Abdulrahman Alazemi

Final project

BNFO301

Clustered repeats and potential regulatory sequences

Introduction;

Regulatory sequences is a segment of a nucleic acid molecule which is capable of increasing and decreasing the expression of a specific genes within an organism(4,6). Regulatory regions can be found in conserved non-coding sequences(4,6). Conserved non-coding sequences are associated with transcription factors binding sites(4,6). Transcription factors are proteins that bind to a specific DNA sequence. Transcription factors does many thing. For example, They are called activators when they activate transcription by facilitating the productive formation between the RNA polymerase and the promoter(7). In the other hand, they are called repressor when they inhibit the transcription by preventing RNA polymerase from forming the productive complex with the promoter(7).

Applying a known method with a known results to other cases of interest looking for similarity and generality between results is a well known approach in research. Indeed, in their paper " Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies", Van Helden J, André B, and Collado-Vides method was based on detection of over represented oligonucleotide and estimate their statistical differences. Then they detect an already known upstream sequence case with a high statistical significance, then they said that unknown motifs with high statistical significance upstream sequences might be a good candidate of a new putative regulatory sequence. After reading their paper which inspired me a lot. I started to ask myself the following question "Can I apply a known method/tool with a known results to other phages and get the same/similar result?", so I started searching and reading about methods. Then, I found the following method “Examine the proven Repressor and Cro binding sites (operators) of Phage Lambda” in Jeff Elhai's website(2). After being introduced to this method, the question I had and the focus of my project became “Would I find some sort of generality between operators of different phages?”.

Method;

To be able to do the unknown, I started with the known case trying to master it. I went to biobike which is a web-based environment enabling biologists with little programming expertise to combine tools, data, and knowledge(3), knowing the 3 operators of phage Lambda, and those operators are OR1 TATCACCGCCAGAGGTA, OR2 TAACACCGTGCGTGTTG, and OR3 TATCACCGCAAGGGATA. The first thing I did is finding phage lambda in biobike and used the following command (ORGANISM/S-NAMED "Lambda"), and I found phage Lambda name which is Escherichia-phage-lambda. The second thing I did is motifs-in. Figure1 shows how I called motifs-in in biobike with all my values and variables. Motifs in shows the most common repeated sequences that might be a good candidates to become operators in their gene inside the genome.

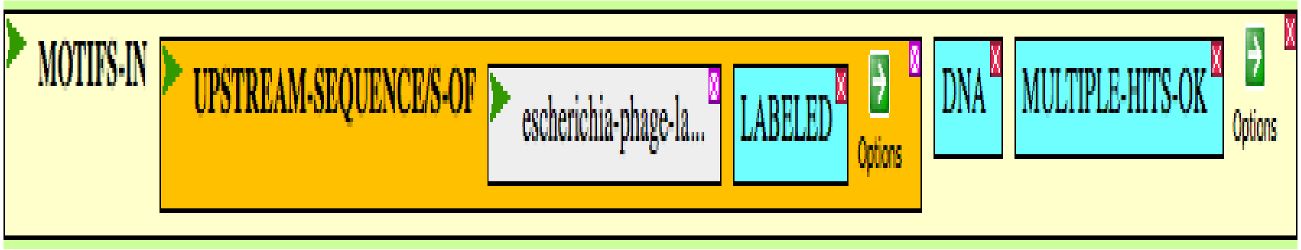
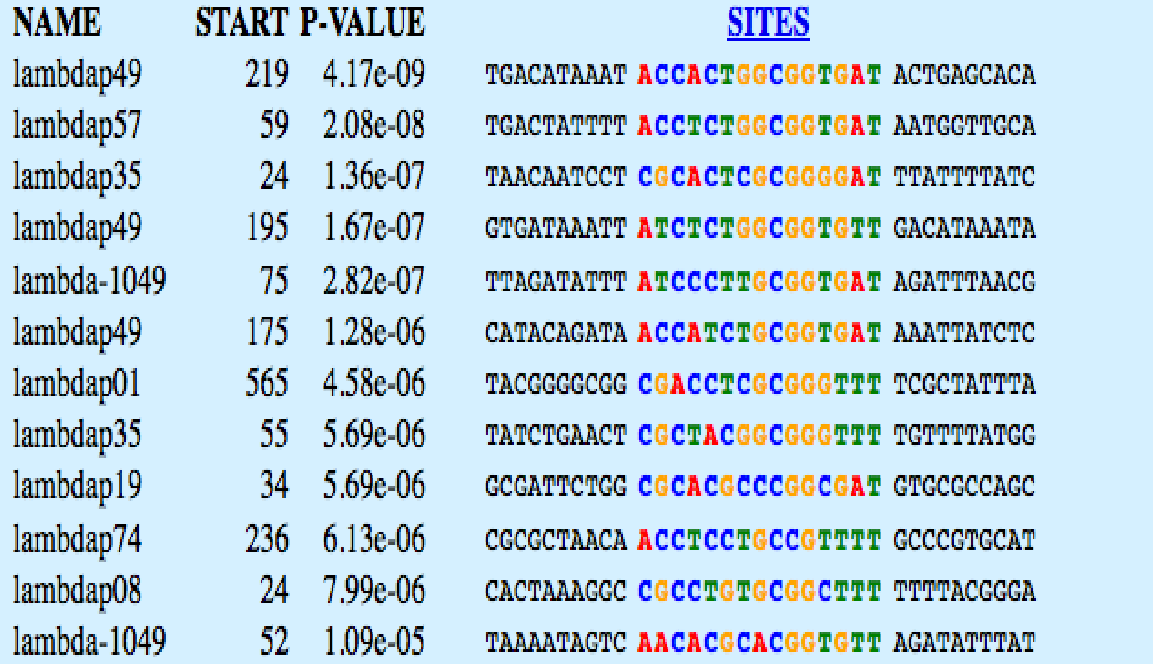


