# Term Project:
# Methicillin-Resistant *Staphylococcus Aureus* (MRSA)

Riley Adams

9th July, 2022

# 1 Introduction

You have probably heard of a "Staph Infection", but why is this seemingly common bacterial skin infection becoming a much bigger deal? The simple answer is that the bacteria which causes these skin infections, *Staphylococcus aureus*, has been evolving to develop resistance against the antibiotics which are used to treat it. (Pantosti and Venditti 2009) This new distinct type of *S. aureus* has been deemed Methicillin-resistant *Staphylococcus aureus*, commonly referred to as MRSA. MRSA first started appearing in healthcare settings, where it caused mortality for hospitalized individuals. However, in the early 1990's a biologically distinct type of MRSA began breaking out in the the greater community, causing severe infections and even mortality among individuals who did not exhibit ill health or risk factors associated with early death. (Kajita et al. 2007) Thus, we now have two types of MRSA: HA-MRSA (healthcare-associated MRSA), and CA-MRSA (community-associated MRSA).

For the studies I conducted along with Ryan Campbell and Aditya Kurkut, we took particular interest in CA-MRSA. Those community members who are at highest risk of infection by MRSA are those in densely populated settings with many shared textiles or possibility of close contact. (Kajita et al. 2007) This can include day cares, sports teams, correctional facilites, homeless populations, military barracks and other similar populations. CA-MRSA has been spreading with increasing intensity across the globe, evolving and becoming more difficult to treat with the passing of time. (StrauÃ et al. 2017)

In an attempt to better understand this ever relevant pathogen, we took a deep dive into mathematically modeling its spread as well as its evolution. To this end, we broke our work into two midterms: Midterm 1 (Section 2) where we use a compartment model and $R_0$ analysis to reproduce results from Kajita et al. (2007), and Midterm 2 (Section 3) where we followed the evolution of Sequence Type 8 (ST8) *S. aureus* on its evolutionary journey from a Methicillin-susceptible ancestor (MSSA) in Europe, to the hyper virulent MRSA strain, USA300, which is now spreading globally. For this second study we attempt to reproduce results from StrauÃ et al. (2017). In this paper, I will summarize our findings from the first two studies, as well as introduce a final study (Section 4) on the mecI gene of *S. aureus*, which plays a significant role in the bacteria's resistance to Methicillin. (Lewis and Dyke 2000)

# 2 Midterm 1: Compartment Model Analysis

## 2.1 Introduction: Midterm 1

For midterm 1 (Adams, Campbell, and Kurkut 2022a), we followed a study called *Modelling an outbreak of an emerging pathogen* by Kajita et al. (2007). This gave us a chance to examine one of those MRSA hotbeds mentioned in Section 1, a jail. Not just any jail, but the largest jail in the world, the LA County Jail. (Kajita et al. 2007) In the course of this study, Ryan evaluated the approximate solution curves of the differential equations shown in Figure 2.1, using the Runge-Kutta Method. With some collaboration from Ryan and I, Aditya implemented the ODE45 Integration Technique. I derived the reproduction number $R_0$, using the methods published by van den Driessche and Watmough (2002) and then simulated 1,000 sample values for $R_0$ based on pseudo-randomly varying parameter values.

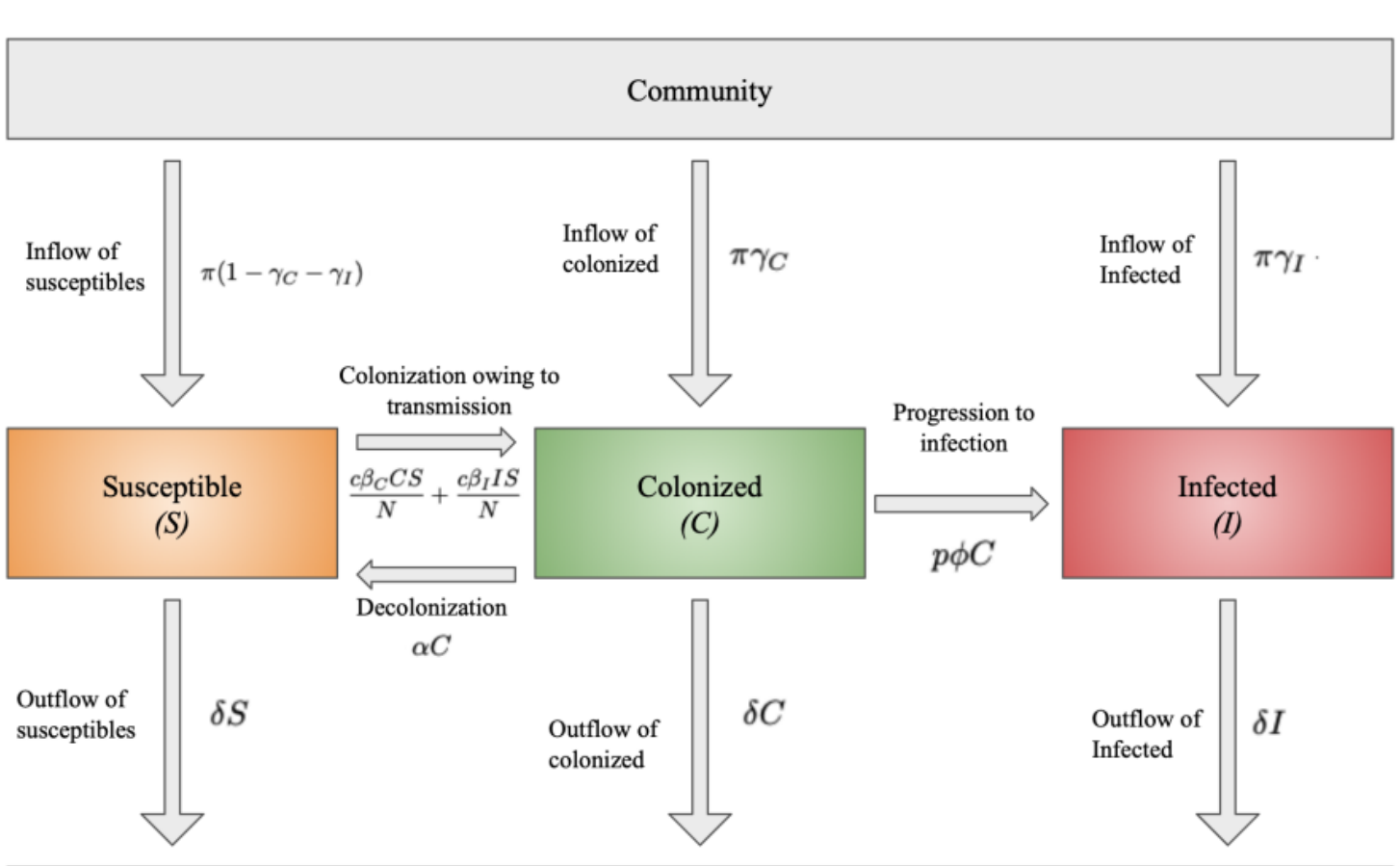

Figure 2.1: CA-MRSA: compartment model (image by Ryan Campbell, model by Kajita, et al)

