Correlating CRISPR3-Cas with Antibiotic Resistance in *Enterococcus faecalis*

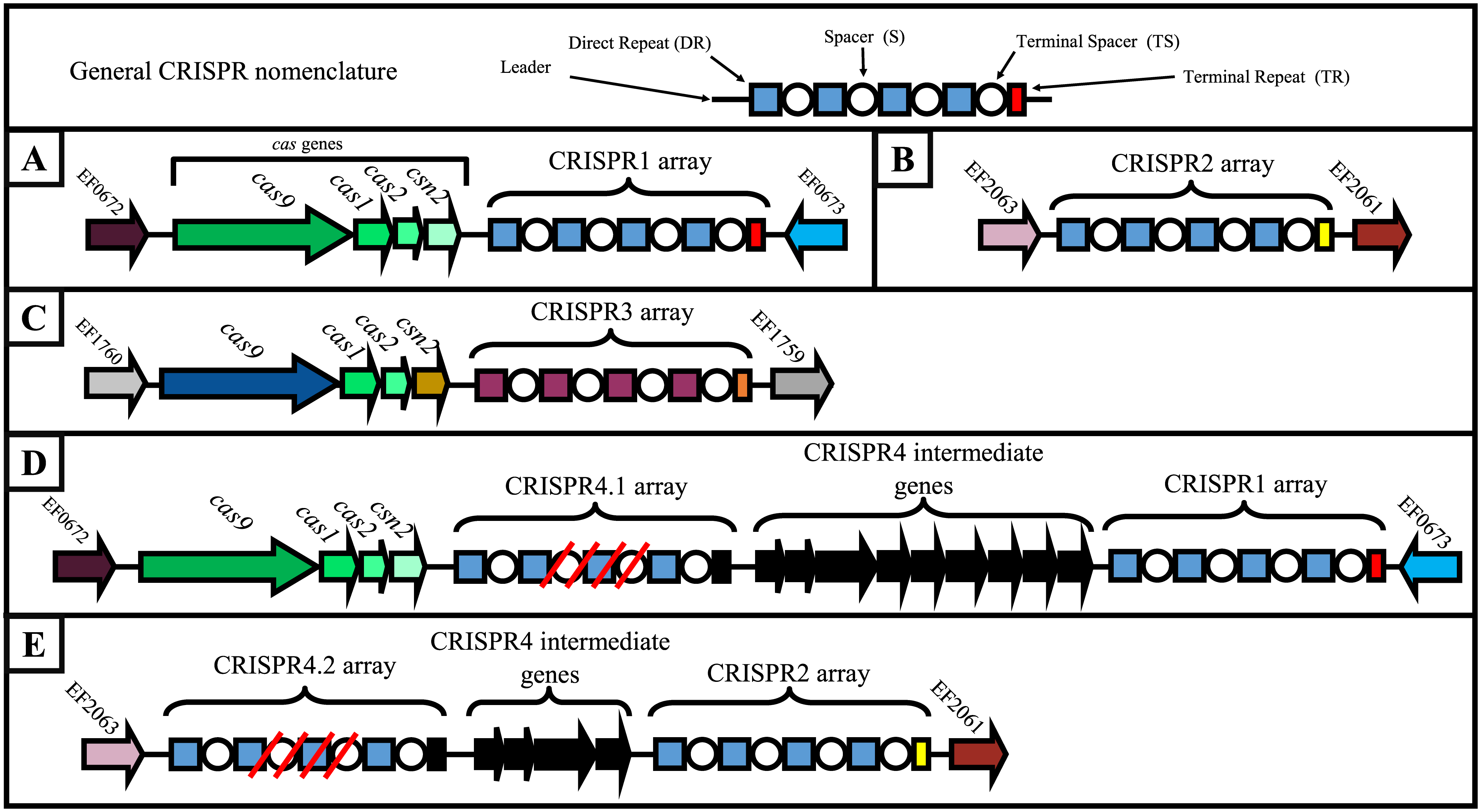
Kevin Mulkerrins

5/29/2019

## Introduction:

Clustered Regularly Interspaced Palindromic Repeats (CRISPR) is an adaptive immune system found in prokaryotes. The system is made up of an array of DNA repeats and spacers of a specific length of base pairs that are homologous to sequences found in viruses, plasmids, and other mobile genetic elements, and CRISPR associated (cas) genes that act to degrade foreign DNA or RNA and to incorporate new spacers for future recognition.[1] Because CRISPR targets foreign DNA, it acts as an impediment to horizontal gene transfer (HGT), including the transfer of genes that may be of selective advantage to the prokaryote. A trade-off exists: complete CRISPR-cas systems have been demonstrated to be negatively correlated with antibiotic resistance in some species.[2,3] However, the presence of CRISPR-cas does not preclude antibiotic resistance.[2] Additionally, there is some evidence that CRISPR-cas temporarily tolerates HGT. CRISPR is also implicated in non-canonical functions.[4] Resistance to glyocopeptides (vancomycin and teicoplanin) is acquired in *Enterococcus faecalis*, and are encoded by the genes *vanA*, *vanB*, *vanE*, and *vanG*.[2]

There are three Type II CRISPR loci in *Enterococcus faecalis*, with a potential fourth locus in a non-conserved location.[5] *E. faecalis* is a common gastrointestinal commensal in both humans and multiple land animals, but it is also ubiquitous in the environment.[2,3] The species has a high degree of intrinsic antibiotic resistance and engages in frequent HGT. Plasmids, pathogenicity islands, and antibiotic resistance genes can make up as much as 25% of the genomes of hospital-adapted strains.[2] As a result, *E. faecalis* has emerged as a major source of nosocomial infection.[2] The three CRISPR loci are found in conserved positions on the *E. faecalis* chromosome (Figure 1), though they don’t always co-occur and all three may be absent entirely.