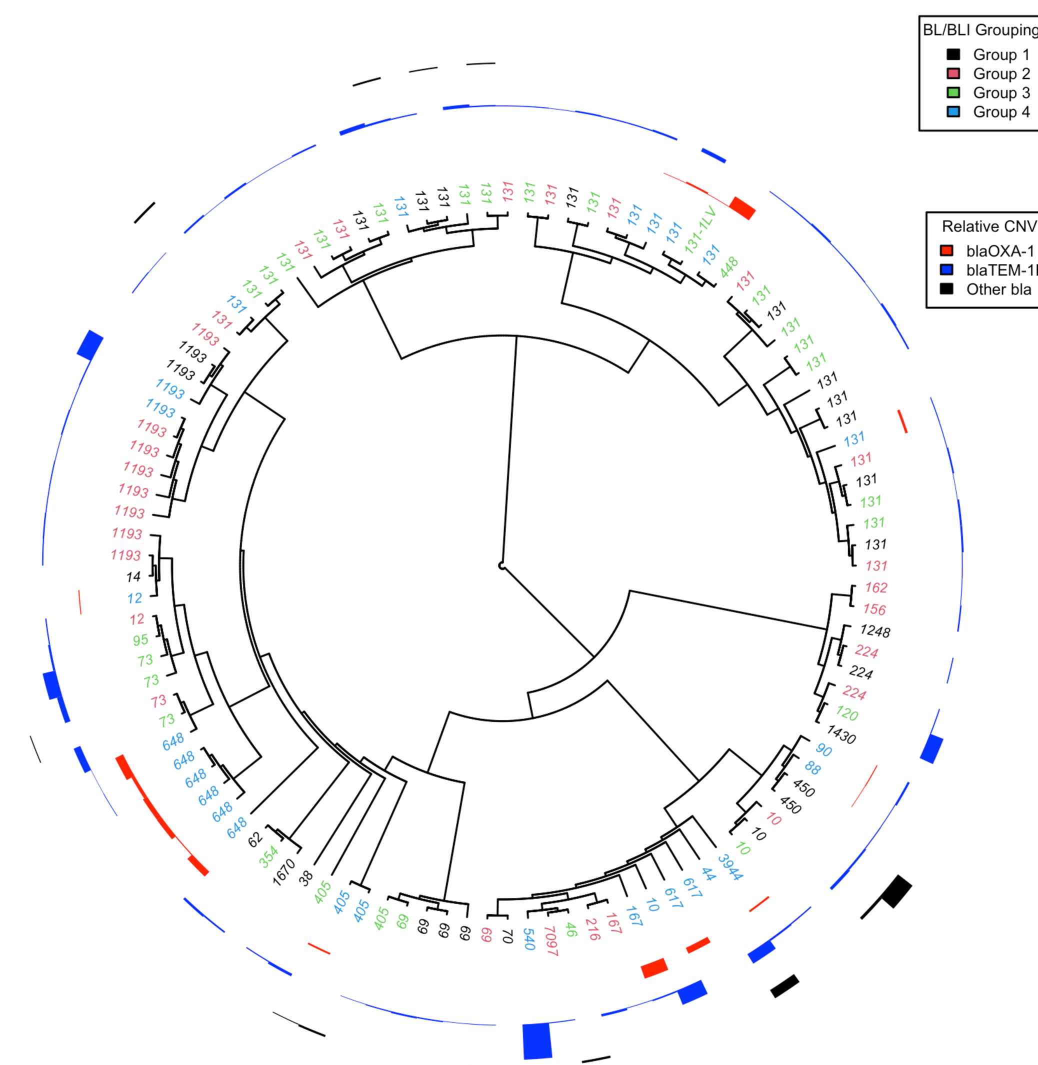

Figure 2. Midpoint rooted Maximum likelihood phylogeny using a 99% core genome with bootstrapping. MLSTs colored by ESRI phenotypic groupings. Relative bla genes overlayed onto the tree bla_{OXA-1} (red), bla_{TEM-1B} (blue), all other *bla* (black).

