BIF701 Lab 6 Gene Prediction: Gene Discovery in a Resistance Plasmid
By Christopher Eeles

Introduction

In this lab we will be exploring the application of the bioinformatics analysis tools NEBcutter and ORFfinder to gene annotation in a highly antibiotic resistant strain of *Enterococcus facieum*.[1] In doing so we will become familiar with the methodology of gene annotation as well as identification of novel or uncharacterized genes for subsequent gene prediction.[1, 2] New England BioLabs' NEBcutter V2.0 tool was designed to identify large, non-overlapping open reading frames (ORFs) from a given gene sequence.[3] It accomplishes this by simulating restriction enzyme digestion with commercially available enzymes which make only a single cut in the provided sequence.[3] These ORFs will then be analyzed via BLASTP to look for homologs in other bacterial strains.[1, 3] Similarly, the NCBIs' ORFfinder tool identifies ORFs which can be analyzed via BLASTP or SMART BLAST for identification of potential protein encoding regions or verification of protein predictions made for uncharacterized DNA segments.[2, 4] Results from both of these tools will be used to identify the potential function of each respective ORF from the *E. facieum* plasmid.

While this strain is a part of the normal intestinal microflora, it has the potential to become an opportunistic infection in individuals with compromised health and/or immunity.[1] As a result of living in the human gut—a diverse and highly competitive microbial ecosystem—*E. facieum* is exposed to myriad microbes, both pathogenic and otherwise, which can serve as a pool of advantageous genes for horizontal gene transfer.[1] Thus, if *E. facieum* is able to enter the human body and establish an infection, there is a significant risk that it already contains numerous antibiotic resistance plasmids which it accumulated over its life cycle in the intestinal microflora.[1] The plasmid studied in this lab was isolated from a patient with a life-threatening post-surgical abdominal infection and serves as a model for the kind of post procedure nosocomial infections common in health care delivery environments such as hospitals and clinics.[1] As antibiotic resistance continues to spread, analyses such as these will become increasing important to develop new and innovative ways to tackle what were less than a century ago simple bacterial infections. Thus, the skills we learn in this laboratory could be essential to future research efforts in this field.

Method

A copy of the FASTA file for the *E. facieum* plasmid was provided by our instructor and used as input into NEBcutter and ORFfinder. The minimum ORF size in both instances was set to 100 AAs (300 NTs) and results were sorted by size, larger being favoured over shorter. This was done to maximize the likelihood that a given ORF would contain an actual protein coding region and not constitute a small non-coding region.[1, 2] Both searches yielded 19 results, but to conserve time only the largest 10 ORFs from each tool were selected for further analysis. These ORFs were then analyzed via BLASTP using the non-redundant protein sequences (nr) database; searches were limited to Bacteria (taxid: 2). The results of each BLASTP search were inspected, with the superfamily and family of the aligned protein being recorded in the results section. Selecting the alignment most useful for this analysis was done non-systematically but, when possible, favoured multispecies protein alignments over single species as well as other non-*E. facieum* alignments over those from *E. facieum.* This was to ensure putative protein functions were shared across species and not unique to the *E. facieum* strain being analyzed; in horizontal gene transfer it is the shared functions that are of interest. The results were recorded in the tables of the sequent section then compared to identify protein alignments shared between the two searches. All unique alignments were collated into a new table and searched in Uniprot to gain more insight into whether the ORFs play a role in antibiotic resistance.

BIF701 Lab 6 Gene Prediction: Gene Discovery in a Resistance Plasmid
By Christopher Eeles

Results

NEB Cutter Results[3]

	Superfamily	Family	Protein
Result 1	BaeS	BaeS	MULTISPECIES: VanA-type vancomycin resistance histidine kinase VanS [Bacteria]
Result 2	Dala_Dala-lig_N	vanB	MULTISPECIES: D-alanine--(R)-lactate ligase VanA [Bacteria]
Result 3	Xf-ISL3 DDE_Tmp_ISL3	COG3464	MULTISPECIES: ISL3-like element ISEfa11 family transposase [Bacteria]
Result 4	FDH_GDH_like	HGDH_LDH_like	MULTISPECIES: D-lactate dehydrogenase VanH-A [Bacteria]
Result 5	Adenyl_tranf	Adenyl_transf	MULTISPECIES: aminoglycoside nucleotidyltransferase ANT(6)-Ia [Bacteria]
Result 6	HTH_21	n/a	IS3-like element ISEnfa3 family transposase [Enterococcus faecium]
Result 7	NT_Pol-beta-like	NTP_transf_2	MULTISPECIES: DNA polymerase [Streptococcus]
Result 8	PKc_like	APH	MULTISPECIES: aminoglycoside O-phosphotransferase APH(3')-IIIa [Bacteria]
Result 9	AdoMet_MTases	RrnaAD	MULTISPECIES: 23S rRNA (adenine(2058)-N(6))-methyltransferase Erm(B) [Bacteria]
Result 10	AdoMet_MTases	Methyltransf_25	MULTISPECIES: class I SAM-dependent methyltransferase [Bacteria]