[5] CRISPR2 is the most common and was once considered ubiquitous. CRISPR1 is the next most common, and CRISPR3 is thus far rare.[2] CRISPR1 and CRISPR2 have arrays with identical repeat sequences and are thought to have a common origin, however, CRISPR2 lacks known cas genes. As a result, it is considered an orphan system with no canonical functionality.[2,3,5] Integrating CRISPR1-cas9 into CRISPR2 strains restores interference functionality.[4] CRISPR3 has four Type II cas genes like CRISPR1, however the cas genes seem to be more distantly related, and the two loci have different repeat sequences.[5] Previous research has shown that CRISPR1 strains are negatively correlated with antibiotic resistance compared to CRISPR2 only strains, however, previous unpublished analysis has suggested that CRISPR3 is more prevalent in clinical than non-clinical strains. It is possible that CRISPR3 strains are permissive to HGT and are therefore gain a selective advantage against CRISPR1 strains by allowing resistance to both antibiotics and to phage infection. It is the expectation of this analysis that CRISPR3 strains will be positively correlated with vancomycin resistance in clinical strains.



**Figure 1.** Genomic mapping of CRISPR loci in *E. faecalis*. Taken from Hullahalli K, Rodrigues M, Schmidt BD, Li X, Bhardwaj P, Palmer KL (2015) Comparative Analysis of the Orphan CRISPR2 Locus in 242 *Enterococcus faecalis* Strains. PLoS ONE 10(9): e0138890. <https://doi.org/10.1371/journal.pone.0138890>

## Methods and Results

To determine whether there is any correlation between acquired antibiotic resistance and CRISPR3 in *E. faecalis*, strains will be screened for CRISPR loci and vancomycin resistance genes. CRISPR3 comparisons will be made using both clinical and wild strains. Strains that are CRISPR2 only will be counted as strains with no CRISPR locus, since the locus is considered to have no canonical CRISPR function without cas genes. Additionally, it will be determined if there is differential resistance to vancomycin based on resistance genotype and if co-occurrence of vancomycin resistance genes are additive in their effect.