Figure1. Motifs-in function in biobike.

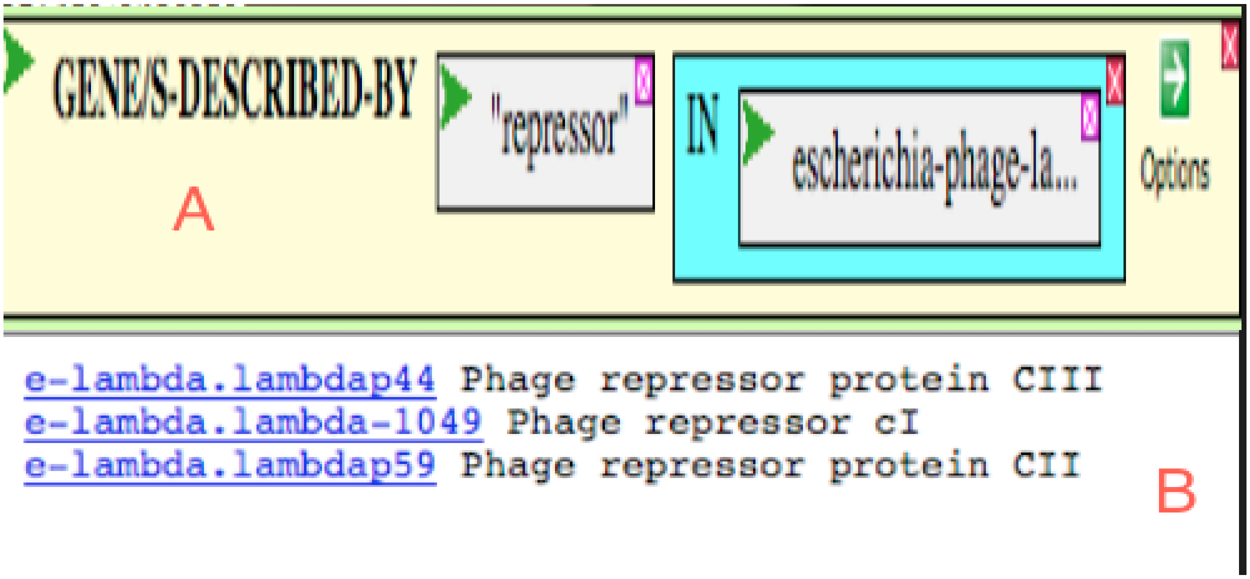
After calling the motifs-in, a new page started that has all the information and the outcome of the function motifs-in. Figure 2 shows the outcome. There are some genes mentioned more than once that has the repeated sequence . For example, the gene lambdap49 is mentioned 3 times which means that the repeated sequence exists 3 times inside lambdap49 which is interesting. On the other hand, Lambda-1049 is mentioned twice and this is interesting too, and now we don't know which gene has the operator and which sequence might be the operator for the repressors because both genes are a good candidates to contain the operators. So something has to be done to solve this!.



