Lucas Rizkalla

January 26th: Annotation of the first gene.

Original Glimmer call @bp 1184 has strength 15.18
SSC: Start: 1184 Stop: 1684 (FWD)
CP: Includes all coding potential
SD: SD score is 735 and is the highest
SCS: Predicted by G and GM
GAP: The first gene, no 5 prime gene
BLAST: 1:1 match of gp61 to phage B4
LO: Longest ORF; 501 bp
ST: 
FS: 
ST: 

January 26th: Annotation of the second gene.

Original Glimmer call @bp 1738 has strength 9.35
SSC: Start: 1738 Stop: 2037 (FWD)
CP: Includes all coding potential
SD: SD score is 651 and is the highest.
SCS: Predicted by G and GM
GAP: Gap of 53 bps from gene 1 and gene 2
BLAST: 1:1 match with gene from Phage Riley
LO: Longest ORF; 300 bp
ST: 
FS: 
ST: 

January 26th: Annotation of the third gene.

Original Glimmer call @bp 2088 has strength 3.76
SSC: Start: 2088 Stop: 2255 (FWD)
CP: Doesn't include all coding potential
SD: SD score of 651 and is the highest
SCS: Product of G and GM
GAP: Gap of 50 bps from gene 2 to gene 3
BLAST: 1:1 match with gene from Phage BigBertha
LO: Longest ORF; 168 bp
ST: 
FS: 
ST: 

February 9th: Annotation of gp218

Original Glimmer call @bp 135511 has strength 3.45
SSC: Start: 135511 Stop: 136416 (FWD)
CP: Covers all coding potential
SD: SD score of 441
SCS: Predicted by G and GM
GAP: 23 bp gap
BLAST: 1:1 match with gp128 in phage B4
LO: Not longest ORF
ST: 
FS: 
ST: 

Annotation of gp219

Original Glimmer call @bp 136930 has strength 7.49
SSC: Start: 136930 Stop: 136424 (RVS)
CP: Covers all coding potential
SD: SD score of 567
SCS: Predicted by G and GM
GAP: 58 bp gap
BLAST: 1:1 match with putative HNH endonuclease in phage BCP8-2
February 4th

Annotation of gp220

Original GeneMark call @bp 136989
SSC: START: 136989 STOP: 137102 (FWD)
CP: Covers all reading potential
SD: Second highest SD score of 494; no difference with highest score
SCS: Predicted by GM
GAP: 58 bp gap
BLAST: 20:20 match with gp220 of phage BigBertha
LO: Longest ORF
ST:
F:
FS:

Annotation of gp221

Original Glimmer call @bp 137092 has strength of 4.38
SSC: Start: 137092 Stop: 139350 (FWD)
CP: Covers all coding potential
SD: SD score of 798 and is highest
SCS: Predicted by GM
GAP: Overlap of 11 bp
BLAST: 1:1 match with FtsK/SpolIIe protein in phages B4, Riley, Troll, BigBertha; 1:1 match with DNA translocase in phage Hoody T
LO: Longest ORF
ST:
F:
FS:

Annotation of gp222

Original Glimmer call @bp 139441 has strength 6.22
SSC: START: 139441 STOP: 139935 (FWD)
CP: Covers all coding potential
SD: SD score of 672 and highest in region of start
SCS: Predicted by G and GM
GAP: Gap of 90 bps
BLAST: 1:1 match with endonuclease in phage BigBertha
LO: Not longest ORF
ST:
F: HNH homing endonuclease
FS: HHpred

Annotation of gp223

Original Glimmer call @bp 140009 has strength 13.09
SSC: START: 140009 STOP: 140239 (FWD)
CP: Covers all coding potential
SD: Highest and only SD score of 609
SCS: Predicted by G and GM
GAP: Gap of 73 bp
BLAST: 1:1 match to phages BigBertha and Troll
LO: Longet ORF
ST:
F:
FS:

