Identification of Endolysin and Glycosyl Hydrolase Family 25 Proteins in *Streptomyces*

Sky Harper, Dr. Christopher Clarke – United States Department of Agriculture – Agricultural Research Services

Research Question

How can endolysins (glycosyl hydrolase family 25 proteins) be identified through database research and genome BLASTing for laboratory synthesis in order to investigate a new form of antimicrobial?

Hypothesis

If annotated endolysins in the NCBI database are found and their genomes are copied, then it is possible to compare against other possible enzyme candidates to determine if they are endolysins by BLASTing against known phages and the construction of a phylogenetic tree and percent identity tables,

Table 1. Variables

Independent	Dependent	Lab-Grade Control
Glycosyl Hydrolase family 25 proteins	Annotated endolysins in NCBI	Annotated endolysins in NCBI

Software/Programs/Materials

- National Center for Biotechnology Information Database
- Notepad++ for FASTA file
- JGI Integrated Microbial Genomes & Microbiomes for custom protein BLAST
- CAZy Website (Carbohydrate Active enZYmes)
- Phagesdb – The Actinobacteriophage Database – identifying phages
- SignalP-5.0 Server – identify secretion pathways
- Inkscape – labeling the phylogenetic tree
- Geneious Prime – a bioinformatics software solution with many fundamental molecular biology and sequence analysis tools
- Molecular Evolutionary Genetics Analysis- computer software for conducting statistical analysis of molecule evolution and for constructing phylogenetic trees

Experimental Procedures

Figure 1. Example entry in NCBI showing where data was found

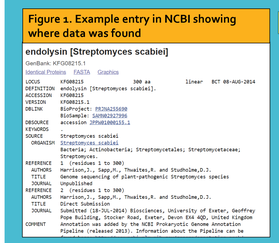

Figure 2.Excel file with first 4 protein entries

Accession/Lo	Genus	specific ep	Strain	Protein Nam	Description/phage	Top Phage	Identical	Alignm	Secretion Path	Probab	Domain
KFG08215	Streptomy	scabiei	NCPPB:4	endolysin		MT310898	53.70%	178-298	Other	0.9758	CW_7
YP_009615357	Streptomy	scabiei	RL-34	endolysin	phage Scap1	MF975637	100.00%	1-289	Other	0.9202	CW_7
ATN93653	Streptomy	scabiei	RL-34	endolysin	phage Scap1	MF975637	100.00%	1-289	Other	0.9202	CW_7, Amidase_2
QJD49887	Streptomy	scabiei	RL-34	endolysin	phage CricKo	MH155877	99.60%	1-280	Other	0.9886	Amidase_2

Figure 3. Running FASTA file with protein sequences

```
>KFG08215.1 endolysin [Streptomyces scabiei] scabiei NCPPB:4066

MPELWMPGAQRLDIGDHAPTDGGPPKAIAHITWDVNATPKAPKDLVPYERLRSYFSGSGKSAAPHVLWDP
FTGRIVQFLPATSRSKSLADRAGGTRTNRAGKVVLQVEALFFPHCRVDGKVYATLADTPCEGWAELNAWI
RSWGVPDVWPMGKPINFLPRRSEATWREEAGWYAHAHVPENDHQDPGSWPAFAKAPTEPSPPATGRPVVD
LSELVKAAKGDPPKRGTPVSYAGVKVVETALVAEGLLARSLADGHFGTATVEAYGLWQRRCGWSGPDADG
IPGKASLTKLGKRRGFDVKE
```

BLASTing against other databases and known phages → Data analysis and manipulation

Literature Review

Streptomyces is an incredibly diverse genus of actinomycete gram-positive bacteria, of which several species are phytopathogens. Specifically, 22 species were considered in this investigation; 22 virulent and nonvirulent species were considered because it was found that some nonvirulent species had the ability to turn virulent. Diseases of potatoes and other root crops caused by these phytopathogens are prevalent and costly.

There are several bacteriophages that are known to exist within the *Streptomyces* genome. Bacteriophages are viruses that infect bacterium. The phage genome is integrated into the genome of the host bacteria, but in order to escape and infect other bacteria, the phage must escape. The phage genome contains a gene that encodes for an enzyme known as an endolysin. Some lytic phages use endolysins (enzymes) to lyse the cell wall of the bacteria, ultimately killing the bacteria in the process.

These enzymes may have use in agriculture as antimicrobials against *Streptomyces*. This is because the enzymes found in the phages within the *Streptomyces* genus have specifically evolved to degrade the cell walls of *Streptomyces*. Although the research is focused primarily on endolysins, any enzyme that can cut peptidoglycan bonds is a potential endolysin and has the potential to serve the purpose of lysing the bacterial cell walls.

The GH25 family is the most promising because many enzymes in that family are known to be capable of completely lysing bacteria or are known to be true endolysins. There are other enzymes and proteins involved with the cell wall, however some are strictly for remodeling.