Conclusions

- β-lactamase genes bla_{OXA-1} and bla_{TEM-1B} differentiate the resistance patterns seen between Groups 4, 3, and 1. Particularly, bla_{OXA-1} differentiates Group 4 from all other groups, while bla_{TEM-1B} CNVs differentiate Group 3 from Group 1. This is evident in the statistical tests done (Table 1) and the unsupervised clustering accomplished with the t-SNE (Figure 1). Moreover, the clustering shows clear separation between Group 4, Group 1, and Group 3 isolates, but within two clusters, Group 3 and 2 are intertwined pointing to more nuanced method of classification that is likely needed, which would expand on the number of groups present. More refined groupings are likely due to mechanisms outside the β-lactamase genes or involve amplifications/deletions of other members of specific drug resistance pathways not accounted for in this single gene model. However, we did see that resistance patterns across the antibiotics, especially isolates seen in Group 2, are not uniform, and would likely require further refining to find intermediate high/low phenotypes.
- Specific MLSTs align with specific phenotypic groupings, but others show that non-core elements alone do not explain the varying distributions of phenotypes. MLST 648 and 1193 fall into the Group 4 and Group 2 isolates respectively, while MLST 131 has a varying distribution without β-lactamase genes explaining the variation.
- Promoter differences are likely not the reason for the phenotypic differences. Our curation of promoters found that a majority of the isolates with curatable promoters (n=81) contained the P3 promoter (n=61). While some (n=3) Group 3 isolates had the strong promoter Pa/Pb, most variation came from P3 promoters with IS26 spacers (n=7), generating hybrid promoters. Smaller spacers, <20-bp (n=2), were found in Group 3, while of the larger spacers, > 20-bp (n=5) , four were in Group 1 and one was in Group 3. This points to the potential effect of the P3-IS26 hybrid promoter in potentiating promoter strength, based on the spacing/promoter length.[18]

References

1. Bonomo, R. A. (2017). β-Lactamases: A Focus on Current Challenges. *Cold Spring Harbor Perspectives in Medicine*, *7*(1), a025239. https://doi.org/10.1101/cshperspect.a025239
2. Bush, K., & Bradford, P. A. (2016). β-Lactams and β-Lactamase Inhibitors: An Overview. *Cold Spring Harbor Perspectives in Medicine*, *6*(8), a025247. https://doi.org/10.1101/cshperspect.a025247
3. Shelburne, S. A., Kim, J., Munita, J. M., Sahasrabhojane, P., Shields, R. K., Press, E. G., Li, X., Arias, C. A., Cantarel, B., Jiang, Y., Kim, M. S., Aitken, S. L., & Greenberg, D. E. (2017). Whole-Genome Sequencing Accurately Identifies Resistance to Extended-Spectrum β-Lactams for Major Gram-Negative Bacterial Pathogens. *Clinical Infectious Diseases*, *65*(5), 738–745. https://doi.org/10.1093/cid/cix417
4. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics*, *70*(1). https://doi.org/10.1002/cpbi.102
5. William C. Shropshire, CONVICT, python version 3.9. . https://github.com/wshropshire/convict
6. Philip T.L.C. Clausen, Frank M. Aarestrup & Ole Lund, "Rapid and precise alignment of raw reads against redundant databases with KMA", BMC Bioinformatics, 2018;19:307.
7. Barret Schloerke, Di Cook, Joseph Larmarange, Francois Briatte, Moritz Marbach, Edwin Thoen, Amos Elberg and Jason Crowley (2021). GGally: Extension to 'ggplot2'. R package version 2.1.2. https://CRAN.R-project.org/package=GGally
8. Jesse H. Krijthe (2015). Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation, URL: https://github.com/jkrijthe/Rtsne
9. Seemann T. *Prokka: rapid prokaryotic genome annotation* Bioinformatics 2014 Jul 15;30(14):2068-9. PMID:24642063
10. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol 21:180.
11. Katoh, Standley 2013 (*Molecular Biology and Evolution* 30:772-780) MAFFT multiple sequence alignment software version 7: improvements in performance and usability.
12. B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, R. Lanfear (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.*, 37:1530-1534. https://doi.org/10.1093/molbev/msaa015
13. Revell, L. J. (2012) phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3 217-223. doi:10.1111/j.2041-210X.2011.00169.x
14. Paradis E. & Schliep K. 2019. ape 5.0: an environment or modern phylogenetics and evolutionary analyses in R. Bioinformatics 35: 526-528.
15. Seemann T, *Abricate*, Githubhttps://github.com/tseemann/abricate
16. Kent, J. (n.d.). *In Silico PCR* (Version 37x1) [Bash; Linux].
17. Lartigue, M. F., Leflon-Guibout, V., Poirel, L., Nordmann, P., & Nicolas-Chanoine, M.-H. (n.d.). *Promoters P3, Pa/Pb, P4, and P5 Upstream from blaTEM Genes and Their Relationship to ߢ-Lactam Resistance*. 3.
18. Klein, C. A., Teufel, M., Weile, C. J., & Sobetzko, P. (2021). The bacterial promoter spacer modulates promoter strength and timing by length, TG-motifs and DNA supercoiling sensitivity. *Scientific Reports*, *11*(1), 24399. https://doi.org/10.1038/s41598-021-03817-4