***Animal fecal samples and hospital isolates***

*E. faecalis* strains were previously isolated from animal feces to provide a natural baseline for both distribution of CRISPR loci as well as acquired antibiotic resistance genes. Droppings from animals not expected to have significantly higher exposure to antibiotics than in the wild were collected from a variety of species and locations. Most samples were collected in Massachusetts, but some were from upstate New York and Connemara, Ireland. Sterile water was added to the fecal samples and the samples were then vortexed. Cotton swabs were then used to streak the liquefied sample onto enterococcosel agar, which is a growth medium selective for enterococci. Colonies were then picked and isolation-streaked onto Brain-Heart Infusion (BHI) agar to obtain pure cultures. Individual colonies were then picked from these plates and cultured in BHI broth. The broth cultures were then processed using a DNA extraction kit (QIAGEN DNeasy UltraClean Microbial). Next, PCR was performed using the 16S ribosomal subunit. This, and all further Sanger sequencing services, was and will be provided by Eton Bioscience, Boston, Massachusetts. 16S sequences were then searched in NCBI’s BLAST to confirm that the strains were *E. faecalis*. After species is confirmed, PCR and sequencing will be performed to characterize CRISPR loci and vancomycin resistance genes. Agarose gel electrophoresis will then be performed on the PCR products to confirm that the strain has a given locus. It is expected that the animal fecal samples will have a low level or complete absence of vancomycin resistance genes. *E. faecalis* strains known to be vancomycin resistant will also be acquired from hospitals in the Greater Boston Area. These strains will then be screened for CRISPR loci and vancomycin resistance genes if they have not already been characterized.

***Kirby-Bauer Antibiotic Resistance Assay***

Kirby-Bauer Antibiotic Resistance Assays are performed by completely streaking a pure culture onto Mueller-Hinton Agar, placing antibiotic diffusion discs onto the agar, and then incubating at optimal growth temperatures. After the bacterial lawn has grown, the strain can be characterized as resistant, intermediately resistant, and susceptible to the specific antibiotics in the diffusion discs based on the diameter around the disc where bacteria were unable to grow. This area of no growth is called the zone of inhibition (ZOI). It is possible for there to be individual colonies within the ZOI, and these are likely representative of a mutation that attenuates the efficacy of the antibiotic rather than an acquired resistance. Because loss of antibiotic effectiveness due to mutation of a gene that controls the targeted phenotype is possible, some strains are expected to have resistance unrelated to antibiotic resistance via HGT. As such, the assay cannot be relied upon to determine why a strain is resistant, rather than to what degree. For this investigation, Kirby-Bauer Assays will be performed to determine patterns of resistance based on the presence or absence of specific vancomycin resistant genes.

***Genomes from public databases***

*E. faecalis* genomes were downloaded from public databases, including NCBI’s GenBank[7], the Broad Institute’s EnteroGenome[8], and the Joint Genome Institute’s Genomes Online Database (GOLD)[9]. Some strain genomes were uploaded to Rapid Annotation using Subsystem Technology (RAST)[10] to obtain reference sequences for CRISPR loci, cas genes, and *van* genes. The sequences for all of these were then saved in a text file. These sequences will be used to BLAST for CRISPR loci and *van* genes in each genome[6]. Since CRISPR1 and CRISPR2 repeats are identical, CRISPR2 flanking genes were used to confirm where on the chromosome the repeats were located. If a strain is shown to have no CRISPR loci, then flanking genes are used to confirm the absence of the locus.