$$\frac{dS}{dt} = \pi(1-\gamma_C-\gamma_I)+\alpha C-\frac{c\beta_C CS}{N}-\frac{c\beta_I IS}{N}-\delta S \tag{1}$$

$$\frac{dC}{dt} = \pi\gamma_C+\frac{c\beta_C CS}{N}+\frac{c\beta_I IS}{N}-\alpha C-p\phi C-\delta C \tag{2}$$

$$\frac{dI}{dt} = \pi\gamma_I+p\phi C-\delta I \tag{3}$$

Figure 2.2: Ordinary Differential Equations for the Model by Kajita, et al

| Notation | Parameter |
|---|---|
| N | Total number of inmates |
| $\pi$ | Number of individuals booked per day |
| $1/\delta$ | Average incarceration time (days) |
| $\gamma_C$ | Probability that an inmate enters the jail colonized with CA-MRSA |
| $\gamma_I$ | Probability that an inmate enters the jail infected with CA-MRSA |
| $1/\alpha$ | Average decolonization time (days) |
| $p$ | Proportion of colonized individuals who progress to infected |
| $1/\phi$ | Average time for colonized individuals to progress to infection (days) |
| $\beta_C$ | Probability that a non-carrier individual would become colonized with CA-MRSA upon contact with a colonized individual |
| $\beta_I$ | Probability that a non-carrier individual would become colonized with CA-MRSA upon contact with an infected individual |
| c | Average number of contacts per day |

Figure 2.3: Parameters (image by Riley Adams model by Kajita, et al)

## 2.2 Results: Midterm 1

$R_0$ **Analysis:** I used the methods developed by van den Driessche and Watmough (2002) to thoroughly derive, and successfully reproduce the equation for the basic reproduction number of this model as found by Kajita et al. (2007) and shown in Equation (2.1). I also explored an alternate method for deriving $R_0$ which lends itself to a convenient biological interpretation. I looked at the three possible ways individuals in the jail population could exit the Colonized compartment of the model, and then multiplied each by how many inmates will take this route. In this sense, each route out of the Colonized class had its own $R_0$ value, and the overall reproduction number was the weighted average of the three. As in Equation (2.2)

$$R_0 = \frac{c\beta_C + (\rho\phi c\beta_I)/S}{\alpha+\rho\phi+\delta} \tag{2.1}$$

$$\begin{aligned} R_0 &= q_1R_0^1+q_2R_0^2+q_3R_0^3 \\ &= \left(\frac{\alpha}{\alpha+p\phi+\delta}\right)\left(\frac{c\beta_C}{\alpha+p\phi+\delta}\right)+\left(\frac{\delta}{\alpha+p\phi+\delta}\right)\left(\frac{c\beta_C}{\alpha+p\phi+\delta}\right)+\left(\frac{p\phi}{\alpha+p\phi+\delta}\right)\left(\frac{c\beta_C}{\alpha+p\phi+\delta}+\frac{c\beta_I}{\delta}\right) \end{aligned} \tag{2.2}$$

I wrote code in R to collect $2,000$ pseudo-random samples ($1,000$ for male, $1,000$ for female) of values for each parameter in the $R_0$ equation. I used the ranges of values estimated by Kajita et al. (2007), as depicted in Figure 2.4. Then, I utilized Equation (2.1) to calculate $1000$ sample $R_0$ and conduct an analysis on the results. See Figure 2.5. We discovered the mean $R_0$ for males was $3.27$, while the mean $R_0$ for females was $0.71$. The results for the females were very close to those achieved by Kajita et al. (2007), while our mean for males was fairly higher.

| Parameter | Males | Females |
|---|---|---|
| $N$ | 16,956 | 2,200 |
| $\pi$ | 341 - 407 | 64 - 81 |
| $1/\delta$ | 42 - 50 | 27 - 34 |
| $\gamma_C$ | $8.8\text{x}10^{-5} - 4.923\text{x}10{-}3$ | $4.43\text{x}10^{-4} - 7.77\text{x}10^{-3}$ |
| $\gamma_I$ | $8.8\text{x}10^{-5} - 4.923\text{x}10{-}3$ | $4.43\text{x}10^{-4} - 7.77\text{x}10^{-3}$ |
| $1/\alpha$ | 30 - 120 | 30 - 120 |
| $p$ | 0.10 - 0.30 | 0.10 - 0.30 |
| $1/\phi$ | 4 - 15 | 4 - 15 |
| $\beta_C$ | $1\text{x}10^{-5} - 1.5\text{x}10{-}3$ | $1\text{x}10^{-5} - 2\text{x}10^{-3}$ |
| $\beta_I$ | $1\text{x}10^{-5} - 1.5\text{x}10{-}3$ | $1\text{x}10^{-5} - 2\text{x}10^{-3}$ |
| $c$ | 5 - 50 | 5 - 50 |

Figure 2.4: Estimated parameter Ranges (image by Adams and Kurkut, estimates by Kajita, et al)

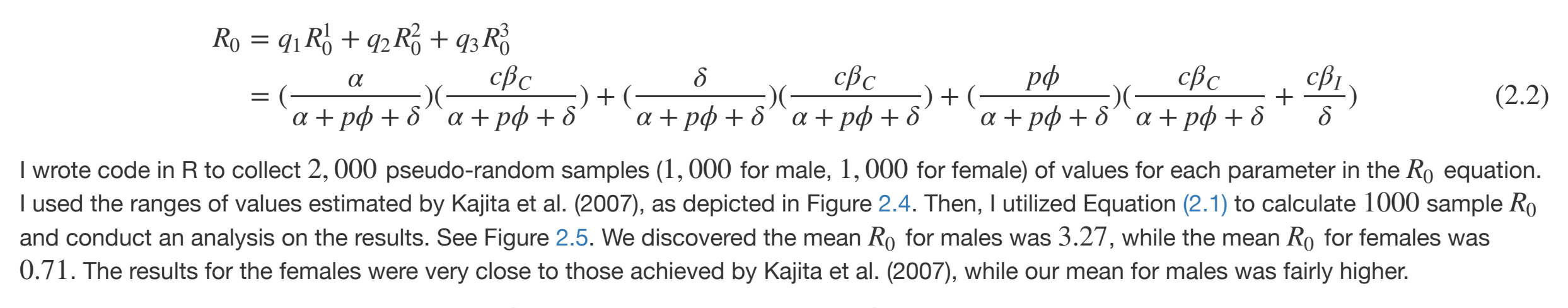

Figure 2.5: Histograms for 1000 Randomly Generated R0 by Varying Parameter Values. Male Population (Left), Female Population (Right)

**ODE45:** Aditya Kurkut implemented the ODE45 function in the MATLAB software to integrate the differential equations in the model. To do so he input the Differential Equations, a range of time and some initial values which we deemed reasonable estimations. The results are depicted in Figure 2.6.