Annotation of gp224
Original Glimmer call @bp 140339 has strength 11.48
SSC: START: 140339 STOP: 141019 (FWD)
CP: Covers all coding potential
SD: highest SD score of 693
SCS: Predicted by G and GM
GAP: Gap of 174 bp
BLAST: 1:1 match with phages; BigBertha, Troll, Spock
LO: Longest ORF
ST:
F: Peptidoglycan endopeptidase RIPA
FS: HHpred
ST:

Annotation of gp225

Original Glimmer call @bp 141147 has strength 7.71; GeneMark calls start at 141126
SSC: START: 141126 STOP: 141455 (FWD)
CP: Covers all coding potential
SD: 420 highest in start region
SCS: Predicted by GM
GAP: Gap of 106 bp
BLAST: 1:8 match with gp225 in phage BigBertha
LO: Second longest ORF
ST:
F:
FS:
ST:

Annotation of gp226

Original Glimmer call @bp 141470 has strength 11.39
SSC: START: 141470 STOP: 141652 (FWD)
CP: Covers all coding potential
SD: SD score of 546
SCS: Predicted by G and GM
GAP: Gap of 14 bp
BLAST: 1:1 match with gp226 in phage Riley
LO: Longest ORF
ST:
F:
FS:
ST:

Annotation of gp227

Original Glimmer call @bp 141642 has strength 10.70
SSC: START: 141642 STOP: 142145 (FWD)
CP: Covers all coding potential
SD: SD score of 672 and is highest
SCS: Predicted by G and GM
GAP: 11 bp overlap
BLAST: 1:1 match with gp231 in phage Troll
LO: Second longest ORF
ST:
F:
FS:
ST:

Annotation of gp228

Original GeneMark call @bp 142158
SSC: START: 142158 STOP: 142256 (FWD)
CP: Covers all coding potential
SD: Highest SD score of 504
SCS: Predicted by GM
GAP: Gap of 12 bp
BLAST: 1:1 match to gp228 in phage BigBertha
LO: Longest ORF
ST:
F:
FS:
ST:
Annotation of gp229

Original Glimmer call @bp 142338 has strength 15.64
SSC: Start: 142338 Stop: 142577 (FWD)
CP: Covers all coding potential
SD: SD score of 630 and is highest
SCS: Predicted by G and GM
GAP: Gap of 81 bp
BLAST: 1:1 match with phage Riley
LO: Not longest ORF
ST:
F:
FS:
ST:

Annotation of gp230

Original Glimmer call @bp 142666 has strength 14.11
SSC: Start: 142666 Stop: 142806
CP: Covers all coding potential
SD: SD score of 455; highest and only
SCS: Predicted by G and GM
GAP: Gap of 88 bp
BLAST: 1:1 match with gp233 in phage Troll
LO: Longest ORF
ST:
F:
FS:
ST:

Annotation of gp231

Original Glimmer call @bp 142799 has strength 13.06
SSC: Start: 142799 Stop: 142957 (FWD)
CP: Covers all coding potential
SD: SD score of 840 and is highest
SCS: Predicted by G and GM
GAP: Overlap of 8 bp
BLAST: 1:1 match with gp234 in phage Troll
LO: Longest ORF
ST:
F:
FS:
ST:

Annotation of gp232

Original Glimmer call @bp 142948 has strength 14.92; GeneMark calls start at 142960
SSC: Start: 142948 Stop: 143169 (FWD)
CP: Covers all coding potential
SD: SD score of 357
SCS: Predicted by G and GM
GAP: Overlap of 10 bp
BLAST: 1:1 match with phage Riley
LO: Longest ORF
ST:
F:
FS:
ST:

Annotation of gp233
Original Glimmer call @bp 143228 has strength 13.10
SSC: Start: 143228 Stop: 143446
CP: Covers all coding potential
SD: SD score of 546
SCS: Predicted by G and GM
GAP: Gap of 58 bp
BLAST: 1:1 match with phage Riley
LO: Not longest ORF
ST:
F:
FS:
ST:

Annotation of gp234
Original Glimmer call @bp 143448 has strength 9.50
SSC: Start: 13448 Stop: 144182
CP: Covers all coding potential
SD: SD score of 651 and is highest
SCS: Predicted by G and GM
GAP: Gap of 1 bp
BLAST: No blast match
L: Longest ORF
ST:
F:
FS:
ST:

Endolysin Sources: March 2nd

Biochemical and biophysical characterization of PlyGRCS, a bacteriophage endolysin active against methicillin-resistant Staphylococcus aureus

Endolysins of staphylococcal bacteria characterize the CHAP catalytic domain as amidases or endopeptidases. The authors predicted a catalytic domain with a single cleavage specificity, but the data showed dual activities in the catalytic domain. A benefit of this dual activity is that cleaving at two locations in the peptidoglycan structure does more destabilizing than increasing enzyme concentration at one location.

This figure shows the locations that PlyGRCS cleaves in the peptidoglycan structure. The two enzymes are the N-acetylmuramoyl-Lalanine amidase which cleaves the amide bond and the D-alanyl-glycyl endopeptidase which cleaves the peptide bond.

Lysins: the arrival of pathogen-directed anti-infectives

Phage lyse its host by using a combination of two proteins, holin and endolysin. Holin works first to create a "hole" in the peptidoglycan structure which then allows endolysin to start cleaving bonds to break down the wall of the bacteria to allow phage to continue infecting bacteria.
The N-terminal enzymically cleaves peptidoglycan bonds while the C-terminal cell wall targeting domain specifies binding to peptidoglycan ligands or secondary cell wall polymers.

The N-terminal domain is characterized by multiple binding sites that can be cleaved by endolysin. This figure shows the sites for endolysin to cleave the peptidoglycan structure.

A two-component, multimeric endolysin encoded by a single gene

The endolysin studied in this article is multimeric, meaning it has two polypeptide chains encoded by a single gene. Multimeric endolysin PlyC is composed of a two CD-containing polypeptide (PlyCA) and eight PlyCB subunits with CWB activity. Lys170 is a typical endolysin with N-terminal amidase CD linked to a C-terminal CWB domain. Expression of Lys170 results in a small protein basically corresponding to the Lys170 CWB domain, and the expected full-length polypeptide. Inspection of the lys170 nucleotide sequence indicates a in-frame translational start that produces the smaller protein. It was concluded that full-length monomeric endolysin needed to increase its amount of CWB motifs to provide the best lytic activity, which was fulfilled by the CWB170 subunit.

Characterization and comparative genomic analysis of bacteriophages infecting members of the Bacillus cereus group

This article consists of a genomic analysis of the B. cereus phages. An analysis of the B. cereus sensu lato phage group I showed that that all members in the family had functional gene clusters located at identical positions. However, a comparative protein analysis of endolysin revealed the existence of four different homologous endolysin groups. Endolysin group I contained N-acetylmuramoyl-L-alanine amidase and an SH3-like domain. Endolysin group II contained cell wall hydrolysis/autolysin and an SH3-like domain. Endolysin group III contained peptidase M15B/M15C and an SH3-like domain. Endolysin group IV contained glycoside hydrolase family 25 and N-acetylmuramoyl-L-alanine amidase. The B. cereus sensu lato phage group I contains all virulent bacteriophages capable of lysing many B. cereus hosts.

My general topic of interest: Comparing functional domains of endolysin among phages and regulation of endolysin and holin when next to each other on the genome versus far apart

My research question: I am interested in the regulation of proteins holin and endolysin because I want to find out if both proteins are expressed at the same time independent of their distance on the genome. I also want to compare functional domains of endolysin in each phage to find out their similarity to create a phylogeny.

The data I will use: Predicted functional domains. Phage genomes that have holin and endolysin next to each other and that have them far apart.
Where I will find my data: Phamerator

The approach/method/tool I will use: I plan on using Phamerator to compare the predicted functional domains and use ClustalW to create a phylogenetic tree. I also plan on looking at the promoter sequences of a phage that has the proteins close together and compare those sequences to that of a phage whose proteins are far apart. I will use DNA Master to analyze.

My goal- I hope I create a figure/table that looks like this:

I will need help in this area: Analyzing the promoter sequences and comparing the sequences between phages.