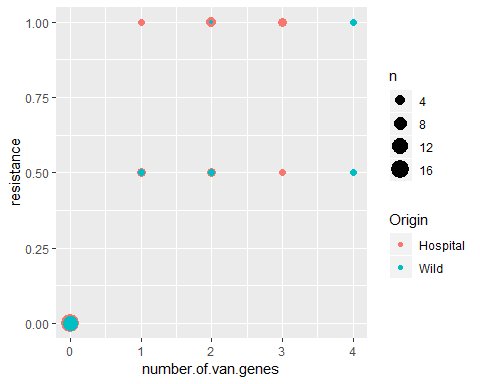
***Statistical analysis in R***

Because strain collection and characterization is not yet complete, the data used for this paper is simulated. The percentage distributions for CRISPR loci based on strain origin were used to assign CRISPR loci to 100 imaginary strains, with 50 being clinical in origin and 50 being wild in origin. Strains with a CRISPR1 locus of any sort were rejected. Strains with only a CRISPR2 locus were counted as having no CRISPR locus (since the CRISPR2 locus is not thought to have canonical function without cas genes). Remaining strains with CRISPR3 loci were counted as being CRISPR3 regardless if there was also a CRISPR2 locus. This resulted in the deletion of 13 clinical strains and 30 non-clinical strains. A gaming D10 was then rolled to assign whether a strain would be susceptible, intermediate, or resistant in a Kirby-Bauer assay. A roll of 4-6 would assign intermediate resistance. A lower range would assign susceptibility. A higher range would assign full resistance. For strains that showed intermediate resistance and resistance, a D4 was rolled to determine how many *van* genes it would have. Finally, the D4 was rolled to determine which *van* genes the strain would have. Strains without *van* genes were assigned to the susceptible category. In reality, resistance may emerge due to mutatation. However this would be controlled for in a real experiment by disregarding strains that showed resistance with an absence of *van* genes.

***Patterns of antibiotic resistance based on number of* van *genes***

The following figure (Figure 2) represents patterns of resistance by number of *van* genes, and color coded by strain origin.

#Data and visualization  
CR\_data <- read.csv("../R/SimulatedData3.csv")  
  
ggplot(CR\_data,  
 aes(x = number.of.van.genes, y = resistance, color = Origin)) +  
 geom\_count()



**Figure 2.** Level of resistance (intermediate = 0.5, full = 1) compared by number of *van* genes.

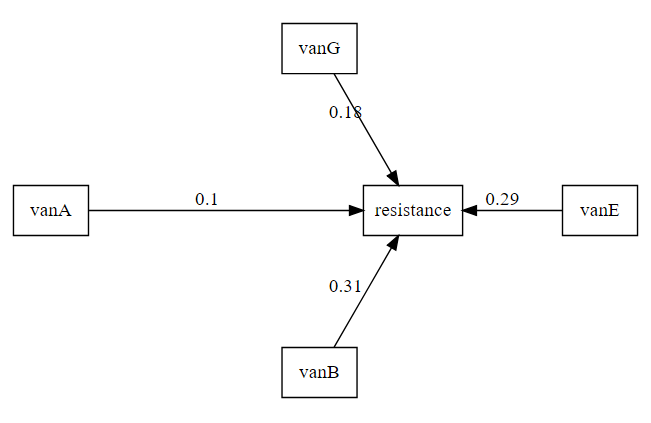
The data show a general trend where antibiotic resistance is far greater in a hospital setting than in wild strains. However, it does not appear that the number of *van* genes is the determining factor in whether or not the strain is phenotypically intermediately or fully resistant.

I attempted to do an ordinal logistic regression for antibiotic resistance using specific *van* genes as ordinals. However, I was unable to get the code to work and was unsuccessful in troubleshooting it. Next, I performed a series of partial mediated modellings to determine whether there were correlations between individual data categories.

***Resistance by genotype***

To determine if any specific *van* genes conferred differential resistance to vancomycin, resistance was compared to all the individual genes using structural equation modeling. (Figure 3)

ResGenoModel <-' resistance ~ vanA + vanB + vanE + vanG'  
ResGenoSEM <- sem(ResGenoModel, data = CR\_data)  
lavaanPlot(model = ResGenoSEM, coefs = TRUE, graph\_options = list(layout ="circo"))

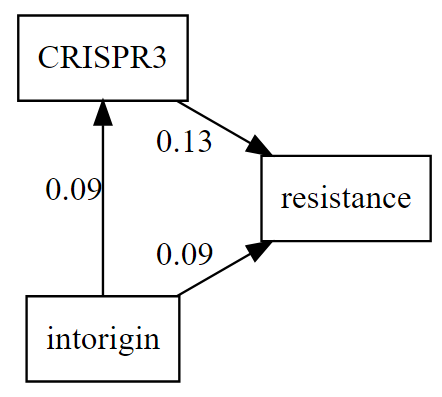


**Figure 3.** Effect of specific *van* gene on level of resistance.

***Correlating CRISPR3 with* van *genes***

Figure 4 tests the the expected model of hospital environment selecting for both CRISPR3 and antibiotic resistance, with CRISPR3 as a mediating factor on acquisition of antibiotic resistance genes.

