

Gordonia Terrae
phage Charianelly
contains 4 novel
genes capable of
infecting 2 out of 3
alternate hosts

Hamilton, Reagan Khan, Ayaan
Mashru, Simran

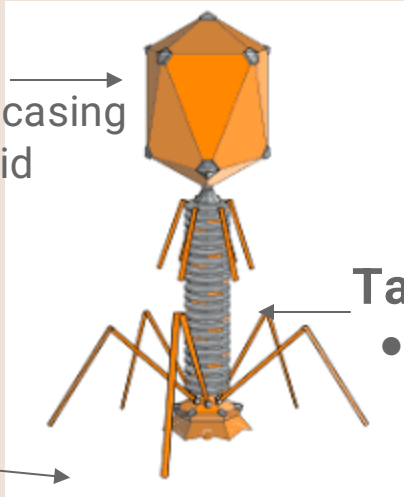
BIOSC 0067 Foundations of Biology 2
Lab Fall 2023 semester

University of Pittsburgh Department
of Biological Sciences

The study of a bacteriophage

Capsid Head

- Protein encasing nucleic acid genome



Tail Sheath

- Channel for genome delivery

Tail Fibers

- Mediate phage binding

Fig. 1: Bacteriophage

SEA-PHAGES program- undergraduate research that focuses on discovery and retention in the biological field

Gordonia Terrae

- Used by pitt sea phage students
- Useful because it is related to *M. smegmatis*
- Found within the soil

- Viruses that infect bacteria
- Can expand host range by new mutations and genes
- Most abundant entity on Earth and widely diverse

Why studying phages is important

Experimental:

- Exploration in host interactions and genetics
- SEA-PHAGES 1 and SEA-PHAGES 2 (Hatfull lab)

Medical:

- Phage therapy to eliminate bacterial infections
- Help in antibiotic resistance

Technological:

- Genome editing: CRISPR inspired by phage
- Development of biosensors

Evolutionary:

- Provide insight to evolutionary dynamics with bacteriophages and their hosts

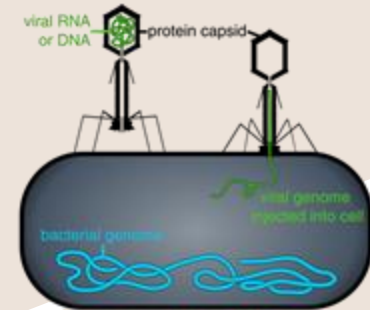


Fig 2: Bacteriophage infection

Genome annotation and its significance

- Genome annotation is the process of trying to identify functional parts within the sequence of a genome, giving it an identity and more context in order to more accurately understand it
- Very useful in modern medical science as well as evolutionary science
 - Medical science has been able to experience breakthroughs due to us being able to identify key functions within genes quicker and more efficiently
 - Evolutionary science has also made breakthrough now that we can use genome annotation to make more accurate inferences about how these sequences might have evolved

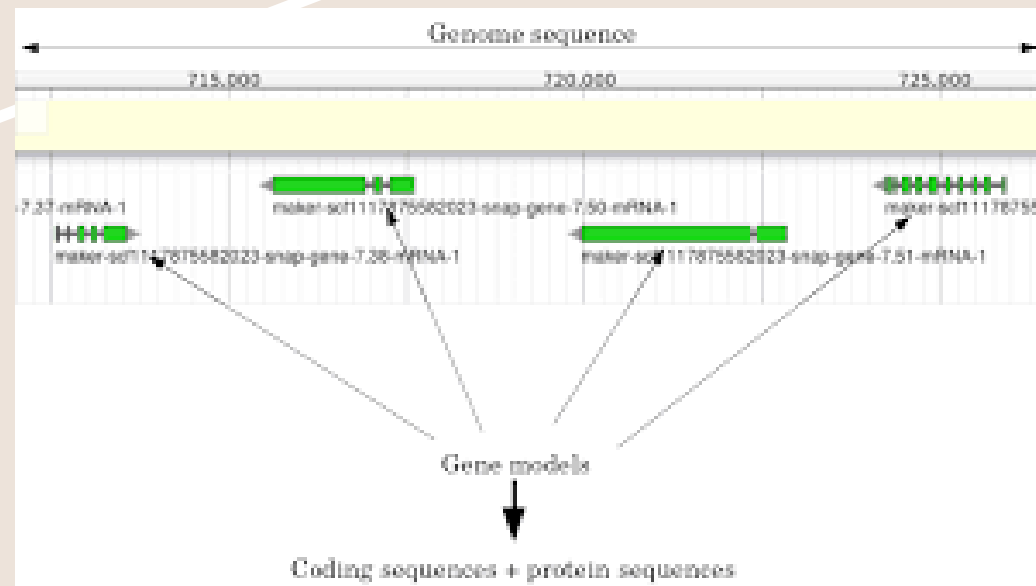


Fig.3: Genome sequence

1. Determine if similar viruses have been identified

The screenshot shows the phagesdb.org homepage. At the top, it says "phagesdb.org Charlan". Below this is a circular image of a phage. To the right of the image is a "Locally Blast Your genome" button. Below the button is a link to "Go to phagesdb.org to locally blast genome of phage Charlianelly and copy entire protein sequence".

