Exploring the Relationship Between Observed Plasmid Host Range and Antimicrobial Resistance Gene Count

Introduction

Plasmids are extrachromosomal DNA molecules that replicate independently from the chromosome and are key players in horizontal gene transfer(HGT)(1). Through HGT, plasmids can be exchanged between bacterial cells allowing for the sharing of genetic data and promoting rapid diversification and adaptation(2). Plasmids can be shared between distinct taxonomic groups, however, the full determinants underlying this host specificity remain poorly understood(3). Generally, plasmids are classified into three categories, conjugative, mobilizable, and non-mobilizable, based on their ability to transfer between bacterial cells(4). Conjugative plasmids can independently facilitate their own transfer, mobilizable plasmids require assistance from other plasmids to transfer, and non-mobilizable plasmids are incapable of transferring between cells under normal conditions.

Plasmids play a pivotal role in bacterial evolution, often carrying genes that enhance host fitness. Notably, they often harbor antimicrobial resistance genes(ARGs) and virulence factors, which, upon transfer, allows the new host to express these genes and can lead to increased survival under selective pressures(5). The global rise of antimicrobial resistance(AMR) has been largely attributed to plasmid-mediated HGT, which accelerates the dissemination of ARGs(1). In fact, plasmid-mediated transfer of ARGs has been linked to a wide array of outbreaks(6-8) with separate, distinct species carrying the same ARG-encoding plasmid.

Utilizing the Plasmid Database (PLSDB), a comprehensive resource containing over 70,000 high-quality, publicly available plasmid sequences as well as its metadata(9), I explored the relationships between predicted host range and the number of ARGs and developed a logistic regression model to predict whether a plasmid exhibits a wide host range.

Results

The PLSDB provides detailed information on the observed range of hosts that harboured a given plasmid based on NCBI annotations. This host range is categorized into six taxonomic classifications: genus, family, order, class, phylum, and multi-phyla. The database also includes information on ARGs present in each plasmid. After filtering for plasmids with data on host range, a total of 63,010 records were found, comprising of 18,362 records at the genus level, 9,129 at the family level, 20,641 at the order level, 6,896 at the class level, 1,426 at the phylum level, and 6,556 at the multi-phyla level. The number of ARGs for each host range is shown here(Figure. 1). One-way ANOVA analysis revealed that not all host range ranks contained the same number of ARGs. Pairwise comparisons with Tukey's adjustment indicated that most rank pairs exhibited statistically

significant differences in ARG counts, with the exceptions of the genus-family and class-phylum pairs. Plasmids with the genus host range had the lowest number of ARGs, while those at the class level harbored the highest counts.
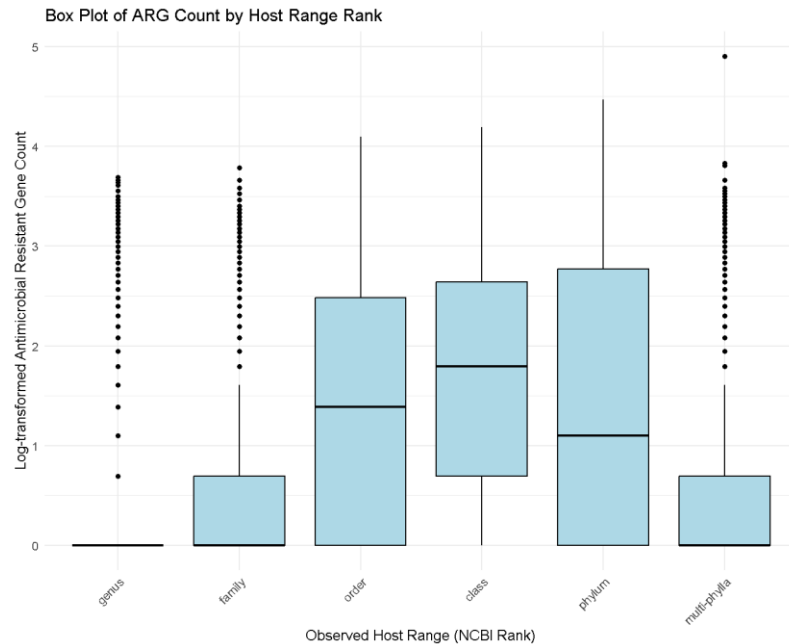

Figure 1. A box plot of the log ARG count for each observed host range.