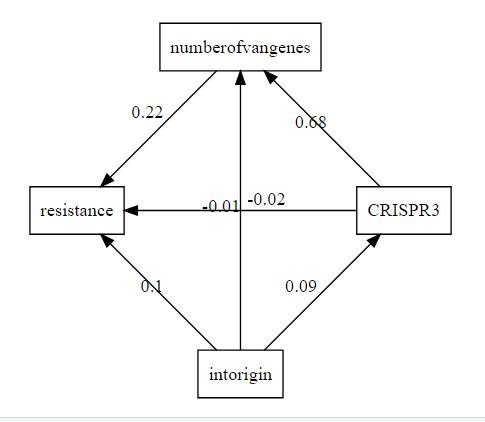
CR3ARModel <-' CRISPR3 ~ int.origin  
resistance ~ CRISPR3 + int.origin'  
CR3ARSEM <- sem(CR3ARModel, data = CR\_data)  
lavaanPlot(model = CR3ARSEM, coefs = TRUE, graph\_options = list(layout ="circo"))



**Figure 4.** Effect of hospital origin on resistance mediated by CRISPR3. Hospital origin in intorigin occur in data as 1. Wild strains occur as 0.

Figure 5 further expands this model by considering whether CRISPR3 has an effect on the number of *van* genes acquired.

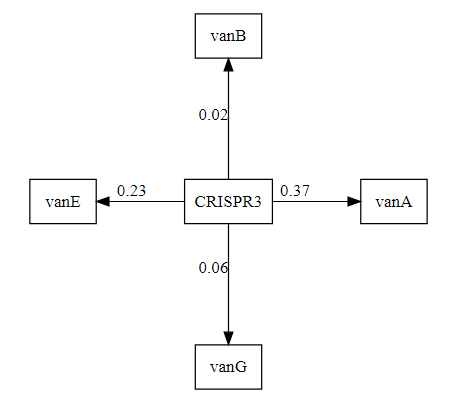
CR3VanModel <-' CRISPR3 ~ int.origin  
number.of.van.genes ~ CRISPR3 + int.origin  
resistance ~ CRISPR3 + int.origin + number.of.van.genes'  
CR3VanSEM <- sem(CR3VanModel, data = CR\_data)  
lavaanPlot(model = CR3VanSEM, coefs = TRUE, graph\_options = list(layout ="circo"))



**Figure 5.** The effect of CRISPR3 on resistance as mediated by number of *van* genes.

Finally, in Figure 6, a mediated model tests whether CRISPR3 preferentially allows or inhibits the acquisition of specific *van* genes.

CRGenoModel <-' vanA ~ CRISPR3  
vanB ~ CRISPR3  
vanE ~ CRISPR3  
vanG ~ CRISPR3'  
CRGenoSEM <- sem(CRGenoModel, data = CR\_data)  
lavaanPlot(model = CRGenoSEM, coefs = TRUE, graph\_options = list(layout ="circo"))



**Figure 6.**  The effect of CRISPR3 on specific *van* gene.

## Discussion

Ordinal logistic regression for antibiotic resistance and logistic regression for probability of CRISPR3 loci were intended but ultimately not carried out due to inability to write good code for it. This was exacerbated by searching online and getting conflicting and confusing examples. A more thorough review of those analyses and working through more examples of how to code for them in R will most likely allow for the inclusion of those analyses when using real data in the future.

The results in Figure 2 suggest generally reflect expectation that hospital strains would have a greater tendency towards antibiotic resistance than wild strains. However, the data also shows an even distribution of resistance based on number of *van* genes. This suggests that resistance is not additive by *van* genes. This may be due to gene expression. Perhaps because the hospital environment has stronger antibiotic selection, expression of *van* genes is upregulated. Another possibility is that certain *van* genes confer greater antibiotic resistance than others. Perhaps it’s even possible that certain combinations of genes attenuate resistance.