Repeat with Minos and analyze results

2. Determine where genes are located in the genome

Navigate to PECAAN and search for phage Chariannely and click on "Genes Tab" and select desired gene number then copy protein sequence from "sequence"

Example of
Charianelly 15
shown

Analyze Blast conservation,
alignment and function

Use Glimmer and compare to Genemark start to predict true start from PECAAN (Example is Chariannelly 15)

Page	Channel	Cluster	CS
Glimmer Start:	Glimmer Score:	Genshark Start:	Plan
16757	13.18	16757	Starline: 114
			100
			Phase: 114

Click on Pham
Statorator to
open up
starterator
report

Record Gap/Overlap and take note of -1 and -4 overlaps. Conclude final start

[illegible]

Paste protein sequence again into HHPred. Observe functions and hits above 80% and 90%

► Navigate to Phamator and Select Phages. Compare Chariannely-Draft with Bianmat and Minos in cluster CS3 and view map

Analyze if there are close relatives between Bianmat and Minos. Record pham numbers and functions. Conclude an overall function.

Analyze if all phages have the same autoannotated start and then see which start was most manually annotated. Compare this for final nucleotide start.

Figure 4: Methods Flow Diagram for Overall Phage Annotation

GeneMark and Glimmer are two programs that are used to predict a gene's start location

GeneMark:

- GeneMark scans the chosen genome for start/stop codons
 - Examines Open Reading Frames (ORFs) for continuous presence of the host's preferred codon (codon bias)
- Genemark correctly calls about 9 out of 10 times
 - Human input needed for the other 10% of annotation
- Doesn't call TTG starts

Codon Bias:

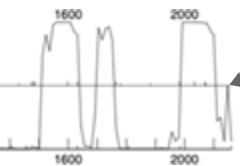
- Phages tend to have codons they prefer, which show up in higher frequency
 - This known as **Codon Usage Bias**
- Codon usage Bias differs among phages

UTG F 0.57	UUG N 0.11	UUG Z 0.53	UUG C 0.42
UUG F 0.43	UUG N 0.11	UUG Y 0.47	UUG C 0.58
UGA L 0.15	UGA N 0.15	UGA A 0.44	UGA N 0.36
UGA L 0.12	UGA N 0.14	UGA N 0.50	UGA W 1.50
CGU L 0.12	CGU F 0.17	CGU N 0.55	CGU N 0.36
CGU L 0.18	CGU F 0.13	CGU N 0.45	CGU N 0.44
CGA L 0.08	CGA F 0.14	CGA Q 0.30	CGA N 0.57
CGA L 0.44	CGA F 0.38	CGA Q 0.75	CGA N 0.67
AGU Z 0.56	AGU Y 0.14	AGU N 0.47	AGU N 0.14
AGU Z 0.35	AGU Y 0.47	AGU N 0.53	AGU N 0.33
AGA Z 0.07	AGA Y 0.13	AGA W 0.73	AGA N 0.52
AGA W 1.00	AGA Y 0.14	AGA W 0.27	AGA N 0.53
GUU W 0.25	GUU A 0.11	GUU N 0.65	GUU G 0.29
GUU W 0.10	GUU A 0.31	GUU N 0.35	GUU G 0.46
GUU W 0.17	GUU A 0.21	GUU N 0.70	GUU G 0.13
GUU W 0.49	GUU A 0.35	GUU N 0.70	GUU G 0.12

[Codon usage/translation per codon per a.a.]
E. coli K12 Data from the Codon Usage Database

- As shown in the image, different codons have different frequencies within a phage

How it works:

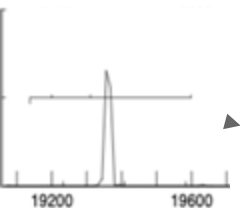


Example of "Strong" coding potential (high intensity waves, more than one) - Charianelly Gene 12

Phage: Charianelly Cluster: CS3

Glimmer Start:	Glimmer Score:	GeneMark Start:
14804	10.79	14804

When GeneMark and Glimmer agree: Strong chance that the start position is correct (Charianelly Gene 12)



Example of "Weak" coding potential (one low intensity wave, nothing else) - Charianelly Gene 12