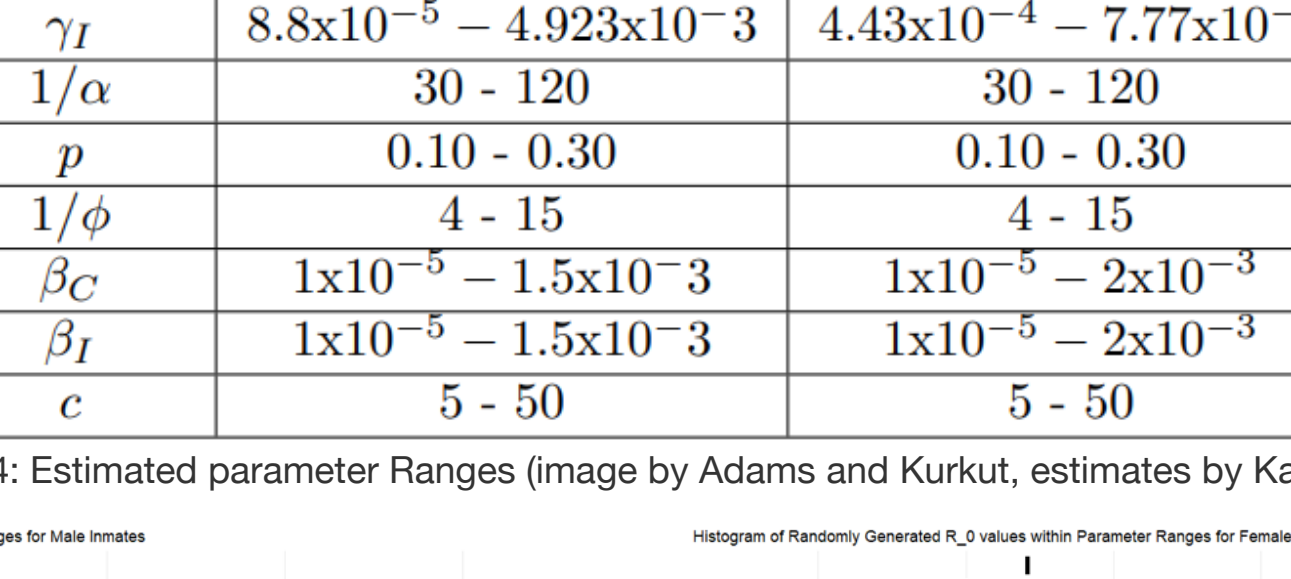

Figure 2.6: Solution Curve for the ODE's in the Model (image by Aditya Kurkut)

**Runge-Kutta:** Ryan Campbell used the Runge-Kutta method to approximate a solution curve for each differential equation. In doing so, a recursive, step-wise process was applied to some reasonable initial conditions for the model. The iterations of this process were made possible by the code Campbell produced in the MATLAB software using a "for loop" over 8 iterations, to represent the time steps at each month of the 9 month long study. The process was carried out seperately for male and female inmates. The results are depicted in Figure 2.7 and Figure 2.8.

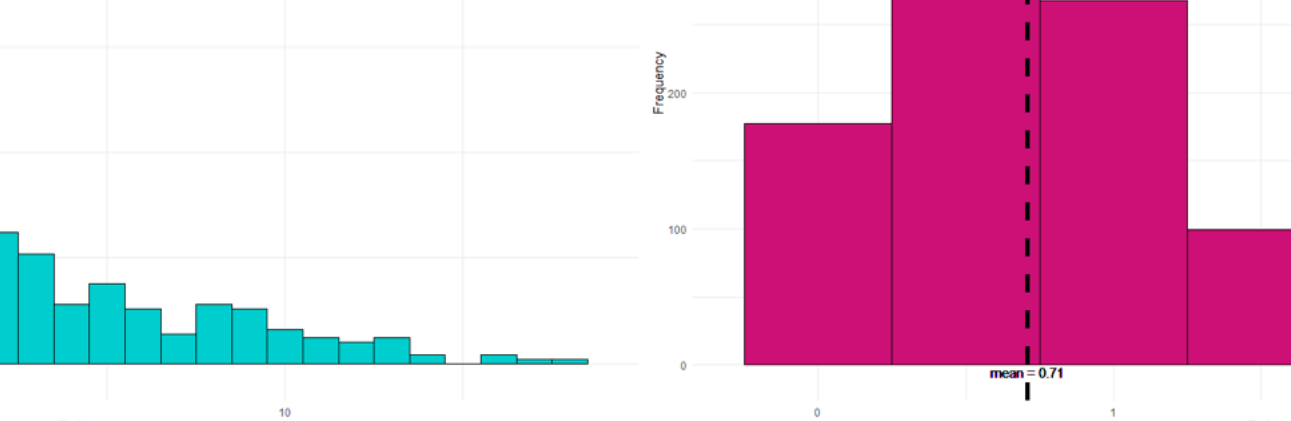

Figure 2.7: Runge-Kutta Numerical Method for Approximating Solution Curves | Male Inmates (image by Ryan Campbell)

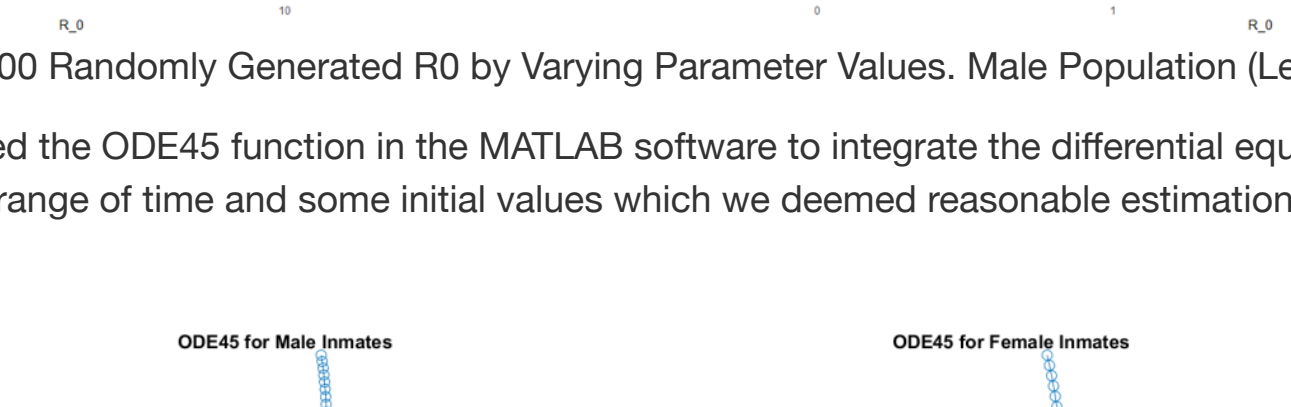

Figure 2.8: Runge-Kutta Numerical Method for Approximating Solution Curves | Female Inmates (image by Ryan Campbell)

## 2.3 Discussion: Midterm 1

The differential equations model and the numerical methods we used to analyze it in Midterm 1 helped us understand how CA-MRSA spreads in susceptible populations. Our $R_0$ analysis indicated how varying values of the parameters in the model can effect the MRSA's ability to reproduce and spread. These parameters could be closely examined to determine which ones we can affect to curb the spread, and inform policy at institutions like correctional facilities. We saw in Ryan Campbell's Runge-Kutta analysis that differing initial conditions can lead to drastically different outcomes, thus further scrutiny of this model and the behavior of MRSA can inform onboarding processes to correctional facilities to minimize introduction of this pathogen into the system. In the same vain, the jail is not a closed system. The average term is short and individuals are released back to the community in one of the 3 compartments of the model. Meaning jails can be a source population for MRSA into the larger population. This study can highlight for the public, an awareness that it is important to consider the conditions (crowding, hygeine, etc.) and standards of living for those in our society who are imprisoned because it has an effect on the health of the greater society.