NAME	START	P-VALUE	SITES
lambdap49	219	4.17e-09	TGACATAAAT ACCACTGGCGGTGAT ACTGAGCACA
lambdap57	59	2.08e-08	TGACTATTTT ACCTCTGGCGGTGAT AATGGTTGCA
lambdap35	24	1.36e-07	TAACAATCCT CGCACTCGCGGGGAT TTATTTTATC
lambdap49	195	1.67e-07	GTGATAAATT ATCTCTGGCGGTGTT GACATAAATA
lambda-1049	75	2.82e-07	TTAGATATTT ATCCCTTGCGGTGAT AGATTTAACG
lambdap49	175	1.28e-06	CATACAGATA ACCATCTGCGGTGAT AAATTATCTC
lambdap01	565	4.58e-06	TACGGGGCGG CGACCTCGCGGGTTT TCGCTATTTA
lambdap35	55	5.69e-06	TATCTGAACT CGCTACGGCGGGTTT TGTTTTATGG
lambdap19	34	5.69e-06	GCGATTCTGG CGCACGCCCGGCGAT GTGCGCCAGC
lambdap74	236	6.13e-06	CGCGCTAACA ACCTCCTGCCGTTTT GCCCGTGCAT
lambdap08	24	7.99e-06	CACTAAAGGC CGCCTGTGCGGCTTT TTTTACGGGA
lambda-1049	52	1.09e-05	TAAAATAGTC AACACGCACGGTGTT AGATATTTAT

Figure 2. The outcome of the function motifs-in in biobike.

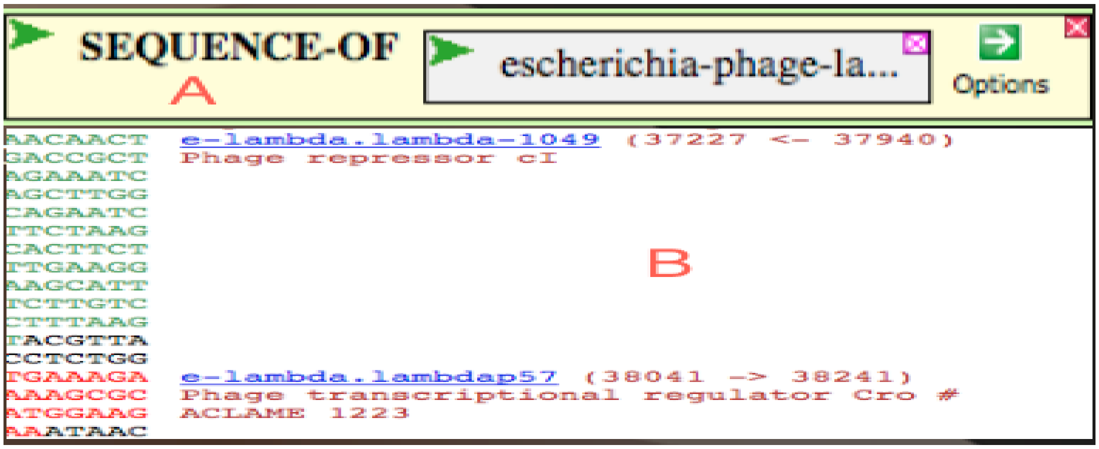
I went back to biobike to call a function that might be helpful in this case, and might give us a clue to, which one of these two genes is the one that has the operators and the function might give us an idea about where to look. This function is called “gene/s-descried-by”. Figure 3A shows how to call the gene/s-descried-by function. I put “repressor” because I wanted the function to look for genes that has repressors in phage Lambda for me. Figure 3B shows the outcome of the function gene/s-descried-by. It mentioned 3 genes that have repressors and those 3 genes are lambdap44, lambdap59, and lambda-1049. Now that we know which genes has repressors, we go back to motifs-in and look for repeated sequences in these specific genes because there is a big chance that those repeated sequences are the operators.



```
e-lambda.lambdap44 Phage repressor protein CIII
e-lambda.lambda-1049 Phage repressor cI
e-lambda.lambdap59 Phage repressor protein CII
```

Figure3A. gene/s-descried-by function in biobike, figure 3B is the outcome of the gene/s-descried-by function.

After going back to motifs-in outcome, we look for repeated sequences for the genes that have repressors. In figure 2, there is no mention for the gene lambda59 and lambdap44. On the other hand, lambda-1049 is mentioned twice. After picking lambda-1049 as the biggest candidate of having the operators, we want to check and see the genes surrounding lambda-1049. Figure 4 A shows a biobike function that helps us view the sequences of each gene in a specific phage or bacterium. Figure 4 B is the outcome of the function “sequence-of”, and the outcome is called sequence-viewer. You can see in figure 4 B, the gene lambda-1049 is next to the gene lambdap57. And they are going on opposite direction of each other and this is interesting. Now, we want to find the operators and their position inside the gene.