Phage: Charianelly Cluster: CS3

Glimmer Start:	Glimmer Score:	GeneMark Start:
56266	8.51	56176

When GeneMark and Glimmer disagree: Glimmer predictions are usually more accurate

Glimmer:

- Coding potential program similar to GeneMark
- Does "self modeling"
 - Scans entire sequence for all the possible ORFs in all frames
 - Samples the longest frames for predicting the phage's preferred codons (codon bias)
 - Calls TTG starts
- Tends to be more accurate than GeneMark

This is the start codon site according to glimmer

How it works:

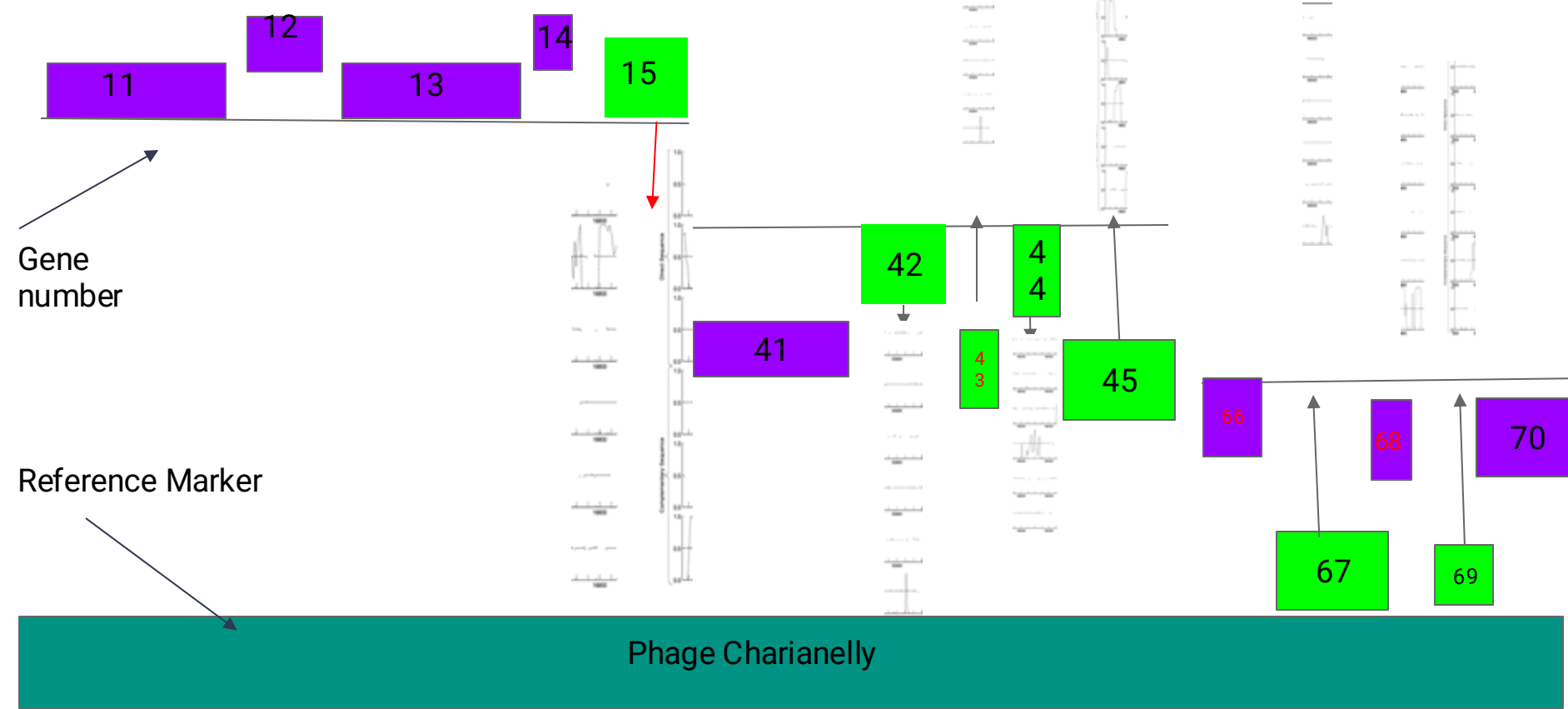
Phage: Charianelly Cluster: CS3

Glimmer Start:	Glimmer Score:	GeneMark Start:
14804	10.79	14804

Green: Weak coding potential
Purple: strong coding potential
Red Text: Glimmer and Genemark do not agree