# 3 Midterm 2: Evolutionary Analysis

## 3.1 Introduction: Midterm 2

For Midterm 2 (Adams, Campbell, and Kurkut 2022b), we examined the evolution of ST8 *S. aureus*. We implemented Multiple Sequence Alignment of the DNA, Phylogenetic Tree Analysis and Topological Data Analysis. In doing so, we attempted (with some success) to recreate results from a study titled "Origin, evolution, and global transmission of community-acquired Staphylococcus aureus ST8" by StrauÃ et al. (2017). By following the evolution of ST8 *S. aureus* we can track its origins in Denmark and see how it branched out and spread through massive immigration events and World Wars into the United States where it circulated and underwent mutations which gave the bacteria its resistance to Methicillin. From the United States, this new highly virulent strain has been exported globally. It has even branched into two distinct main types known as USA300-NAE (North American Epidemic) and USA300-SAE (South American Epidemic), which are now circulating heavily as far as Africa, and Australia. (StrauÃ et al. 2017)

## 3.2 Results: Midterm 2

To analyze the phylogeny of ST8 *S. aureus*, we used the data studied by StrauÃ et al. (2017), which contained 224 samples from the DNA of different ST8 *S. aureus* strains. Since StrauÃ et al. (2017) had already aligned the sequences against the chromosome of the *S. aureus* TCH1516 ST8 reference genome, using the Burrows-Wheeler Aligner, we did not get a chance to actually perform the sequence alignment ourselves. Although, I ran the `AlignSeq()` function from the R software package "DECIPHER" anyways, because it was a necessary step in our code. This did not have any effect of the positioning of our already aligned DNA sequences. Using The R software, guidance from the github repository authored by RussellGrayxd (2020) and various packages which can be found in the appendix of Adams, Campbell, and Kurkut (2022b), I first computed a distance matrix of the pairwise distance between each of the 224 sequences as visualized in Figure 3.1

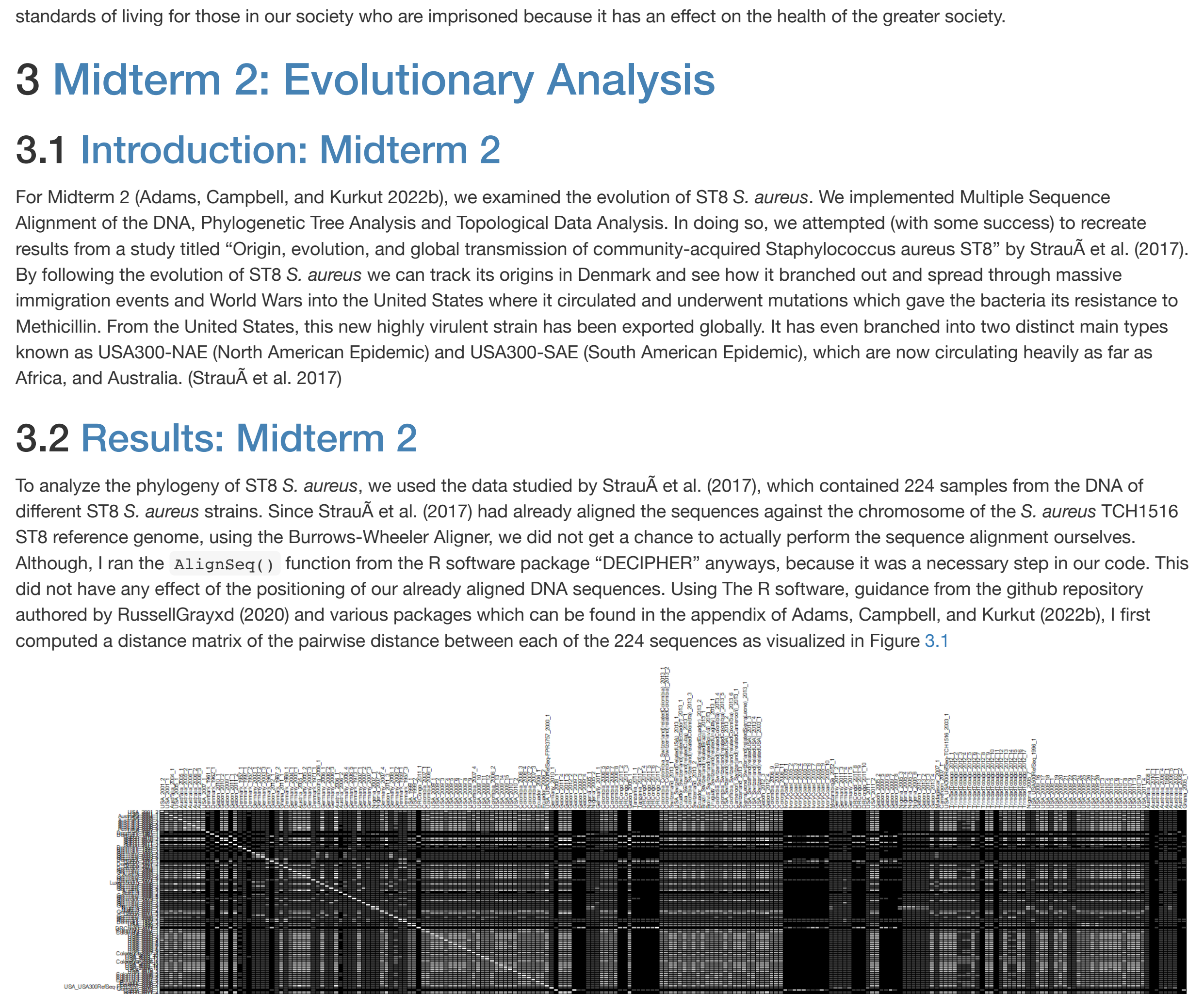

Figure 3.1: Distance Matrix: darker grey is more distant

I then implemented the tree estimation method created by Saitou and Nei (1987) known as "Neighbor-Joining". This was performed using the `ape::nj()` function in R, but the *neighbor-joining matrix* $D^*$ is defined as the matrix which, given an $n \times n$ distance matrix $D$, is

$$D^*_{i,j} = (n-2)\cdot D_{i,j} - TotalDistance_D(i) - TotalDistance_D(j)$$

, where $TotalDistance_D(i)$ is the sum of distances from $i$ to all other leaves.

This matrix $D^*$ was then run through various plotting functions to produces the horizontal phylogram in Figure 3.2 and circular phylogram in Figure 3.3. The phylograms were not an exact match to those produced by StrauÃ et al. (2017), however, they were quite similar. Both phylogenies had the highly virulent USA300 strain peppered throughout. Further, Denmark strains remaine near the root. Two distinct African clades were apparent, and one had close relation between Gabon and Nigeria strains with USA300 as a recent ancestor. We also had distinctly North American and South American clusters, depicting the 2 main USA300 epidemic types (NAE and SAE).