A logistic regression model was fit to investigate the factors influencing plasmid host specificity. The predicted outcome, host range, was transformed into a binary classification of narrow (restricted to a single genus) or wide (a host range of family or wider). The model included ARG count, mobilization type (non-mobile, mobile, or conjugative), and plasmid size quartile as predictors. In the model, linearity was observed between ARG count and the log odds of wide host range(Supplementary Figure 1). All coefficients were statistically significant, with p-values less than 2e-16 (Figure 2). The analysis revealed that an increase in ARG count significantly increased the odds of a plasmid having a wide host range, with each additional ARG count contributing a 1.191 increase in odds(95% CI 1.183 to 1.198). Plasmids in the smallest size quartile exhibited the highest likelihood of being widespread, with a difference in odds of 3.190(95% CI 3.030 to 3.387) compared to the second size quartile. Furthermore, the non-mobilizable plasmids had the lowest likelihood of being widespread with a difference of odds of 0.183 (95% CI 0.174 to 0.192) compared to conjugative plasmids.
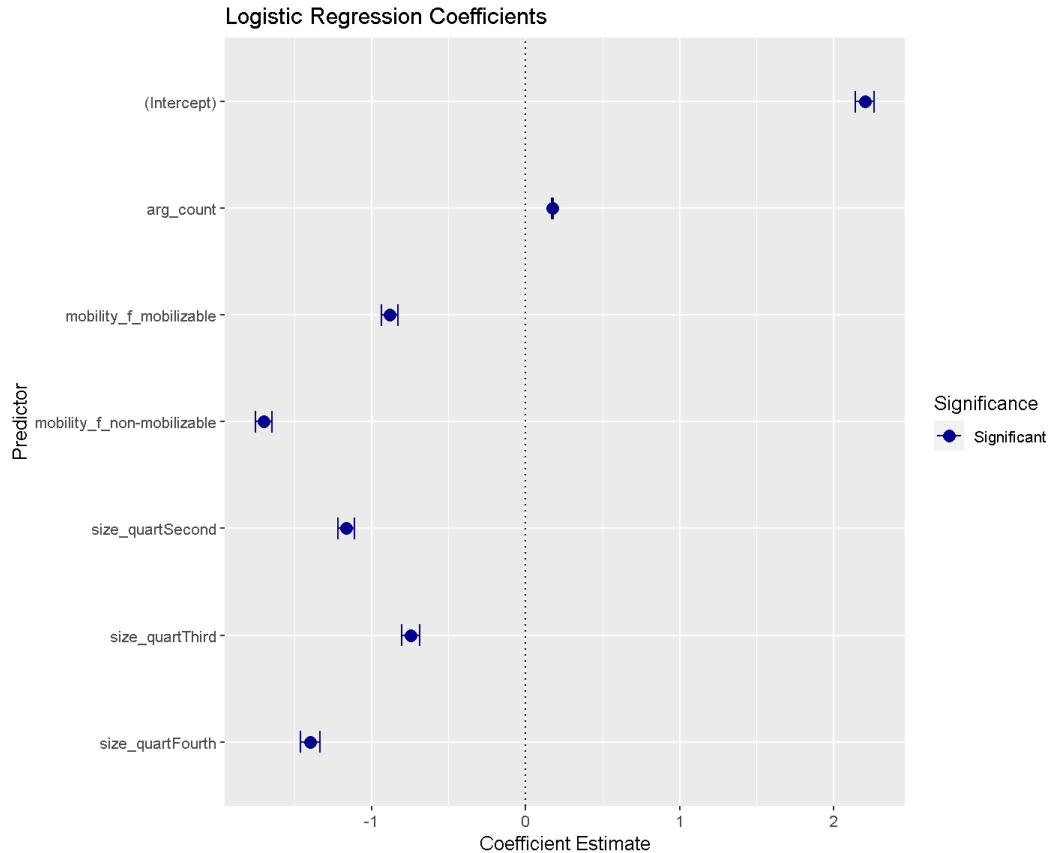
Figure 2. A dot plot of the directionality and magnitude of the coefficients for each predictor from the logistic regression. Dots are colored blue if statistically significant(p-value <0.05) and a 95% confidence interval is shown.