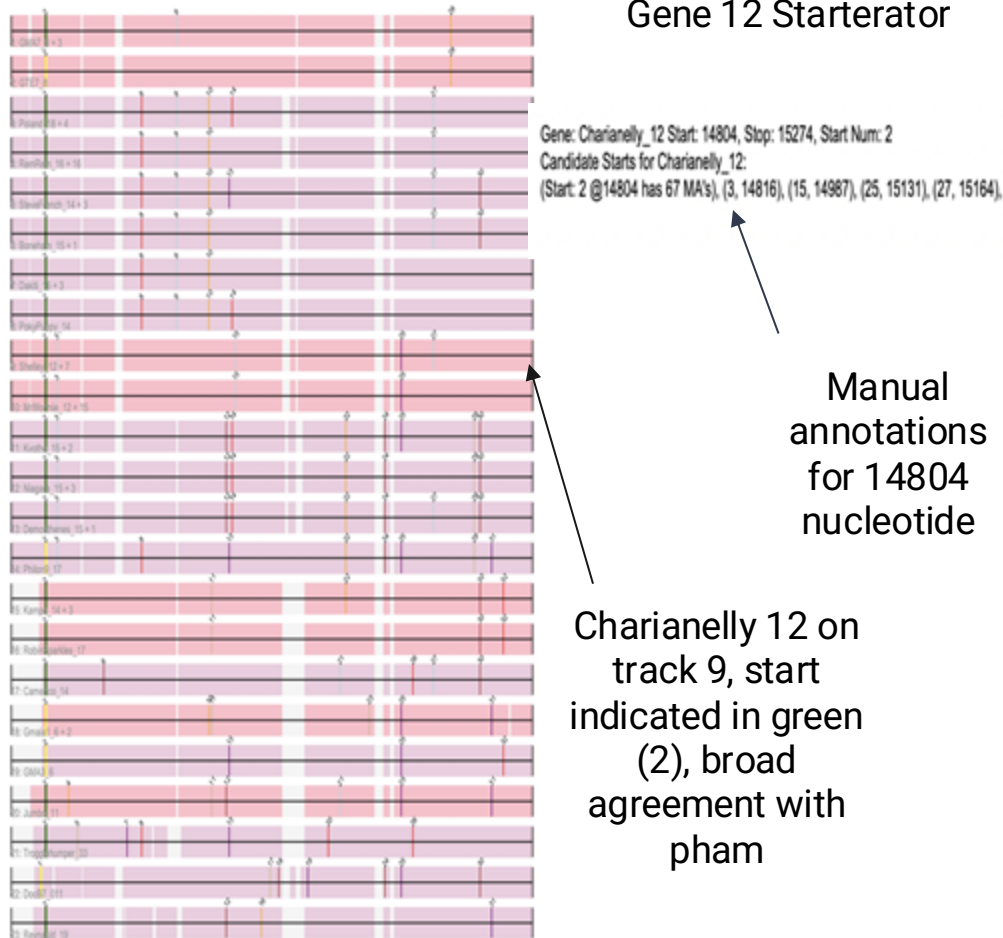
- Below is the size and chromosomal location of 15 genes from Charianelly that our group annotated
- Use of both Glimmer and Genemark

Figure 5: Coding Potential for Charianelly



Coding potential intensity is varied for different gene predictions

Figure 6: Charianelly
Gene 12 Starterator



- Starterator compares each predicted gene within phage's genome with sequence of all members of same pham
- Determines where gene likely starts in consensus
- Able to view how many times nucleotide coordinate has been manually annotated
- Choosing correct start important to differentiate each functional domain

How it works

1. Open starterator report
2. View if phage tracks (right) have same color and same start in green
3. Determine nucleotide coordinate by seeing how many manual annotations there are (More MA = likely start)
4. Observe if there is agreement with cluster and subcluster

Pink = same cluster, Purple = same subcluster, Yellow = called my auto-annotation, Green = Manual Annotations

Start codons are selected using specific annotation programs in relation to similar phages.

- Phage genomes are densely packed with not much overlap measured in base pairs
- Gap of 100+ and Overlap of 30+ rare
- Overlap of 1 or 4 base pairs are best start sites
 - 1 overlap should ALWAYS be selected
- Why it is a strong indicator: Ribosome slide back 1 bp or forward 2bp to translate next gene
- Useful because lower overlap predicts start site better due to packed genome

EXAMPLES

4 bp overlaps:

ATGA

GTGA

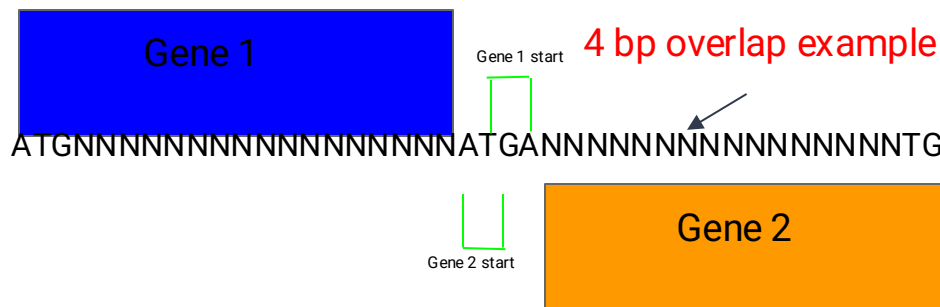
TTGA

1 bp overlaps:

TGATG

TAATG

Charianelly 12 Gap:
Only 2 starts below acceptable range of 100, No -1 or -4 overlap



Show 10 entries

Search:

Direction	Start	Stop	Length	Gap	Spacer	Z-score	Final Score	LORF	Start Codon	All GM Coding Capacity	Selected Gene
Forward	14804	15274	471	36	12	2.929	-2.786	TRUE	ATG	Select	<input checked="" type="checkbox"/>
Forward	14816	15274	459	48	14	1.745	-5.786		GTG		<input type="checkbox"/>
Forward	14987	15274	288	219	10	2.142	-4.299		GTG		<input type="checkbox"/>
Forward	15131	15274	144	363	16	1.303	-7.165		GTG		<input type="checkbox"/>

Figure 7: Charianelly Gap/Overlap

The gaps in genes also guide start site selection

- No final start calls were changed
- Gene 43, 66 and 70 had an inconclusive start position
- Glimmer and GeneMark for gene 43, 66, 68 do not agree on the start position
- Gene 67 had no geneMark start site data, so the position is based on Glimmer analysis

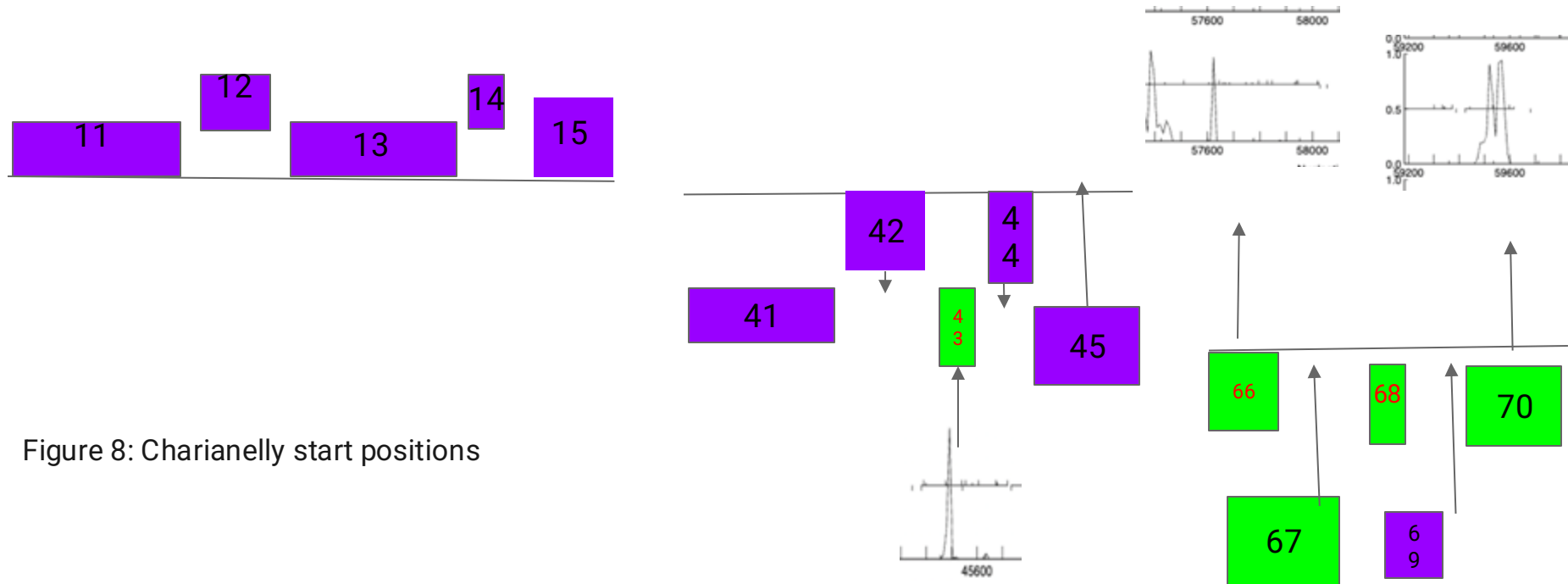


Figure 8: Charianelly start positions

Phage Charianelly

Created with BioRender.com

Based on our results the Glimmer and GeneMark start codons were either changed or left as they were originally annotated

HHpred

- Alternate option to BLAST to determine sequence similarity
 - Useful to help us find similarity to phages that have established functions that are most likely present in our phage as well
- Based on structural similarity
- Consists of both similarity probability as well as a visualization of said similarity