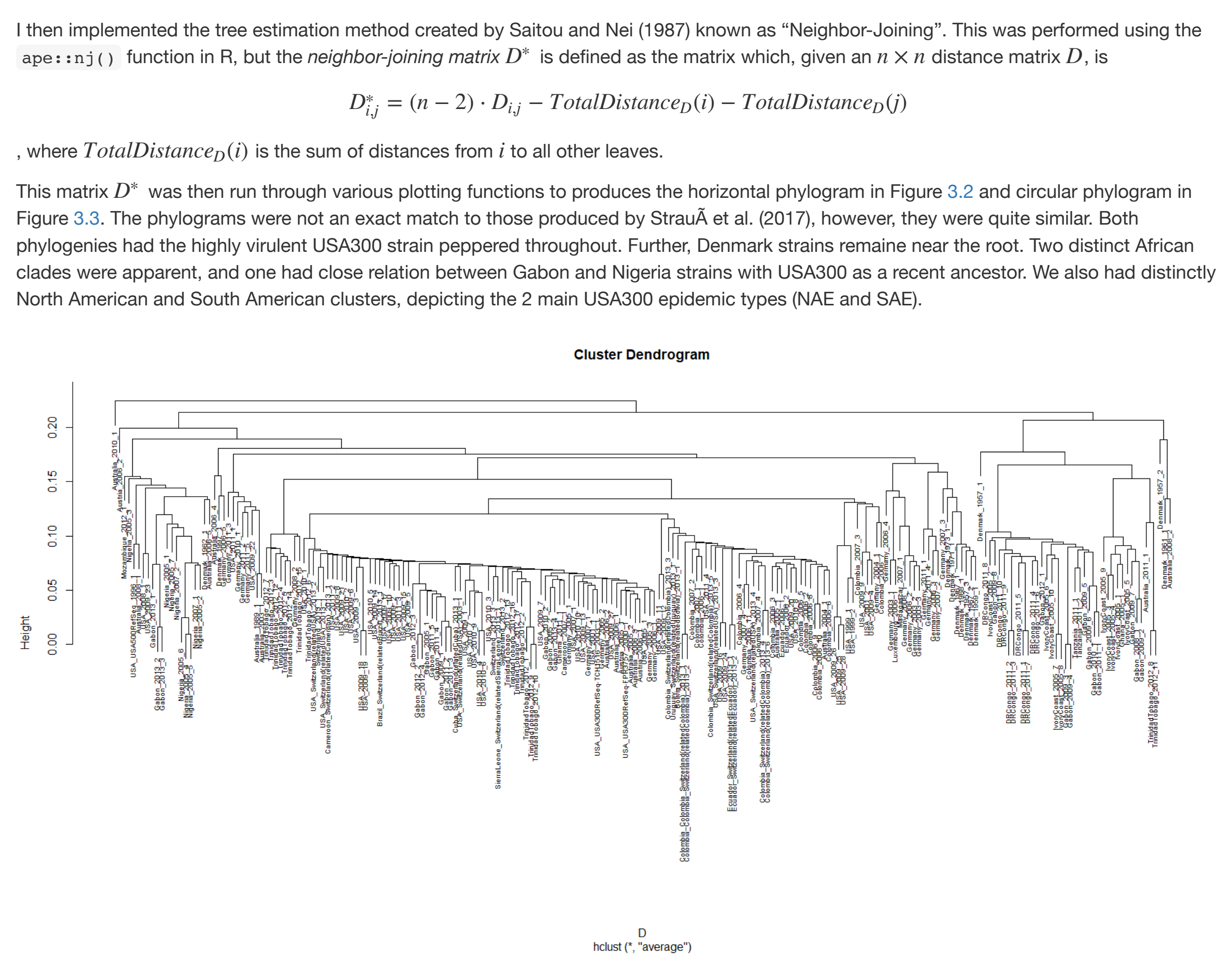

Figure 3.2: Phylogram

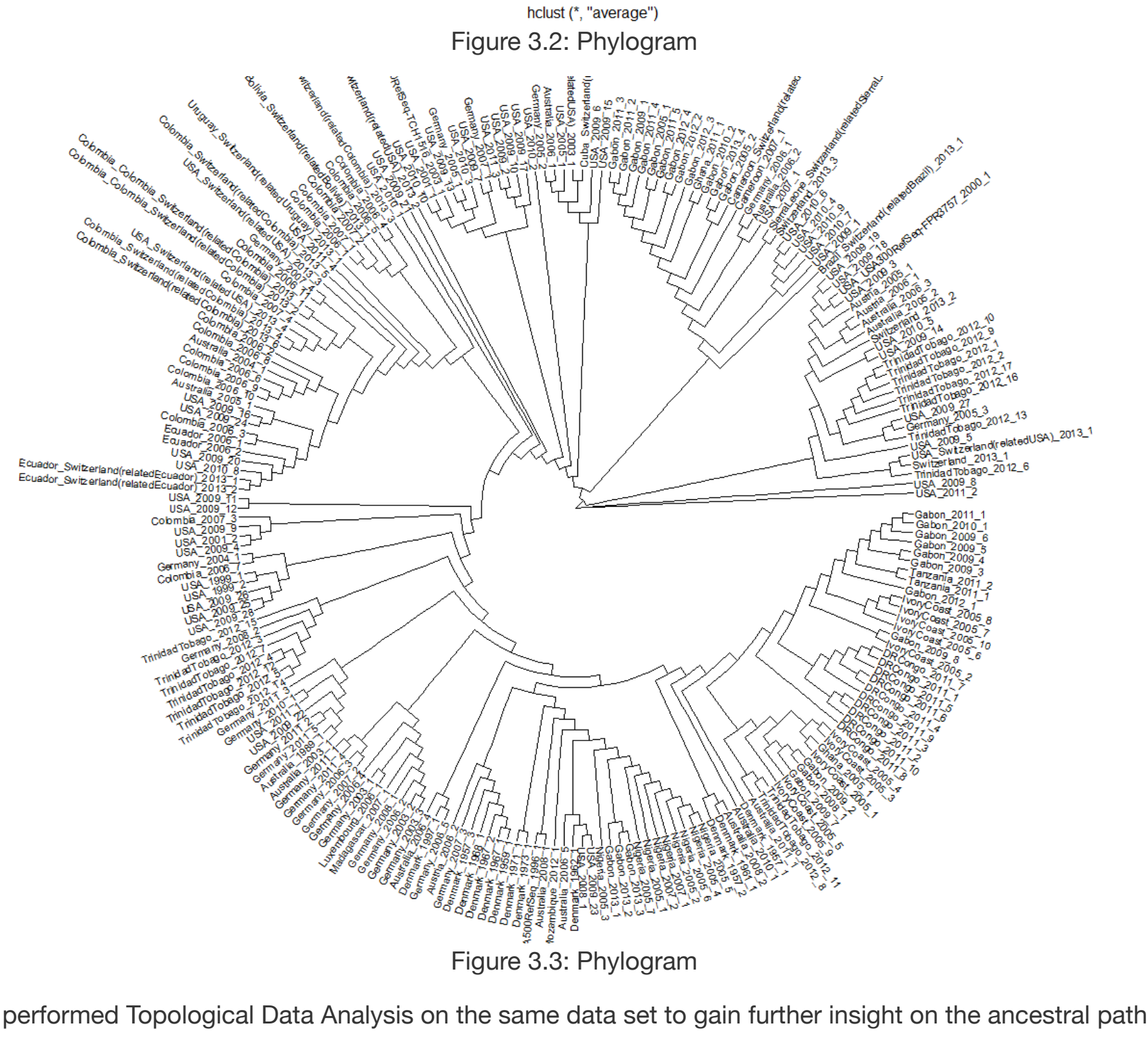

Figure 3.3: Phylogram

Next, Ryan Campbell performed Topological Data Analysis on the same data set to gain further insight on the ancestral paths of ST8. This analysis was also performed using the R software and the same distance matrix visualized in Figure 3.1, this time using the "TDAstats" package with `calculateHomology()` function. Ryan set the dimensions to $1$, threshold to $-1$, format to 'distmat' and standardization to 'FALSE', which gave us the calculated homologies of the 224 strains of ST8. From there, Campbell produced the plots in Figure 3.4. The left plot is the barcode plot which is varified by the same information in the persistence plot on the right.