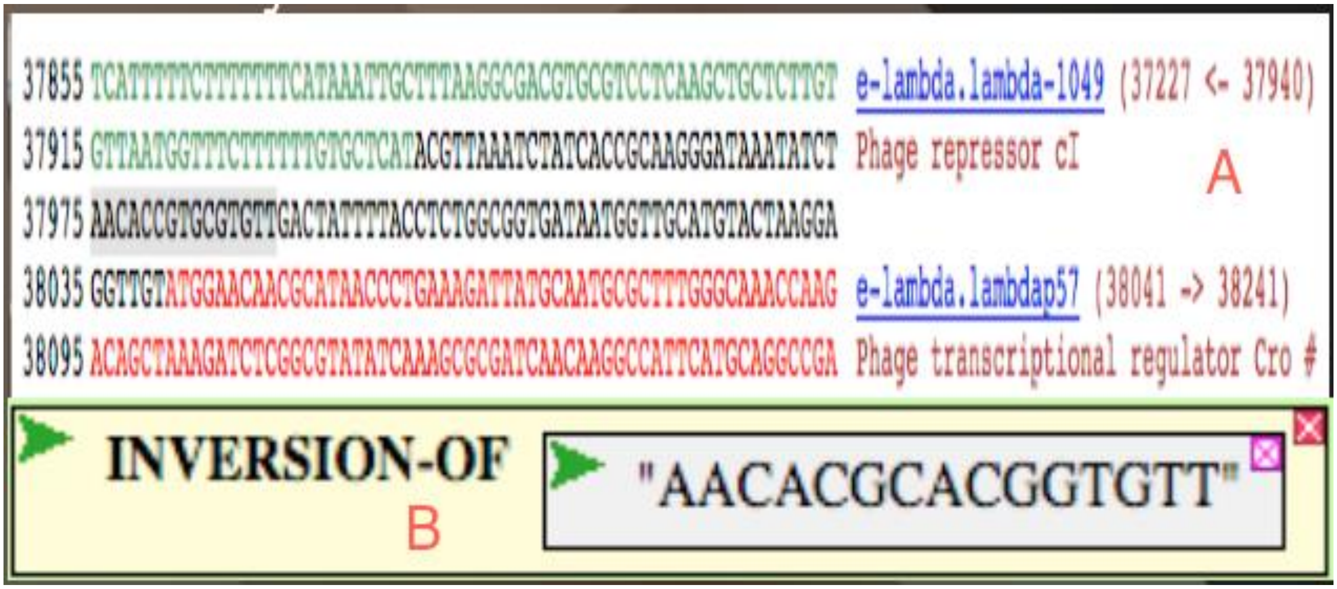


SEQUENCE-OF escherichia-phage-la...

```
AACAACT   e-lambda.lambda-1049 (37227 <- 37940)
GACCGCT   Phage repressor cI
AGAAATC
AGCTTGG
CAGAATC
TTCTAAG
CACTTCT
TTGAAGG
AAGCATT
TCTTGTC
CTTTAAG
TACGTTA
CCTCTGG
TGAAAGA   e-lambda.lambdap57 (38041 -> 38241)
AAAGCGC   Phage transcriptional regulator Cro #
ATGGAAG   ACLAME 1223
AAATAAC
```

Figure4A. The function sequence-of in biobike, figure4B the sequence-viewer which is the outcome of the function sequence-of.

We go back to motifs-in in figure 2 for the last time and pick the repeated sequences of gene lambda-1049 in the first place then pick the repeated sequences of lambdap57. So we pick ATCCCTTGCGGTGAT from lambda-1049 , AACACGCACGGTGTT from lambda-1049, and ACCTCTGGCGGTGAT from lambdap57. Now we go to the sequence-viewer showed in figure 5 A to find these sequences. I found the sequences of lambda57 directly in the sequence-viewer. When I tried to search for the sequences of lambda-1049, I didn't find them directly, so I used a function of biobike called “Inversion-of” showed in figure 5 B to have the complementary sequences of these sequences. Then, I went back and searched for them and I found them directly as it is showing in figure 5 A in grey after using the “Inversion-of”.