Discussion

The global AMR crisis is closely linked to plasmids, which play a significant role in the dissemination of ARGs. Plasmid-encoded ARGs provide a fitness advantage to their bacterial hosts, potentially enhancing the plasmid's ability to persist and spread. ANOVA analysis revealed significant differences in ARG counts across host specificity ranks, with most pairwise comparisons showing increasing ARG counts as host range broadens until the phylum host range where it starts to decrease. This suggests that the accumulation of ARGs may be a critical factor in expanding plasmid host range, providing a benefit to the host cell. However, the decrease in the number of ARGs in phylum or multi-phylla ranks suggests that, in extremely widespread plasmids, there is a potential fitness cost of ARGs. While ARGs have been shown to put metabolic strain on bacterial cells(10), more work needs to be done to study the specific trade-off of ARG accumulation in widespread plasmids.

The fitted logistic regression model predicted whether a plasmid has a narrow host range (can only be spread within a single genus) or a widespread range(can be spread amongst a family or wider). It demonstrated a positive relationship between the number of ARGs and the likelihood of a

plasmid being widespread, reinforcing the idea that ARGs may play a role in plasmid persistence and mobility across diverse hosts. Smaller plasmids, in the first quartile, were found to have greater odds of having a wide host range compared to the larger size quartiles. This observation aligns with previous studies(11), which suggest that larger plasmids face greater challenges during conjugation, potentially limiting their spread. Additionally, conjugative plasmids, known for their ability to transfer between hosts, were unsurprisingly associated with a broader host range in the logistic regression model compared to mobile and non-mobile plasmids.

These findings underscore the need for further research into plasmid host specificity and its role in ARG dissemination. Understanding the factors that drive plasmid distribution and mobility is crucial for addressing the growing AMR crisis and developing strategies to mitigate its impact.

Methods

All relevant code for this project is available at: https://github.com/mayee123/Stat-605-Project. Briefly, the metadata from the PLSDB v. 2024_05_31_v2 was downloaded. The number of antimicrobial resistance genes was counted and the metadata tables were merged using the *data_wrangling.py* file. A one-way ANOVA and pairwise comparisons using Tukey's adjustment were performed using the "car" and "emmeans" library in R and a logistic regression was fit using the glm function. The complete statistical analysis, including data processing and visualization, is documented in the *analysis_and_figures.R* file
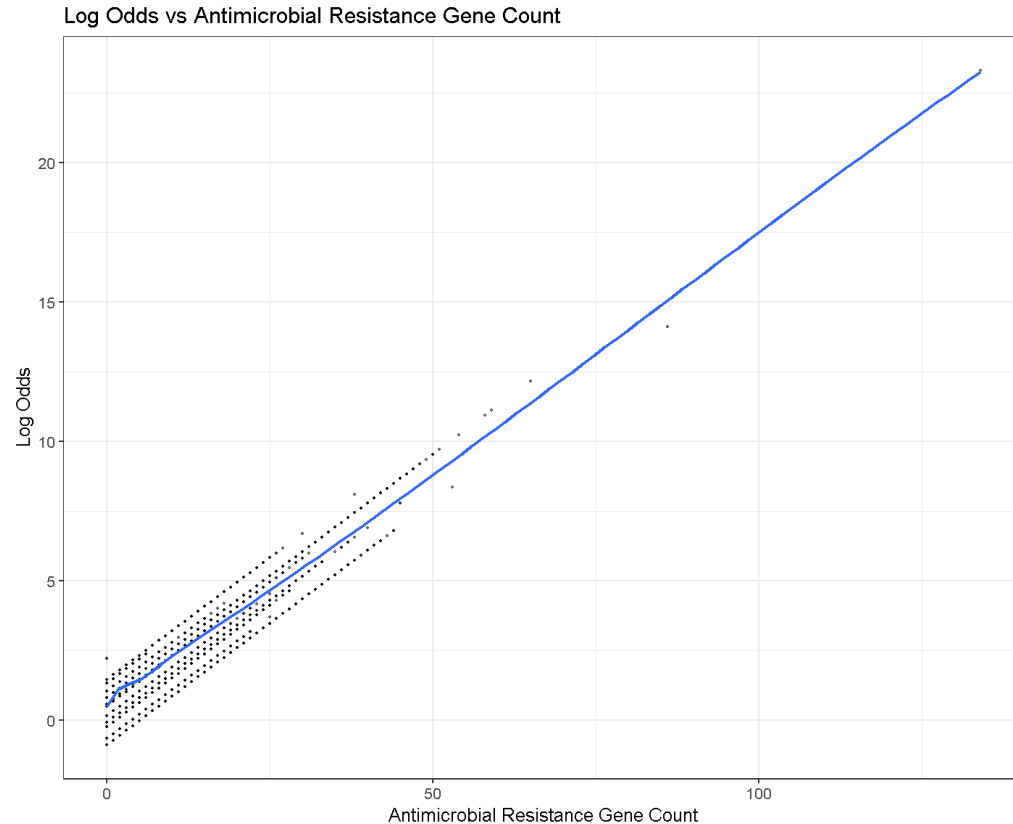
Supplemental Figure 3. Scatter plot of the log odds of wide plasmid host range vs antimicrobial resistance gene count. A LOWESS curve is shown in blue.