Figure 9:
HHpred for
Charianelly

BLASTp

- Helps determine sequence similarity, along with HHpred.
- Runs the amino acid sequence of predicted proteins through the NCBI database
 - Our phage's sequence gets measured against every other phage protein sequence in the database
- Generates both a "similarity" value and an "E value"
 - Based on BLOSUM matrix which weighs how likely an amino acid occurs in reality

Descriptions	Graphic Summary	Alignments	Taxonomy
Sequences producing significant alignments			
Download Select columns Show 100			
select all 23 sequences selected			
Description	Scientific Name	Max Score	Total Score
trv00000001 SEA_A1000000_13 (Gordonia shakle A1000000)	Gordonia shakle A1000000	189	189
trv00000001 SEA_SHELLEY_13 (Gordonia shakle Shelley)	Gordonia shakle Shelley	185	185
trv00000001 SEA_MWORMIE_13 (Gordonia shakle MWormie)	Gordonia shakle MWormie	184	184
trv00000001 SEA_NUCLEUS_13 (Gordonia shakle Nucleus)	Gordonia shakle Nucleus	184	184
trv00000001 SEA_STICKER17_13 (Gordonia shakle Sticker17)	Gordonia shakle Sticker17	182	182
trv00000001 SEA_LUKER_13 (Gordonia shakle Luker)	Gordonia shakle Luker	181	181
trv00000001 SEA_NEED_13 (Gordonia shakle Need)	Gordonia shakle Need	181	181
trv00000001 SEA_HAZZETT_13 (Gordonia shakle Hazzett)	Gordonia shakle Hazzett	178	178
trv00000001 SEA_HELLO_13 (Gordonia shakle Hello)	Gordonia shakle Hello	172	172
trv00000001 SEA_ARMANDER_13 (Gordonia shakle Armander)	Gordonia shakle Armander	124	124
trv00000001 SEA_LIGAND_13 (Gordonia shakle Ligand)	Gordonia shakle Ligand	124	124
trv00000001 SEA_JAMM_13 (Gordonia shakle Jamm)	Gordonia shakle Jamm	121	121
head-tail connector system (Gordonia shakle X000000)	Gordonia shakle X000000	120	120
head-tail connector system (Gordonia shakle B1000000)	Gordonia shakle B1000000	73.6	73.6
trv00000001 SEA_R000000_13 (Gordonia shakle R000000)	Gordonia shakle R000000	73.2	73.2
head-tail connector system (Gordonia shakle S1000000)	Gordonia shakle S1000000	73.2	73.2

Figure 10: BLASTp for Charianelly

What did we use to determine sequence similarity?

- Genes 11,12, and 13 had a function identified.
 - Both HHpred and Pharmerator both showed positive results for them to be identified as such