37855 TCATTTTTCTTTTTTTCATAAATTGCTTTAAGGCGACGTGCGTCCTCAAGCTGCTCTTGT e-lambda.lambda-1049 (37227 <- 37940)
37915 GTTAATGGTTTCTTTTTTGTGCTCATACGTTAAATCTATCACCGCAAGGGATAAATATCT Phage repressor cI
37975 AACACCGTGCGTGTTGACTATTTTACCTCTGGCGGTGATAATGGTTGCATGTACTAAGGA
38035 GGTTGTATGGAACAACGCATAACCCTGAAAGATTATGCAATGCGCTTTGGGCAAACCAAG e-lambda.lambdap57 (38041 -> 38241)
38095 ACAGCTAAAGATCTCGGCGTATATCAAAGCGCGATCAACAAGGCCATTCATGCAGGCCGA Phage transcriptional regulator Cro #

INVERSION-OF "AACACGCACGGTGTT"

Figure5A. Is the sequence-viewer of the intergenic sequence between gene lambda-1049 and lambdap57, figure 5B is a function of biobike called “Inversion-of”.

In figure 6, is the genes map of phage lambda, you can see the gene lambda-1049 in blue going the opposite way of the gene lambdap57 in red and the area of interest that has the operators is what between them.

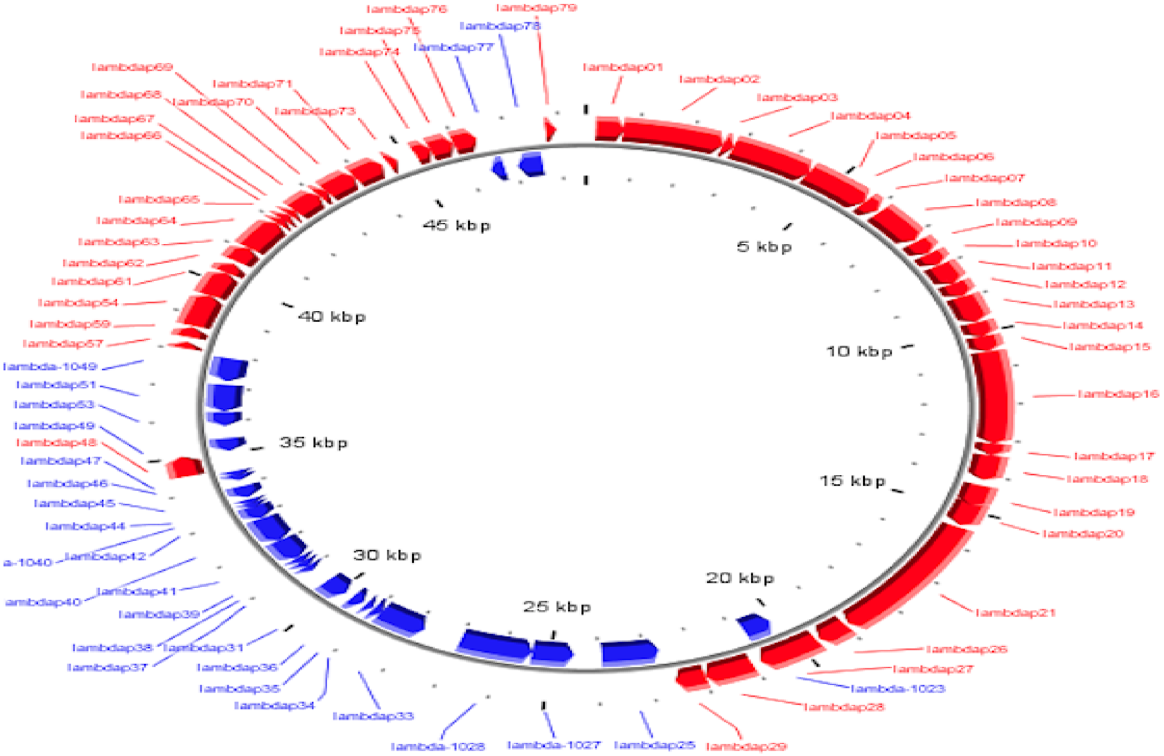


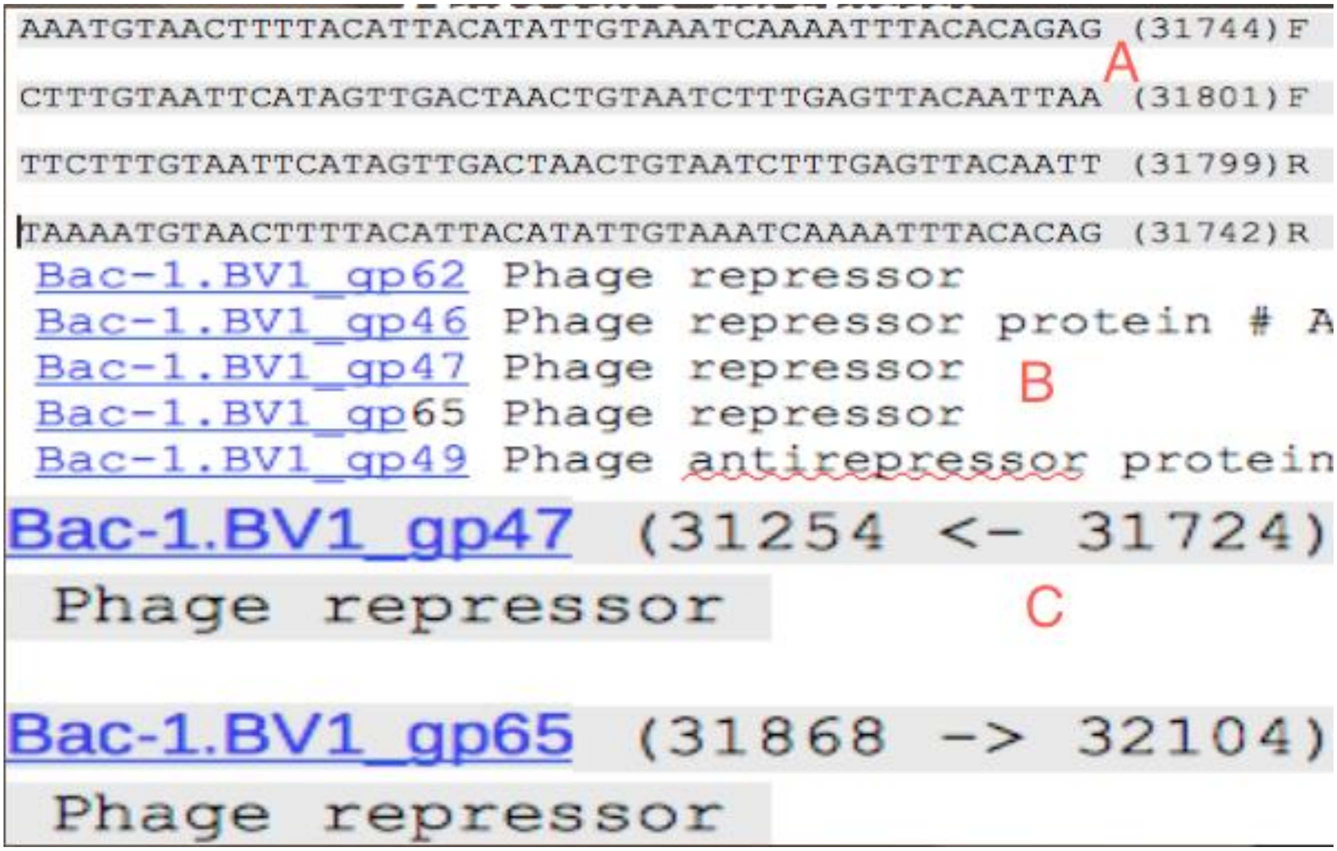
Figure6. The genes map of phage lambda.

After having an experience with the known case, I did the exact same steps and process with 21 different phages of different families. Those phages are Acinetobacter-phage-Ac42, Aeromonas-phage-25, Bacillus-phage-1, Burkholderia-phage-Bcep1, Chlamydia-phage-2, Clostridium-phage-39-O, Enterobacteria-phage-13a, Escherichia-phage-933W, Klebsiella-phage-K11, Lactobacillus-johnsonii-prophage-Lj928, Lactococcus-I34phage-bIL170, Listeria-phage-A006, Mycobacterium-phage-244, Pseudomonas-phage-119, Salmonella-phage-c341, Staphylococcus-phage-187, Streptococcus-phage-2972, Synechococcus-phage-S-PM2, Vibrio-phage-K139, Xanthomonas-phage-Cf1c, and Yersinia-phage-Berlin.