References

1. Castañeda-Barba, S., E. M. Top, and T. Stalder. 2024. Plasmids, a molecular cornerstone of antimicrobial resistance in the One Health era. *Nature Reviews Microbiology* 22: 18-32.
2. Benz, F., and Alex R. Hall. 2023. Host-specific plasmid evolution explains the variable spread of clinical antibiotic-resistance plasmids. *Proceedings of the National Academy of Sciences* 120: e2212147120.
3. Aytan-Aktug, D., P. T. L. C. Clausen, J. Szarvas, P. Munk, S. Otani, M. Nguyen, J. J. Davis, O. Lund, and F. M. Aarestrup. 2022. PlasmidHostFinder: Prediction of Plasmid Hosts Using Random Forest. *mSystems* 7: e01180-01121.
4. Ramsay, J. P., S. M. Kwong, R. J. Murphy, K. Yui Eto, K. J. Price, Q. T. Nguyen, F. G. O'Brien, W. B. Grubb, G. W. Coombs, and N. Firth. 2016. An updated view of plasmid conjugation and mobilization in Staphylococcus. *Mob Genet Elements* 6: e1208317.
5. Newbury, A., B. Dawson, U. Klümper, E. Hesse, M. Castledine, C. Fontaine, A. Buckling, and D. Sanders. 2022. Fitness effects of plasmids shape the structure of bacteria-plasmid interaction networks. *Proc Natl Acad Sci U S A* 119: e2118361119.
6. Getino, M., M. López-Díaz, N. Ellaby, J. Clark, M. J. Ellington, and R. M. La Ragione. 2022. A Broad-Host-Range Plasmid Outbreak: Dynamics of IncL/M Plasmids Transferring Carbapenemase Genes. *Antibiotics (Basel)* 11.

7. Mathers, A. J., H. L. Cox, B. Kitchel, H. Bonatti, A. K. C. Brassinga, J. Carroll, W. M. Scheld, K. C. Hazen, and C. D. Sifri. 2011. Molecular Dissection of an Outbreak of Carbapenem-Resistant Enterobacteriaceae Reveals Intergenus KPC Carbapenemase Transmission through a Promiscuous Plasmid. *mBio* 2: 10.1128/mbio.00204-00211.

8. Wan, Y., A. C. Myall, A. Boonyasiri, F. Bolt, A. Ledda, S. Mookerjee, A. Y. Weiße, M. Getino, J. F. Turton, H. Abbas, R. Prakapaite, A. Sabnis, A. Abdolrasouli, K. Malpartida-Cardenas, L. Miglietta, H. Donaldson, M. Gilchrist, K. L. Hopkins, M. J. Ellington, J. A. Otter, G. Larrouy-Maumus, A. M. Edwards, J. Rodriguez-Manzano, X. Didelot, M. Barahona, A. H. Holmes, E. Jauneikaite, and F. Davies. 2024. Integrated Analysis of Patient Networks and Plasmid Genomes to Investigate a Regional, Multispecies Outbreak of Carbapenemase-Producing Enterobacterales Carrying Both blaIMP and mcr-9 Genes. *The Journal of Infectious Diseases* 230: e159-e170.

9. Galata, V., T. Fehlmann, C. Backes, and A. Keller. 2019. PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res* 47: D195-d202.

10. Rajer, F., and L. Sandegren. 2022. The Role of Antibiotic Resistance Genes in the Fitness Cost of Multiresistance Plasmids. *mBio* 13: e0355221.

11. Li, C. c., R. Hu, X. m. Hua, Y. x. Ni, L. Ge, L. Zhang, W. Yu, N. x. Hao, H. Xia, Q. Fang, and Z. y. Tao. 2023. Construction and functional verification of size-reduced plasmids based on TMP resistance gene dfrB10. *Microbiol Spectr* 11: e0120623.