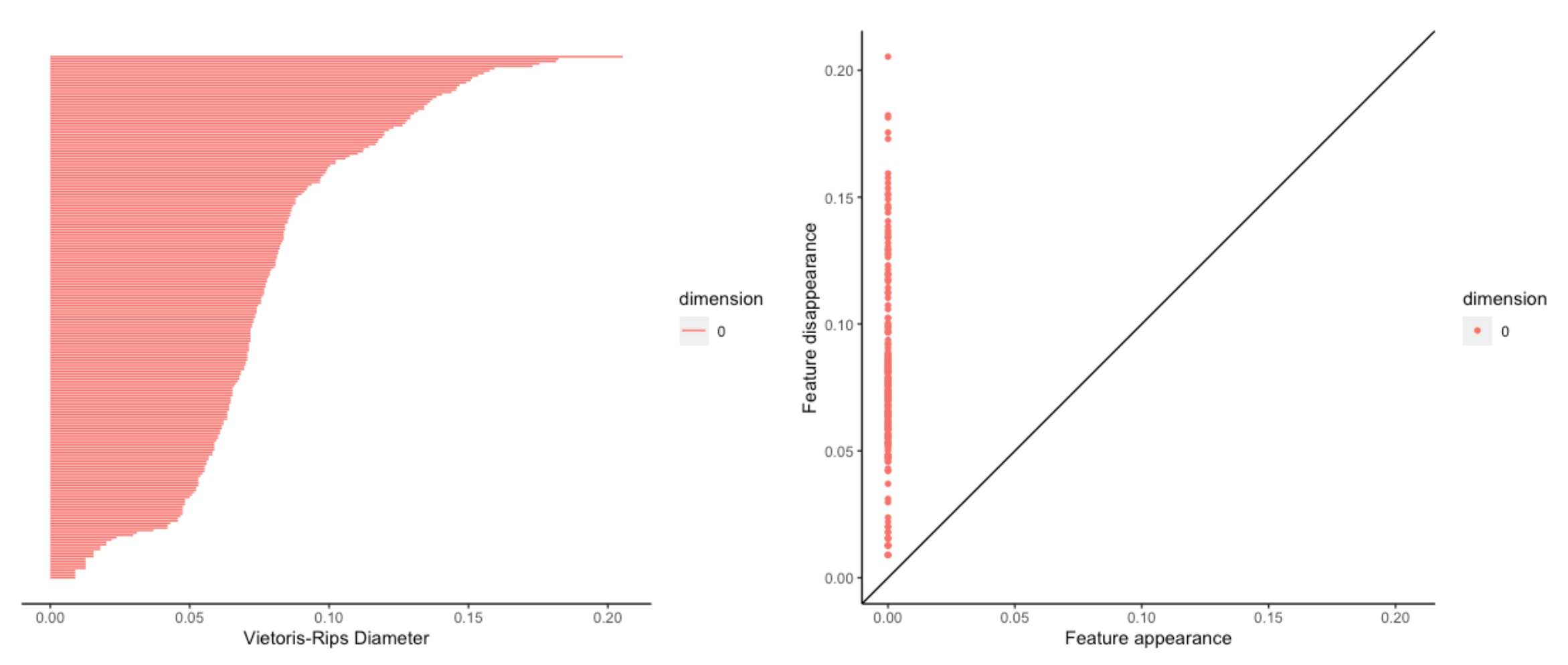

Figure 3.4: Barcode Plot (left) and persistance plot (right)

## 3.3 Discussion: Midterm 2

The analysis performed on the phylogeny of ST8 gave us insight into the history of this bacteria and how highly virulent CA-MRSA has risen to global prominence. Our phylogenetic tree illustrates how USA300 has given rise to many more antibiotic resistant strains of ST8 MRSA which are now circulating on other continents. The topological data further illustrates an acceleration in the mutation of CA-MRSA strains, particularly around the center of the graph where the "slope" is especially steep. As resistance to antibiotics becomes more and more prevelant, we will have to find new ways to treat infections caused by bacteria. Scrutiny should be placed on the allocation of antibiotics to avoid over using them and creating stronger resistance. Further, more research can be conducted into the mechanisms of the bacteria which cause its resistance to antibiotics. (We will examin one such mechanism in Section 4).

In an exciting new frontier, Nick et al. (2022) very recently published a study in which mycobacteriaphages we employed to successfully treat an infection in the lung of a male patient. These phages (essentially viruses for bacteria) were engineered in such a way as to more effectively lyse the *M. abcessus* causing the infection. Although this type of treatment has had varying success, with bacteria also developing restance to phages, (Nick et al. 2022) further research in this area could provide us with more tools to fight this epidemic of MRSA as the problem grows more severe.

# 4 Final: 3D Protein Structure

For the final portion of this MRSA study I will obtain the 3D structure for the Methicillin Resistance Regulatory protein MecI encoded for by the mecI gene, and discuss the function and relevance of this protein to Methicillin Resistance in *S. aureus*. To obtain the structure of the protein, I accessed the UniProt database and downloaded the FASTA file which codes for the amino acid sequence translated from the mecI gene. I then used the AlphaFold2 software tool developed by Mirdita et al. (2022) to simulate the 3d structure and produce Figures 4.1, 4.2 and 4.3. These show a computer generated model of what the protein looks like and the confidence of the predictions made by the program. Cooler colors in the model indicate higher confidence in the model, where warmer colors are lower confidence, measured in per-residue confidence score (pLDDT), this information is diplayed in plots in Figure 4.3.