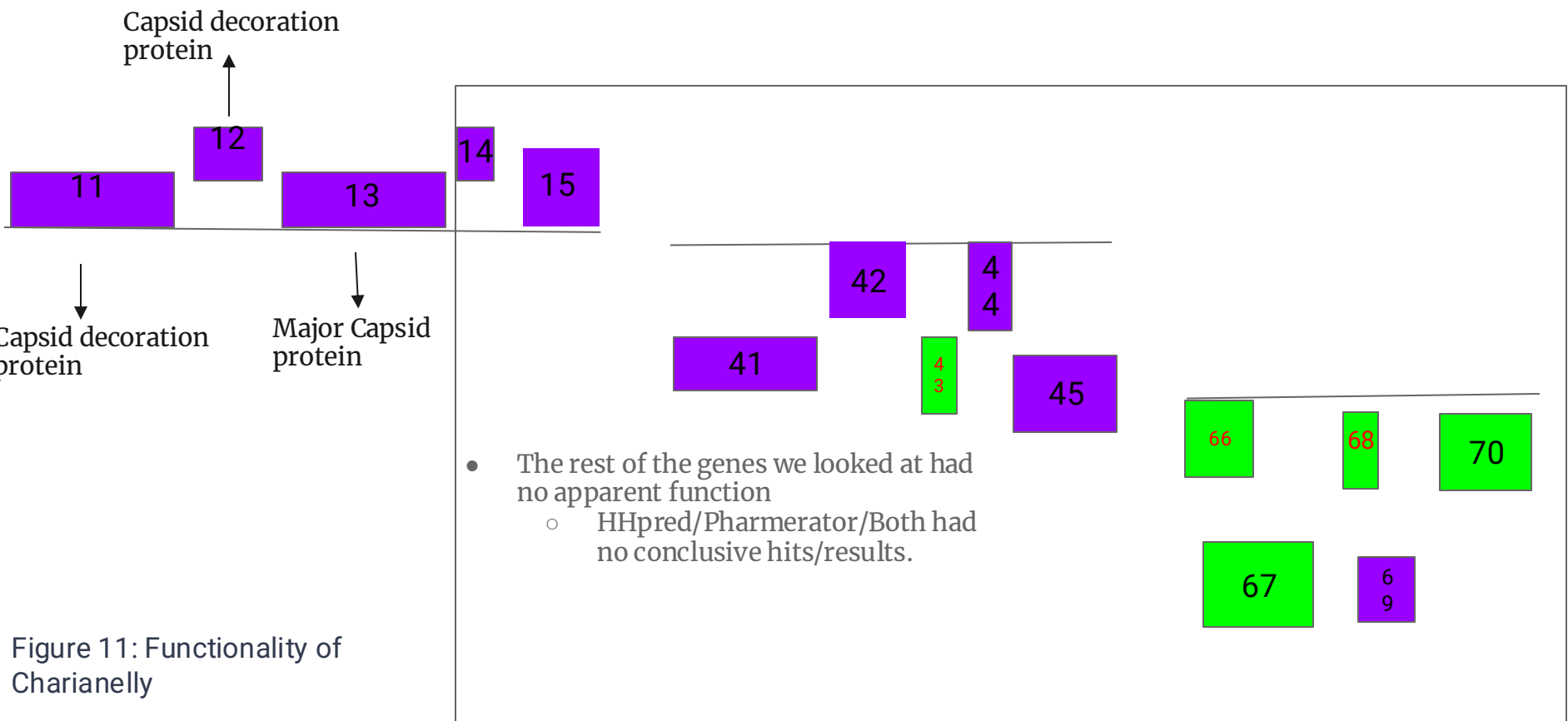


Figure 11: Functionality of Charianelly

Functionality of our various genes for Charianelly varied between various proteins and the majority having no function.

Host range is described as the range of cell types a host species is able to infect

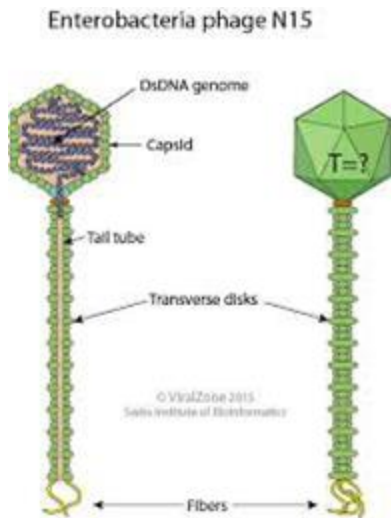
The importance of knowing the host range

The background research of annotating the charianelly phage, and understanding what classifies as a gene laid the groundwork for the host range project

Minor tail proteins are a crucial part in the assembly of the tail in the bacteriophage that makes it easier for viruses to infect their hosts

Importance

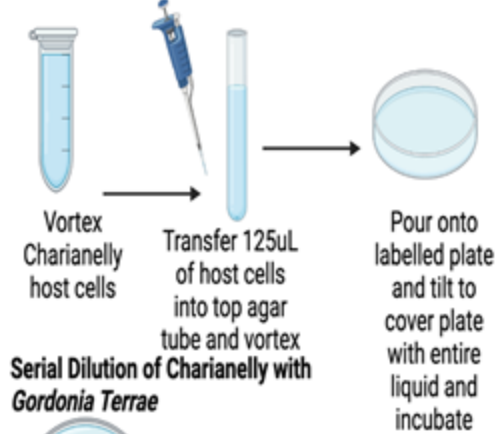
- The minor tail protein aids the phage identify in the correct host, and then break the surface so the DNA can be injected.
- Host range can show the dynamic of the survival of pathogens which can be a leading factor in understanding evolution patterns