March 23rd

Phages with holin and endolysin far apart on genome.

Phages with holin and endolysin next to each other.
March 25th

Endolysin

Taylor_31

MKKVTLDAHGKGDPAGVGNLKEKDLTLEIAKQTKSYLESNYSGVSVQLTRSTDFLELPERRAAIANKNKSDFVSIHINSAGGNTGFETLTYN
KLSAKSPTKSDQVLHASILNIESFGVANRKEDADLSVLRTNMSAILTELFINNPADAKLLDDKSFVKAJSVHAKGIAKVLGLKAKKAPESPVK
APSKPSTPKDGYQKVGDLTYGIARQHGMSSVDDLKLNLGKSDIIRVGQTLKVQSSSTYTKVKGDLTYGIAKDGHTTVANIKNKLNLKDLNIGD
TLRVK

SPO1_1021

MGAPFTLQELIDKSNKRLGVSLNKVYYESAIEVIKRAYKEIWQYQSYAEGYREAOQALGYAQRTKPGISVTNARGGYNHNFGLAODYFLYDDN
GKAHWVNSDWWKRVAQAIKDLGEWGGDWKSDYDAPHLEMSTGLSTALRAGKRKPILVKVKNPYSPGTSSSSSGSSKKNYLSKGDNSSAVK
TMQEKLNAAGFSGVQKADGFQAKTESLKAFQKSVGISADGLYGPTSAKLESYKPSISSSKSKGTIVLPKGVVSASSHDIVKBATATLSYP
DKGAKNGIDGYWPGKTQDAIRRQSTKSL68KTDGIGYPATRKALEKDLKEAGYTOK

Phrodo_56

MAMALQTLIDKANRLKINSGMKEKQDVRADTRAPQIQMQHAIQCVAGQFSAEQDALEYAOQRTKPSNVTNARGGGSNHNYAVDLCLYQDQG
SDVITWTEGNSFRKVIAMKGQKGFKGWDWVSFKEFHFEVLVDVGGKPPADNGGAVDNNGGSSGSGSTGGGTDGDYDSSTWFTKETGT
FTTNTSKLRPAFTSAGVATLPAGSVNVHGYGEYDGYMVIRQPRSNSGYGLATGESKGGKRQNYWGTFK

SageFayge_73

MNINTQVLTDPILKVGPNWMNPTEITFBNTYNDASAAEVRNVRNNGTSTSFHTAVDDFEVQVVPFDNWAHAGDGTYGAGNRRNSIGEVIC
YMSMSSGERYRLAENIALAEIHISDMVFNFIPISKVTQERRNGKYPHRMLDGEHGRVFWKAEERRANEKRNNGCGTNPEPKDPPKPEPTKPP
SGDYDSSSTWFTKETGFVTFNTTITKLRPAFTSAGVATLPAVTNVHGYGEYDGYMVIRQPRSNSGYGLATGESKGGKRQNYWGTFKZ

March 30th

Endolysin

Waukesha_68

MEIRKINLVDASKYTGKCPYPYTNNPEITWNYTNDATANNEVAYMIRNDQVSHTAVDDKEAVQIPLERNWWHCDDGQGGNENKSIQVEVIC
YLSGGYDKAEHNAIVVAGLMLKQNIPISKVRTHQSWGKYCHPRMLAEGRNWFIERVQNYNGGSVPMTTPRPSNQGTKVAYINGDNVNLRK
GPQGTYAVIRAKLGKCECYQWGEGNWNLNGGDQWVYNDSYIRYTGENAPSPKPSNDDGIVVTITADVLRVTRPGNTNYGVKYNVYQERYQ
SWGYRDGVYNVGGDQWVSGEYVKFEK