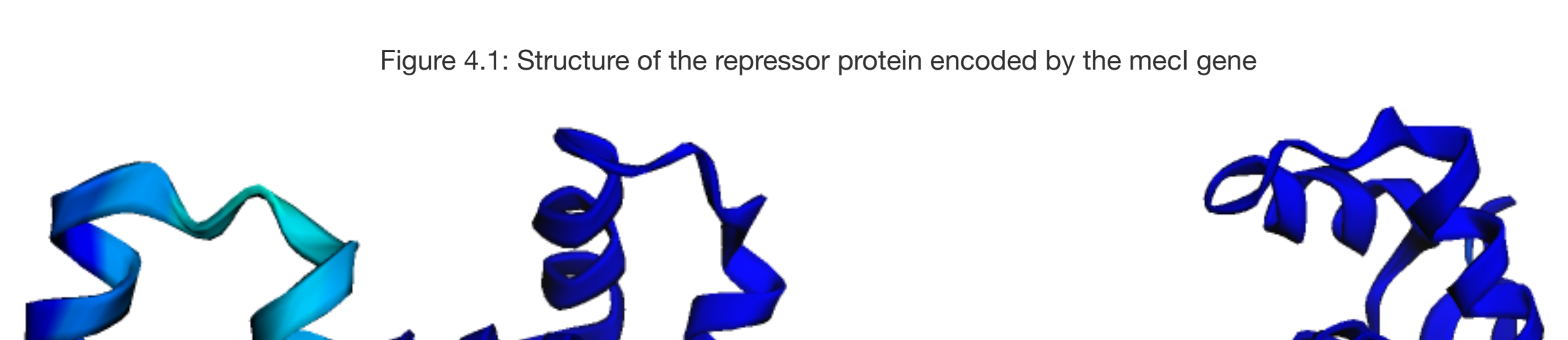

Figure 4.1: Structure of the repressor protein encoded by the mecI gene

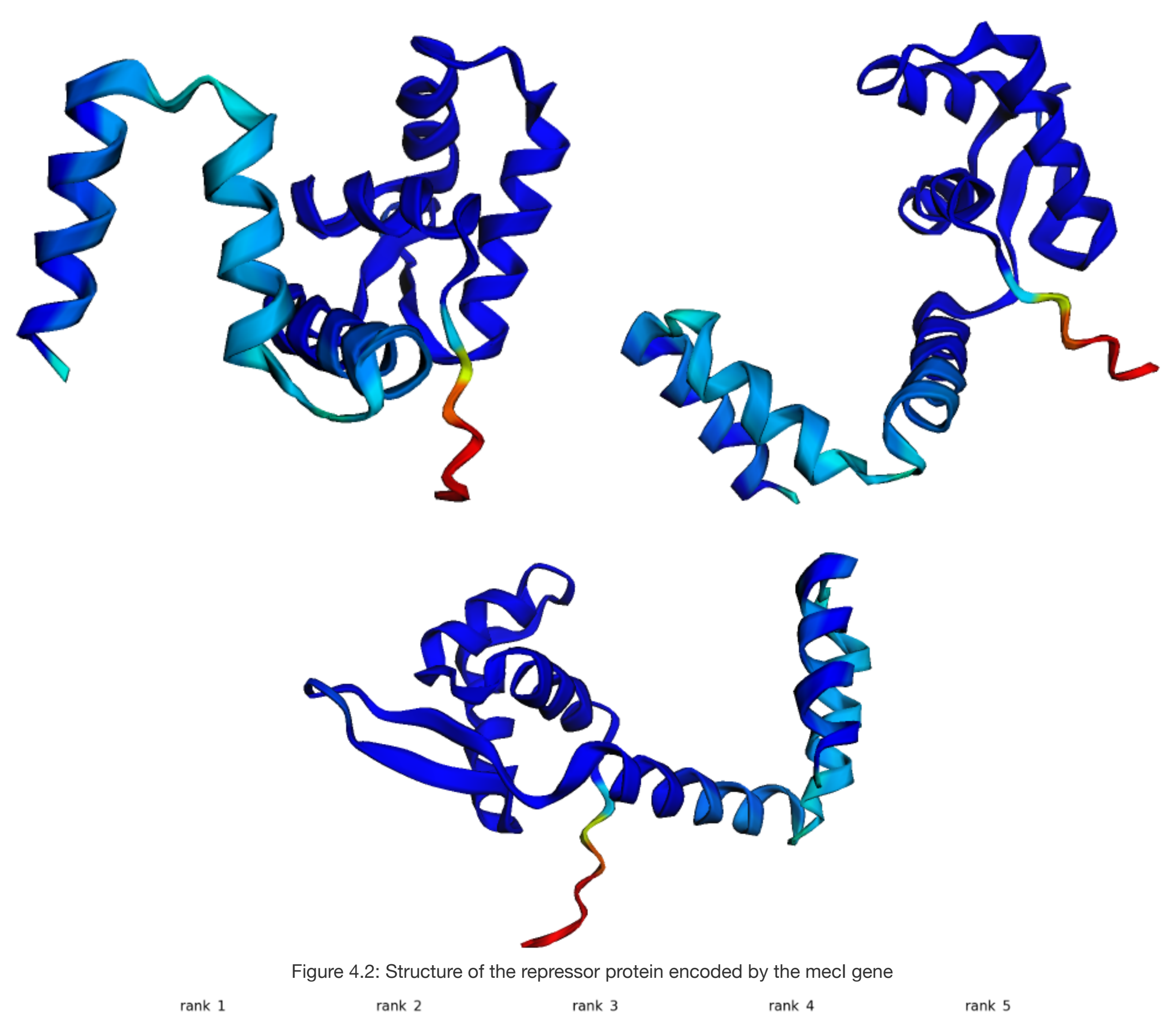

Figure 4.2: Structure of the repressor protein encoded by the mecI gene

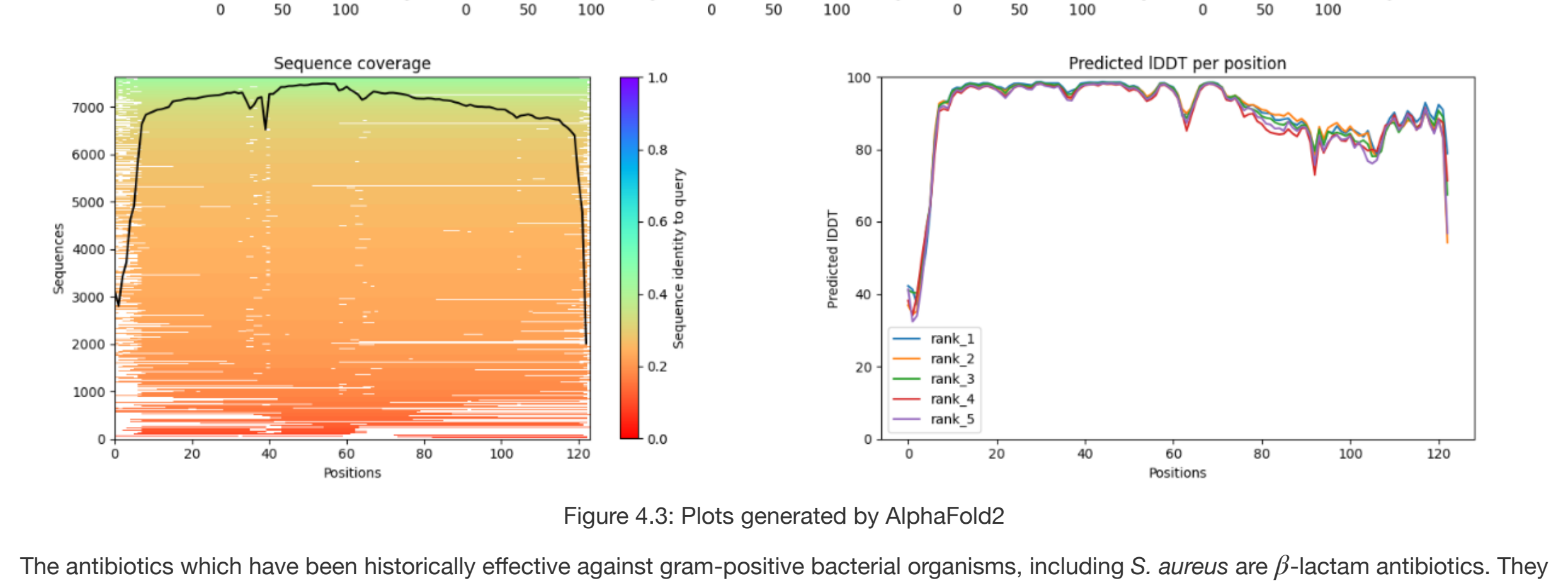

Figure 4.3: Plots generated by AlphaFold2

The antibiotics which have been historically effective against gram-positive bacterial organisms, including *S. aureus* are $\beta$-lactam antibiotics. They are called $\beta$-lactam antibiotics because they contain a $\beta$-lactam ring in their chemical structure. (Holten and Onusko 2000) They are effective at treating against these bacterial infections by "interfering with the structural crosslinking of peptidoglycans in bacterial cell walls." (Holten and Onusko 2000) However, MRSA strains have evolved to deal with this in a number of ways, the main way of resisting $\beta$-lactams in *S. aureus* is accomplished by the production of the enzyme $\beta$-lactamase. (Lewis and Dyke 2000) However, the mechanism which relates to the mecI gene and methicilin resistance, is production of a protein known as PBP2a (penicillin-binding protein 2a).

This protein, PBP2a, is encoded for by the mecA gene. MecA is part of the mobile genetic element known as Staphylococcal Chromosomal Cassette *mec* (SCCmec). (Shalaby et al. 2020) Two other important genes are a part of this element: mecR1 and the one we are examining, mecI. mecI is a regulatory gene which represses expression of the mecA gene. In turn, mecR1 represses expression of the mecI gene, allowing production of the PBP2a protein which results in antibiotic resistance. (Shalaby et al. 2020)

Understanding the mecI gene is an important step to revealing the mechanisms which render our medicines ineffective against this pathogen. All MRSA type *S. aureus* either has no mecI gene to regulate expression of the mecA gene, or has mutations in place which prevent proper transcription of mecI. (Weller 1999) We know that mecI is a homodimer, which means it is a protein consisting of two identical polypeptide chains. Each monomer of the protein "consists of a compact N-terminal winged-helix domain" (Safo et al. 2006). We see this pictured in Figure 4.2. This is what is used to bind to the DNA (to repress transcription of mecA). Each monomer also contains "a loosely packed C-terminal helical domain, which intertwines with its counter-monomer." (Safo et al. 2006) MecI is heavily related to the BlaI gene, which is the main regulator in synthesis of $\beta$-lactamase. Lewis and Dyke (2000) indicate that it is probable that both of these genes are controlled in a coordinated manner. Discovering more about this control structure may provide us tools to combat MRSA's resistance to our treatments and ultimately prevent this disease from continuing on its trajectory to becoming our next pandemic.

# References

Adams, Riley, Ryan Campbell, and Aditya Kurkut. 2022a. "Modeling 2002 Outbreak of CA-MRSA in LA County Jail," April.

———. 2022b. "MAT 124 - MRSA Project: Midterm 2," June.