Data and Graphs (All Images credit of researcher)

Figure 4. Alignment of protein sequences using MEGA

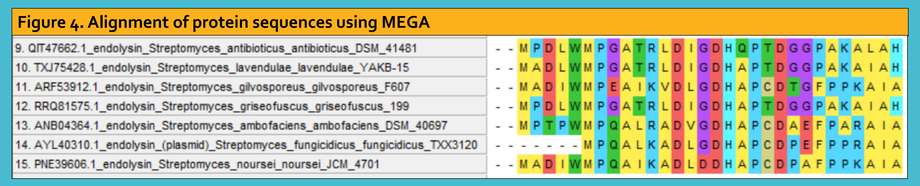

Figure 5. Codon Optimization Window for Bacillus using Geneious

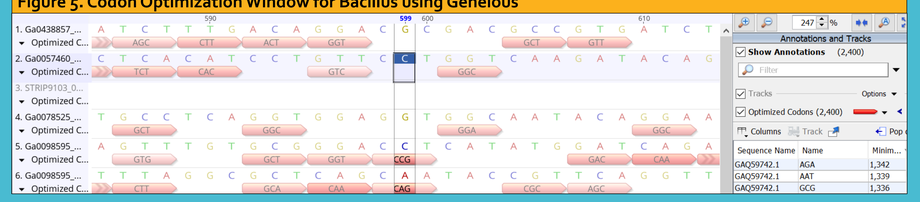

Figure 6. Codon Optimization Window for E. Coli using Geneious

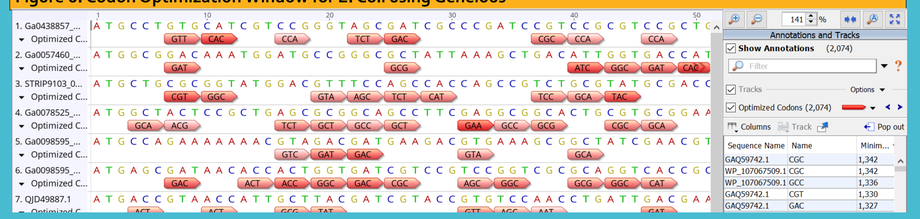

Figure 8. Portion of the finalized (labeled) phylogenetic tree

Data Analysis

Using the NCBI database and CAZy database, a total of 342 endolysin candidates were identified. From these 342 candidates, 13 total clades were formed and 15 endolysins were identified for synthesis in the laboratory setting. Synthesis was determined by the domains and other information found, such as Phages.

Clade 1 - 56 proteins, 32.00% match. 2 proteins were chosen for synthesis.
Clade 2 – 2 proteins, 100.00% match. 1 protein was chosen for synthesis.
Clade 3 - 52 proteins, 1.77% match. 2 proteins were chosen for synthesis.
Clade 4 - 2 proteins, 92.72% match, 1 was chosen for synthesis.
Clade 5 - 2 proteins, 63.64% match. No proteins were synthesized.
Clade 6 – 43 proteins, 77.61% match. No proteins chosen.
Clade 7 – 6 proteins, 87.00% match. 1 protein chosen.
Clade 8 – 2 proteins, 94.00% match. No proteins chosen.
Clade 9 – 52 proteins, 25% match. 1 chosen for synthesis.
Clade 10 – 3 proteins, 28.9% match, 2 chosen for synthesis.
Clade 11 – 6 proteins, 37.67% match. No proteins chosen.
Clade 12 – 26 proteins, 87.39% match. 2 chosen for synthesis.
Clade 13 – 39 proteins, 16.35% match. 2 chosen for synthesis.
Protein QFP97359 was not contained within a clade and was chosen for synthesis.
The phylogenetic tree chosen had a total of 342 proteins and was built using 200 bootstrap.

Conclusions

Overall, this investigation ultimately met its goal to identify potential proteins for synthesis. The 15 proteins that were identified were carefully picked from 342 candidates. This was done through the use of the Basic Local Alignment Search Tool to BLAST the GH25 proteins against already sequenced phages. Furthermore, the proteins domains and secretion pathway were also identified. Ultimately, a phylogenetic tree was created and used to show the relationships between the proteins.

The protein codons were then optimized to aid in the laboratory synthesis using E.Coli or Bacillus.

This research endeavor was conducted in hopes of identifying a new category of antimicrobials that can be applied to the agricultural industry.

References

- *The Actinobacteriophage Database: Streptomyces*. The Actinobacteriophage Database| Streptomyces. (n.d.). https://phagesdb.org/hosts/genera/4/.
- *Glycoside Hydrolase Family 25*. CAZypedia. (n.d.). https://www.cazypedia.org/index.php/Glycoside_Hydrolase_Family_25.
- Love, M. J., Bhandari, D., Dobson, R. C. J., & Billington, C. (2018, February 27). *Potential for Bacteriophage Endolysins to Supplement or Replace Antibiotics in Food Production and Clinical Care*. Antibiotics (Basel, Switzerland). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5872128/.
- National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2017 Apr 06]. Available from: https://www.ncbi.nlm.nih.gov/
- Nelson, D. C., Schmelcher, M., & Rodriguez-Rubio, L. (n.d.). Endolysins as Antimicrobial. *Advances in Virus Research*, *83*.
- Schmelcher, M., Donovan, D. M., & Loessner, M. J. (2013, August 1). *Bacteriophage endolysins as novel antimicrobials*. Future microbiology. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3563964/.
- Thapa, S. P., Davis II, E. W., & Lyu, Q. (2019). The Evolution, Ecology, and Mechanisms of Infection by Gram-Positive, Plant-Associated Bacteria. *Annual Review of Phytopathology*, *57*.