April 8th
B4_6
N-10-136, C-80-252

Troll_59
N-16-136, C-177-249

Taylor_31
N-1-188, -198-300
### Gamma_17

N-1-168, C-188-234

### Nigalana_74

N-18-157, C-193-265

### SPO1_1021

N-19-135, C-177-333
April 13th

DIGNKC_69

MNINTQYLVTDPRLEVIGPYPLNPTEITFHNTYNDASASEVRNVRNSKGTSSHTAVDDFEVQQVFVQRFNHWADGDNGAGNRSIGVEICYSMSGGERYRKAEAIHISDLMRVFNPISPSKVKTQERNKGKCPHRMLDEGRGVWKAEERCERRANEKRNNGGGGTPTEP KPDPAPKPPSGDYDSSWFKETGTFVTNMTKLRTAPFTSAGVIATLPAGSTVNYNGFIEYGDGYVRQPR SysNYGYLATGEGSKGGKVRNYYWGFKZ

Zuko_70

MNINTQYLVTDPRLEVIGPYPLNPTEITFHNTYNDASASEVRNVRNSSTGTSFHATVDDFEVQQVFVQRFNHWADGDNGAGNRSIGVEICYS LGGEGERYRKAEAIHISDLMRVFNPISPSKVKTQERNKGKCPHRMLDEGRGVWKAEERCERRANEKRNNGGGGTPTEP KPDPAPKPPSGDYDSSWFKETGTFVTNMTKLRTAPFTSAGVIATLPAGSTVNYNGFIEYGDGYVRQPR SysNYGYLATGEGSKGGKVRNYYWGFKZ

Jugalone_56

MAMALQTLIDKANRGKINSGMRKDVAADRTAVITQMHMQIYICVAQGRFSAEQDALYAOGRTPGIVTNARGGQSNHYGVAVDLCLYTQDG SDVIVTVEGFRKVIAMKGQGSKFWGDWFSKFDYPHELEYDVGOGGKPPADNGGAVDNGGGSQCGSNGGCTGGGTGDYDSSWFKETGFTTNMTKLRTAPFTSAGVIATLPAGSVVNYNGFIEYGDGYVRQPR SysNYGYLATGEGSKGGKVRNYYWGFKZ

For Promoters:

Nigalana_74 Start: 27949 Stop: 27122 REV

Phrodo_56 Start: 22799 Stop: 22020 REV

April 15th

Phrodo Promoter:

Sigma 32-27,084
-10: TACTAATA
-35: CCCGC

Nigalana Promoter:

Sigma 32-32,802
-10: TGTAACAC
-35: CCCGC

April 16th

Endolysin functional domains.docx

April 17th

Vinny_63
April 20th

Finn_30

MKKVTLDAGGGDGAQNGLEKDLTLEIAKOTKSYLENSYSGVSVQLTRSTDKFIELPERAAIANKNSDLFVSIHINSAGGTNGTGFETFTYN
KLASSQTDQKVHLHASLNEIAFVGTVNEKADLQLRNTMSAITEFLINNPAADKLLKDFKAVGSHAKGIAKVGLKAKKPEPSK
APSKPSTPKGDTKYVKGDTDLYGIAKQHGMSVDDLKKNLNGDFTLGYGIAKQHTTVANIKKLNLKSDLINIGDVLRKV

Eoghan_31

MKKVTLDAGGGDKPGAVNGNLDKDLTLEIAKOTKSYLENSYSGVSVQLTRSTDKFIELPERAAIANKNSDLFVSIHINSAGGTNGTGFETFTYN
KLASSPTQDKVLHASLNEIAFVGTVNEKADLQLRNTMSAITEFLINNPAADKLLKDFKAVGSHAKGIAKVGLKAKKAPESPK
APSKPSTPKGDTKYVKGDTDLYGIAKQHGMSVDDLKKNLNGDFTLGYGIAKQHTTVANIKKLNLKSDLINID VLRKV

Polaris_32

MKKVTLDAGGGDKGAQNGLEKDLTLEIAKOTKSYLENSYSGVSVQLTRSTDKFIELPERAAIANKNSDLFVSIHINSAGGTNGTGFETFAEK
KLASSQTRKDQKLHASLNEIAFVGTVNEKADLQLRNTMSAITEFLINNPAADKLLKDFKAVGSHAKGIAKVGLKAKKPEPSK
APSKPSTPKGDTKYVKGDTDLYGIAKQHGMSVDDLKKNLNGDFTLGYGIAKQHTTVANIKKLNLKSDLINIDVL RKV