Holten, K, and E Onusko. 2000. "Appropriate Prescribing of Oral Beta-Lactam Antibiotics." *American Family Physician* 62 (September): 611–20.

Kajita, Emily, Justin T. Okano, Erin N. Bodine, Scott P. Layne, and Sally Blower. 2007. "Modelling an Outbreak of an Emerging Pathogen." *Nature Reviews Microbiology* 5 (9): 700–709. https://doi.org/10.1038/nrmicro1660.

Lewis, Richard A., and Keith G. H. Dyke. 2000. "MecI Represses Synthesis from the ã-Lactamase Operon of Staphylococcus Aureus." *Journal of Antimicrobial Chemotherapy* 45 (2): 139–44. https://doi.org/10.1093/jac/45.2.139.

Mirdita, Milot, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. 2022. "ColabFold: Making Protein Folding Accessible to All." *Nature Methods*, May. https://doi.org/10.1038/s41592-022-01488-1.

Nick, Jerry A., Rebekah M. Dedrick, Alice L. Gray, Eszter K. Vladar, Bailey E. Smith, Krista G. Freeman, Kenneth C. Malcolm, et al. 2022. "Host and Pathogen Response to Bacteriophage Engineered Against Mycobacterium Abscessus Lung Infection." *Cell* 185 (11): 1860–1874.e12. https://doi.org/https://doi.org/10.1016/j.cell.2022.04.024.

Pantosti, A., and M. Venditti. 2009. "What Is MRSA?" *European Respiratory Journal* 34 (5): 1190–96. https://doi.org/10.1183/09031936.00007709.

RussellGrayxd. 2020. "Phylogenetics." *GitHub Repository*. https://github.com/RussellGrayxd/Phylogenetics; GitHub.

Safo, Martin K., Tzu-Ping Ko, Faik N. Musayev, Qixun Zhao, Andrew H.-J. Wang, and Gordon L. Archer. 2006. "Structure of the MecI Repressor from Staphylococcus Aureus in Complex with the Cognate DNA Operator of Mec." *Acta Crystallographica Section F Structural Biology and Crystallization Communications* 62 (4): 320–24. https://doi.org/10.1107/s1744309106009742.

Saitou, Naruya, and Masatoshi Nei. 1987. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Molecular Biology and Evolution*, July. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Shalaby, Menna-Allah W., Eman M. E. Dokla, Rabah.A. T. Serya, and Khaled A. M. Abouzid. 2020. "Penicillin Binding Protein 2a: An Overview and a Medicinal Chemistry Perspective." *European Journal of Medicinal Chemistry* 199: 112312. https://doi.org/https://doi.org/10.1016/j.ejmech.2020.112312.

StrauÃ, Lena, Marc Stegger, Patrick Eberechi Akpaka, Abraham Alabi, Sebastien Breurec, Geoffrey Coombs, Beverly Egyir, et al. 2017. "Origin, Evolution, and Global Transmission of Community-Acquired Staphylococcus Aureus St8." *Proceedings of the National Academy of Sciences* 114 (49): E10596–604. https://doi.org/10.1073/pnas.1702472114.

van den Driessche, P., and James Watmough. 2002. "Reproduction Numbers and Sub-Threshold Endemic Equilibria for Compartmental Models of Disease Transmission." *Mathematical Biosciences* 180 (1): 29–48. https://doi.org/https://doi.org/10.1016/S0025-5564(02)00108-6.

Weller, Timothy M. A. 1999. "The Distribution of mecA, mecR1 and mecI and Sequence Analysis of mecI and the Mec Promoter Region in Staphylococci Expressing Resistance to Methicillin." *Journal of Antimicrobial Chemotherapy* 43 (1): 15–22. https://doi.org/10.1093/jac/43.1.15.