ORFfinder Results[4]

	Superfamily	Family	Protein
ORF8	BaeS	Baes	MULTISPECIES: VanA-type vancomycin resistance histidine kinase VanS [Bacteria]
ORF1	Dala_Dala_lig_N	vanB	MULTISPECIES: D-alanine--(R)-lactate ligase VanA [Bacteria]
ORF18	Xf-ISL3 DDE_Tmp_ISL3	COG3464	MULTISPECIES: ISL3-like element ISEfa11 family transposase [Bacteria]
ORF13	FDH_GDH_like	HGDH_LDH_like	MULTISPECIES: D-lactate dehydrogenase VanH-A [Bacteria]
ORF12	Adenyl_transf	Adeny_transf	MULTISPECIES: aminoglycoside nucleotidyltransferase ANT(6)-Ia [Bacteria]
ORF11	NT_Pol-beta-like	NTP_transf_2	MULTISPECIES: DNA polymerase [Streptococcus]
ORF17	PKc_like	APH	MULTISPECIES: aminoglycoside O-phosphotransferase APH(3')-IIIa [Bacteria]
ORF3	AdoMet_MTases	RrnaAD	MULTISPECIES: 23S rRNA (adenine(2058)-N(6))-methyltransferase Erm(B) [Bacteria]
ORF16	YesN	OmpR	MULTISPECIES: VanA-type vancomycin resistance DNA-binding response regulator VanR [Bacteria]
ORF7	AdoMet_MTases	Methyltransf_25	MULTISPECIES: class I SAM-dependent methyltransferase [Bacteria]

BIF701 Lab 6 Gene Prediction: Gene Discovery in a Resistance Plasmid
By Christopher Eeles

UniprotKB Results[5]

BLASTP Result	UniProtKB ID	Molecular Function	Protein Function
VanA-type vancomycin resistance histidine kinase VanS	Q06240	Histidine Kinase	Sensor protein VanS: Member of two-component regulatory system VanS/VanR which result in vancomycin resistance.
D-alanine--(R)-lactate ligase VanA	P25051	Ligase	Vancomycin/teicoplanin A-type resistance protein VanA: Required for high-level resistance to glycol-peptide antibiotics; prevents vancomycin binding to cell wall peptidoglycan.
ISL3-like element ISEfa11 family transposase	Q8KSD8	Nucleotidyltransferase	ISL3 family transposase: Involved in catalyzing integration or removal of transposable elements in the plasmid.
D-lactate dehydrogenase VanH-A	Q05709	Oxireductase	D-specific alpha-keto acid dehydrogenase: Required for high level resistance to glycol-peptide antibiotics; prevents vancomycin binding to cell wall peptidoglycan.
Aminoglycoside nucleotidyltransferase ANT(6)-Ia	P08881	Nucleotidyltransferase	Streptomycin 3-adenylyltransferase: Mediates bacterial resistance to streptomycin and spectomycin.
IS3-like element ISEnfa3 family transposase	Q3XWK3	DNA Integrase	Involved in catalyzing integration of transposable elements into the plasmid.
DNA Polymerase	n/a	DNA Polymerase	A DNA polymerase responsible for replication of the plasmid in the cytosol.
Aminoglycoside O-phosphotransferase APH(3')-IIIa	P00551	Kinase/Transferase	Aminoglycoside 3’-phosphotrasnferase: Resistance to kanamycin and structurally-related aminoglycosides, including amikacin.
23S rRNA (adenine(2058)-N(6))-methyltransferase Erm(B)	P75864	Methyltransferase	Ribosomal RNA large subunit methyltransferase K/L: Methylates guanine at position 2445 and 2069 of 23S rRNA; required for ribosomal subunit binding.
Class I SAM-dependent methyltransferase	P0A8I5	Methyltransferase	tRNA guanine-N-methyltransferase: Methylates guanine at position 46 of tRNA.
VanA-type vancomycin resistance DNA-binding response regulator VanR	Q06239	DNA Binding	Sensor protein VanR: Member of two-component regulatory system VaS/VanR which result in vancomycin resistance.

BIF701 Lab 6 Gene Prediction: Gene Discovery in a Resistance Plasmid
By Christopher Eeles

Analysis/Discussion

Differences between the results from the two tools utilized may be explained by utilization of different restriction enzymes to generate the ORFs. For example, NEB declares that as a default it uses only those restriction enzymes which it produces, potentially explaining the minor differences observed between the searches.[3] Other factors may play a role, but it is difficult to identify them without a closer inspection of the methodology employed within each tool. It is worth noting, however, that all but two of the eleven unique ORFs identified were common to both tools.