The data in Figure 3 further explores whether there is a link between level of resistance and specific *van* genes. Modest resistance seems to correlate with the presence of *vanA* with *vanG* conferring greater resistance, and *vanB* and *vanE* correlating the strongest with increased resistance. However, since the data in Figure 2 suggests that resistance is not additive, the co-occurrence of multiple *van* genes may be a confounding factor. The data in Figure 4 tests the basic model of hospital environment selecting for CRISPR3 and antibiotic resistance, with CRISPR3 as a mediating factor on antibiotic resistance. Hospital environment is equally correlated with CRISPR3 and antibiotic resistance, with slightly greater mediating effect of CRISPR3 on antibiotic resistance. However, the slopes of each of those regressions are gentle, meaning that the correlations are not particularly strong.

This model is expanded to include number of *van* genes in Figure 5. The mediation of CRISPR3 on resistance disappears, but there is a much stronger correlation between CRISPR3 and number of *van* genes. This makes sense, since CRISPR3 itself does not confer antibiotic resistance. The effects of hospital origin remain roughly the same on both CRISPR3 and antibiotic resistance. Interestingly, there is virtual no correlation between hospital origin and number of *van* genes. This may be because the availability and horizontal transfer of *van* genes in a hospital setting are random, but there is an intriguing suggestion here that CRISPR3 somehow facilitates HGT of *van* genes. This opens new avenues of research. Possible correlations between CRISPR3 and specific *van* genes are explored in Figure 6. CRISPR3 has negligible correlations with *vanB* and *vanG* but seems more permissive of *vanE* and even more permissive of *vanA*. Taken altogether, these data suggest that CRISPR3 mediates antibiotic resistance in *E. faecalis* by favoring horizontal gene transfer of *vanE* and *vanA*.

As mentioned, the data used in this paper was simulated and is therefore not necessarily reflective of reality. Particularly, the use of gaming dice in determining resistance and resistance genes may account for the results in figure 2 showing no correlation between number of *van* genes and level of resistance. This may have also contributed to the difficulty in building a prediction model of *van* gene locus with antibiotic resistance. However, it has generated additional relationships to explore when working with experimental data in the future. The possibility suggested in these analyses that CRISPR3 may selectively facilitate antibiotic resistance acquisition is an interesting one, as it would suggest non-canonical function for the locus.

## References

1. Rath, D., Amlinger, L., Rath, A. & Lundgren, M. The CRISPR-Cas immune system: Biology, mechanisms and applications. Biochimie 117, 119–128 (2015).
2. Palmer, Kelli L.; Gilmore, M. S. Multidrug-Resistant Enterococci Lack CRISPR-cas. 1, 1–10 (2010).
3. Lyons, C., Raustad, N., Bustos, M. A. & Shiaris, M. Incidence of type II CRISPR1-Cas systems in Enterococcus is species-dependent. PLoS One 10, 1–11 (2015).
4. Hullahalli, Karthik; Rodrigues, Marinelle; Nguyen, Uyen Thy; Palmer, K. An Attenuated CRISPR-Cas System in Enterococcus faecalis Permits DNA Acquisition. 9, 1–16 (2018).
5. Hullahalli, K. et al. Comparative analysis of the orphan CRISPR2 locus in 242 Enterococcus faecalis strains. PLoS One 10, 1–14 (2015).
6. T., M. The BLAST Sequence Analysis Tool. in The NCBI Handbook [Internet]. (2003).
7. Benson, D. A. et al. GenBank. Nucleic Acids Res. 46, D41–D47 (2018).
8. Enterococcus II initiative, Broad Institute (broadinstitute.org). Available at: <https://olive.broadinstitute.org/projects/enterogenome>.
9. Mukherjee, S. et al. Genomes OnLine database (GOLD) v.7: updates and new features. Nucleic Acids Res. 1–11 (2018). <doi:10.1093/nar/gky977>
10. Brettin, T. et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5, (2015).