Results;

In order for me to be able to find generality between operators of different phages, I had done all the steps and process of phage lambda above to 21 different phage. But, the results weren't what I expected. For example, eight phages out of the 21 phages don't have repressors, and I eliminate them from the study. Those phages are Chlamydia-phage-2, Enterobacteria-phage-13a, Klebsiella-phage-K11, Lactococcus-I34phage-bIL170, Mycobacterium-phage-244, Pseudomonas-phage-119, Xanthomonas-phage-Cf1c, and Yersinia-phage-Berlin. In addition to that, two phages of the 21 didn't give me a map of their genes for the following reason “ Can't at present display maps of linear replicons”. And those phages are Aeromonas-phage-25 and Staphylococcus-phage-187. So I've left with 11 phages to work with. Out of those 11 phages, just 3 phages have similarity to phage lambda. And those phage are Bacillus-phage-1, Listeria-phage-A006, and Lactobacillus-johnsonii-prophage-Lj928.

Bacillus-phage-1 is the first similar phage to phage lambda. It similar to phage lambda because it went through all the process and steps that phage lambda went through, and it gave me similar outcome. For example, as you can see in figure 7 A, those are the most common repeated sequences of Bacillus-phage-1from motifs-in. Then, you can see in figure 7 B what genes has repressors in Bacillus-phage-1. And you can see in figure 7 C the two genes that has the intergenic sequences that have the opreators in Bacillus-phage-1, are going on opposite side of each other as in phage lambda and you can compare it with figure 4 B of phage lambda. I tested the sequences and check them in sequence-viewer and they are similar to operators of phage lambda. Anyone can go back to the method and do all the steps that I did with phage lambda to Bacillus-phage-1 and he/she will get similar outcome.



```
AAATGTAACTTTTACATTACATATTGTAAATCAAAATTTACACAGAG (31744)F
CTTTGTAATTCATAGTTGACTAACTGTAATCTTTGAGTTACAATTAA (31801)F
TTCTTTGTAATTCATAGTTGACTAACTGTAATCTTTGAGTTACAATT (31799)R
TAAAATGTAACTTTTACATTACATATTGTAAATCAAAATTTACACAG (31742)R
```

```
Bac-1.BV1_gp62 Phage repressor
Bac-1.BV1_gp46 Phage repressor protein # A
Bac-1.BV1_gp47 Phage repressor
Bac-1.BV1_gp65 Phage repressor
Bac-1.BV1_gp49 Phage antirepressor protein
```

Bac-1.BV1_gp47 (31254 <- 31724)

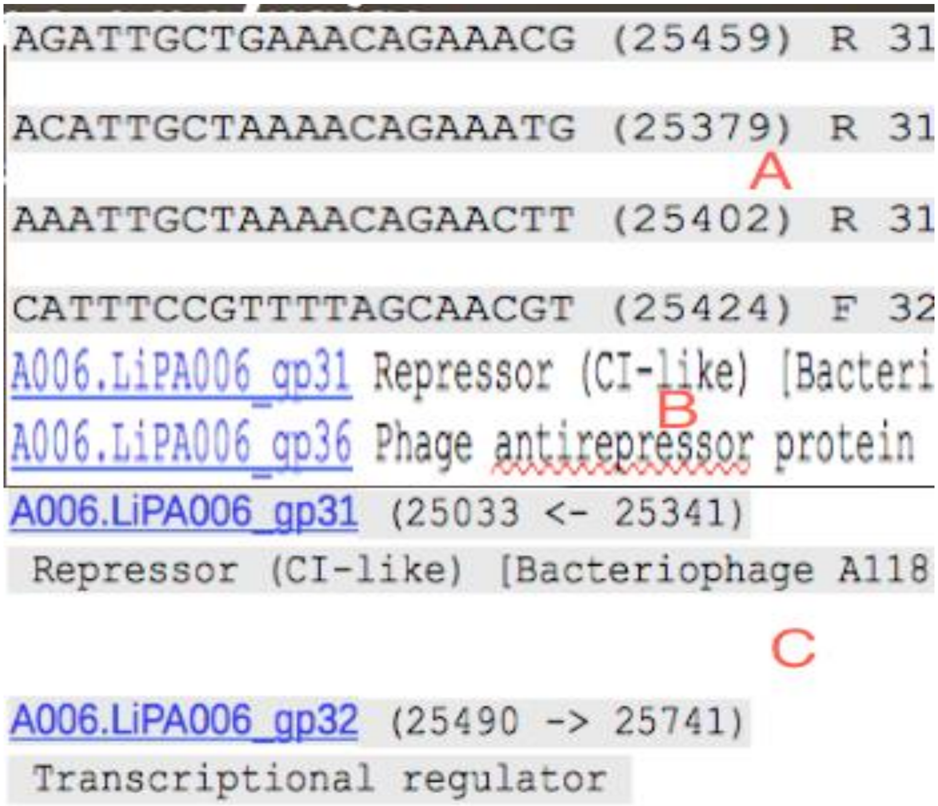
Phage repressor

Bac-1.BV1_gp65 (31868 -> 32104)