The results from UniProtKB have yielded clear evidence of antibiotic resistance factors on the *E. faceium* plasmid used for this analysis. Six out of the eleven unique proteins identified from the ORFs are directly involved in antibiotic resistance, especially to aminoglycoside antibiotics (i.e., those with he suffix -mycin). Four of the six antibiotic resistance genes are part of the glycopeptide resistance pathway common to many *Enterococcal* strains and species.[6] This pathway modifies the structure of the bacterial cell wall to reduce the binding affinity of aminoglycoside antibiotics, thereby decreasing their effectiveness against microbes possessing this genotype.[6] While resistance is most increased to vancomycin, cross resistance is also displayed to other aminoglycoside antibiotics as the entire class functions via a similar mechanism—namely, interruption of cell wall biosynthesis.[7] VanR, the response regulator, and VanS, the histidine kinase sensor, form a two component regulatory system which triggers a phosphorylation cascade in response to perturbation of the cell wall.[6] This response results in upregulation of the vancomycin gene cluster, increase concentration of the VanH and VanA proteins which catalyze the formation of the modified cell wall peptidoglycan responsible for the manifestation of aminoglycoside resistance.[7] Given the numerous proteins involved in the glycopeptide resistance pathway, it is possible that the remaining two antibiotic resistance ORFs were unsuccessfully identified members of the system. Alternatively, the functional specialization in kanamycin resistance may indicate they are part of a separate but complementary pathway which further bolsters general aminoglycoside resistance in *E. faceium*.

The four proteins not directly involved in antibiotic resistance still may play a role as virulence factors. As discussed in Lab 4, proteins not involved directly in the development of antibiotic resistance and pathogenicity may act to facilitate the efficiency of others which are. For example, the methyltransferases identified in the results are involved in modification of rRNA and tRNA; this may increase the efficiency of translation of the six antibiotic resistance ORFs and thereby indirectly augment antibiotic resistance. The two transposases identified may enable transfer of the resistance genes between the genome and plasmids, or between plasmids within the cytosol thereby enhancing the expression and spread of these genes while under selection pressure from vancomycin or other aminoglycosides. Even the DNA polymerase may be specialized to increase the copy number of the plasmid within *E. facieum* cells and thus enable enhanced expression of the proteins encoded therein.

Conclusion

While many of these results merit further investigation to validate the putative functions identified in the analysis, it is clear that currently available bioinformatics tools provide an adequate method for annotating newly sequenced DNA elements; as well as for proposing potential structural and functional homologs in existing DNA and protein databases. These finding validate the premise of this lab and demonstrate the utility even superficial database searches can provide when exploring novel or uncharacterized genes. Inferring function of newly identified genes may not be as simple identifying

homologs, but this information certainly provides guidance on the direction further analyses should proceed upon. In some cases, a close enough match may in itself justify tentative conclusions about the function of a gene product. Through this analysis we have successfully identified genes which encode antibiotic resistance factors from simple sequence data, in the form of a FASTA file, and use of publicly available bioinformatics resources. We were able to propose several hypotheses for the function of genes less clearly involved in antibiotic resistance, which may be supported or disproven by deeper analyses. Overall, given the limited resources and time dedicated to this analysis, we have generated a wealth of information about the composition of the *E. facieum* plasmid. A team with more resources and expertise could undoubtedly accomplish much more, and such information could prove vital to further theoretical and clinical solutions to the impending crisis of antibiotic resistance.

References

1. School of Biological Sciences and Applied Chemistry. (2018). Lab 6 Gene Prediction: Gene Discovery in a Resistance Plasmid. Seneca College: Toronto, ON.
2. School of Biological Sciences and Applied Chemistry. (2018). Topic 6 Gene Prediction. Seneca College: Toronto, ON.
3. Vincze, T., Posfai, J. and Roberts, R.J. (2003). NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res. 31*: 3688 -3691. Retrieved from http://www.labtools.us/nebcutter-v2-0/
4. NCBI. (N.D.). Open Reading Frame Finder (ORFfinder). Retrieved from https://www.ncbi.nlm.nih.gov/orffinder/
5. UniProt Constortium. (2018). UniProt. European Molecular Biology Lab (EMBL): Mutiple locations, EU. Retrieved from https://www.uniprot.org/
6. Hong, HJ *et al.* (2008). Vancomycin resistance VanS/VanR two-component systems. *Adv Exp Med Biol. (631)*: 200-213. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18792691.
7. Eliopoulus, G.M., Gold, H.S. (2001). Vancomycin-Resistant Enterococci: Mechanisms and Clinical Observations. *Clinical Infectious Diseases (33)*2: 210-219. https://doi.org/10.1086/321815.