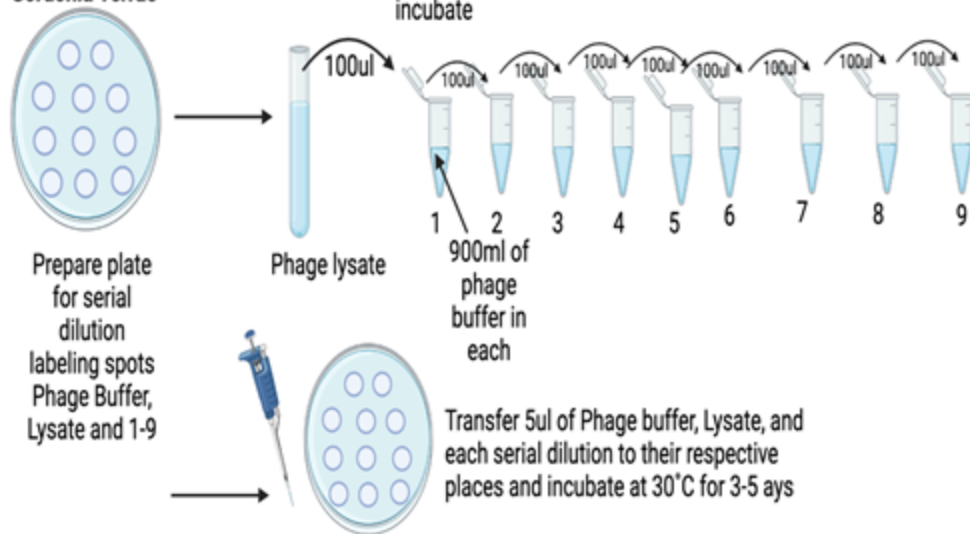
The host range project conducted consisted of 6 different phages each tested on three different strains of *G. terrae*. Our group used the phage Capybara and tested the infection patterns on *G. lacunae*, *G. westfalia*, and *G. rubripertincta*. We hypothesized that all the phages would have similar infection patterns on the different hosts

Figure 12: Phage N15

Titer Plate of Charianelly with *Gordonia Terrae*



Serial Dilution of Charianelly with *Gordonia Terrae*



Titer Plate of Capybara with various host cells

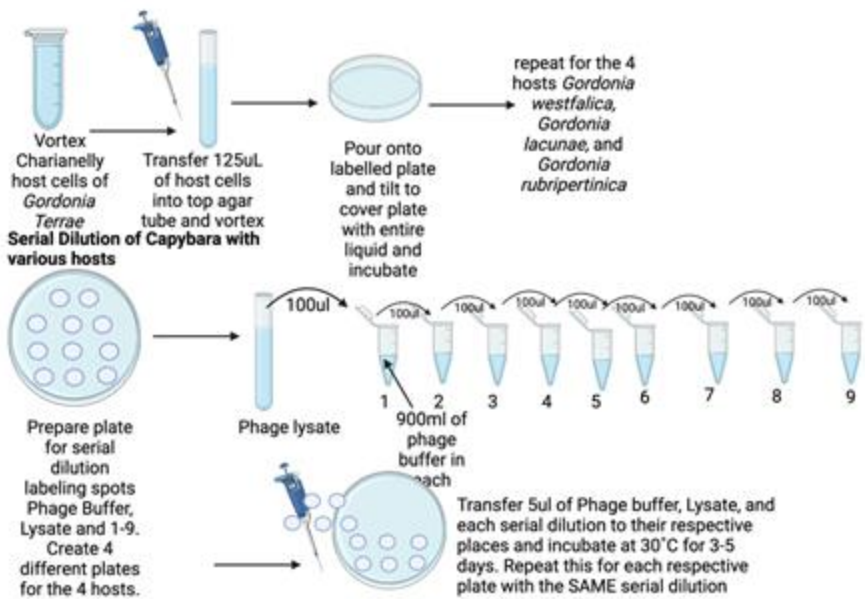


Figure 16: Methods Flow Diagram for Phages Charianelly and Capybara

Phage capybara was able to be infected by all host other than *G. lacunae*

- Tested the Capybara phage in the DJ cluster
- Phage capybara infected all hosts other than *G. lacunae*, with roughly the same efficiency between the other three
- Similar infection pattern with all other phages except belphegor where *G. lacunae* was not able to infect
- Phage capybara shares no other minor pham tails with any other Pitt F23 pages

The significance of these findings is that there is no similarity between the other Pitt F23 phages and that 3 out of the 4 hosts were able to infect



G. terrae



G. westfalia



G. lacunae



G. rubripertincta

Figure 17: Serial dilution with alternate hosts using phage Capybara

Conclusions

- Charianelly cluster: CS
- Subcluster: CS3, 24 phages in this subcluster (large)
- GC content: 59.1%
- Total Genes Predicted: 93 (Bianmat and Minos 94)
- **Similar with other Gordonia phages because minor tail phams match up**

Mechanism: If there is not an appropriate host identity, then the minor tail protein cannot break the surface and no DNA can be injected leading to no injection.

- Annotated 15 genes
 - Started changed:
 - Functions assigned: 3
- 2 minor tail proteins on 30 and 31 for Capybara but did not match up with Charianelly so cannot compare infection patterns
- Charianelly had a broader host range; Capybara infection pattern different
- **Minor Tail proteins:** aids the phage identify in the correct host, and then break the surface so the DNA can be injected.

Author Contributions

Simran: completed slides 4, 6, 9, 10, 15, 18

Ayaan: completed slides 3, 7, 8, 12, 13, 18

Reagan: completed slides 5, 11, 14, 16, 17, 18

We all practice our slides and reviewed the groups slides together