Gemini_32

MKKVTLDAGGGDKGAQNGLEKDLTLEIAKOTKSYLENSYSGVSVQLTRSTDKFIELPERAAIANKNSDLFVSIHINSAGGTNGTGFETFAEK
KLASSQTRKDQKLHASLNEIAFVGTVNEKADLQLRNTMSAITEFLINNPAADKLLKDFKAVGSHAKGIAKVGLKAKKPEPSK
APSKPSTPKGDTKYVKGDTDLYGIAKQHGMSVDDLKKNLNGDFTLGYGIAKQHTTVANIKKLNLKSDLINIDVL RKV

Curly_32

MKKVTLDAGGGDKGAQNGLEKDLTLEIAKOTKSYLENSYSGVSVQLTRSTDKFIELPERAAIANKNSDLFVSIHINSAGGTNGTGFETFAEK
KLASSQTRKDQKLHASLNEIAFVGTVNEKADLQLRNTMSAITEFLINNPAADKLLKDFKAVGSHAKGIAKVGLKAKKPEPSK
APSKPSTPKGDTKYVKGDTDLYGIAKQHGMSVDDLKKNLNGDFTLGYGIAKQHTTVANIKKLNLKSDLINIDVL RKV

Andromeda_32

MKKVTLDAGGGDKGAQNGLEKDLTLEIAKOTKSYLENSYSGVSVQLTRSTDKFIELPERAAIANKNSDLFVSIHINSAGGTNGTGFETFAEK
KLASSQTRKDQKLHASLNEIAFVGTVNEKADLQLRNTMSAITEFLINNPAADKLLKDFKAVGSHAKGIAKVGLKAKKPEPSK
APSKPSTPKGDTKYVKGDTDLYGIAKQHGMSVDDLKKNLNGDFTLGYGIAKQHTTVANIKKLNLKSDLINIDVL RKV

Curly_32
Endolysin region in Jugalone (top) and Phrodo. We noticed the terminase and endolysin proteins are in the same region. Both are required late in the phage lifecycle. Terminase isn’t needed until the head and tail are made, and the head is ready to be stuffed with DNA before connecting the tail. Then endolysin can cleave the cell wall, releasing fully mature phages. The region around 27000 is 'suspicious' as a promoter region for transcription from right to left of the genes on the reverse strand.

We found a sigma 32 promoter in this region. The -10 is highlighted in yellow and the -35 is highlighted in green:
Holin region in Jugalone and Phrodo. I looked for an area that is homologous (purple shading) between Jugalone and Phrodo, to the left of the holin gene. 106,600 region in phrodo stood out as the closest area with significant non-coding region, and present in both phages.

We found a sigma 70 promoter in this region. I didn't find any proteins in this area with functional predictions that would let me know when the genes here should be transcribed. The -10 is in yellow and the -35 is in green.

Here is the map region for B. anthracis phage gamma. Holin is gp 16, endolysin is gp17. The suspicious region is the gap around 7,700.

We found a sigma 32 promoter in this region.
Could you craft an argument that holin and endolysin that aren't near each other don't need the same promoter? I think it would be more important to regulate endolysin production because that is the one that actually does the cleaving. And I think holin is the protein that 'moved' some time way back. It must have plopped down in front of a sigma 70 promoter and been o.k.?

May 8th

Final Paper-Lucas Rizkalla.docx

Endolysin_Final.fasta.txt

Here is the SPO1 gene area for holin and endolysin and a pic from their promoter prediction table:

![Diagram of gene area with promoter prediction table]

Found a paper describing promoters in phage Fah, where endolysin and holin (gp15/16) are expressed after 20 minutes of infection (not early). "For example, expression of genes 5, 10, and 16 (and presumably other genes that belong to the structural/lysis gene neighborhood) was undetectable during the first 20 min post infection but increased dramatically afterwards"

Need to compare gamma and fah promoters, SPO1, and the compare to Phrodo to see if our predictions make any sense.