Phage repressor

Figure7A. The most repeated sequences in motifs-in of Bacillus-phage-1, figure7B is the genes of Bacillus-phage-1 that have repressors, figure 7C is the genes that have repressors and their coordinates in sequence-viewer of biobike.

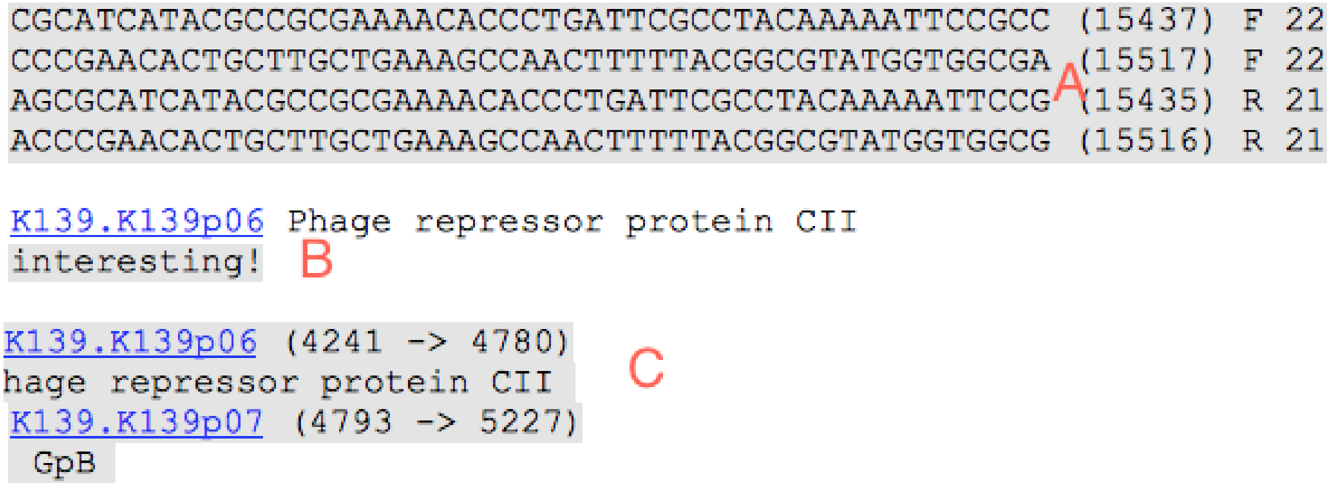
The second similar phage to lambda is Listeria-phage-A006. What I said in the paragraph above, is the same with Listeria-phage-A006 . Listeria-phage-A006 similar to phage lambda because it went through all the process and steps that phage lambda went through, and it gave me similar outcome. For example, as you can see in figure 8 A, those are the most common repeated sequences of Listeria-phage-A006 from motifs-in. Then, you can see in figure 8 B what genes has repressors in Listeria-phage-A006. And you can see in figure 8 C the two genes that has the intergenic sequences that have the opreators in Bacillus-phage-1, are going on opposite side of each other as in phage lambda and you can compare it with figure 4 B of phage lambda. I tested the sequences and check them in sequence-viewer and they are similar to operators of phage lambda.



```
AGATTGCTGAAACAGAAACG (25459) R 31

ACATTGCTAAAACAGAAATG (25379) R 31

AAATTGCTAAAACAGAACTT (25402) R 31

CATTTCCGTTTTAGCAACGT (25424) F 32
A006.LiPA006_gp31 Repressor (CI-like) [Bacteri
A006.LiPA006_gp36 Phage antirepressor protein
A006.LiPA006_gp31 (25033 <- 25341)
 Repressor (CI-like) [Bacteriophage A118

A006.LiPA006_gp32 (25490 -> 25741)
 Transcriptional regulator
```

Figure8A. The most repeated sequences in motifs-in of Listeria-phage-A006, figure 8B is the genes of Listeria-phage-A006 that have repressors, figure 8 C is the genes that have repressors and their coordinates in sequence-viewer of biobike.

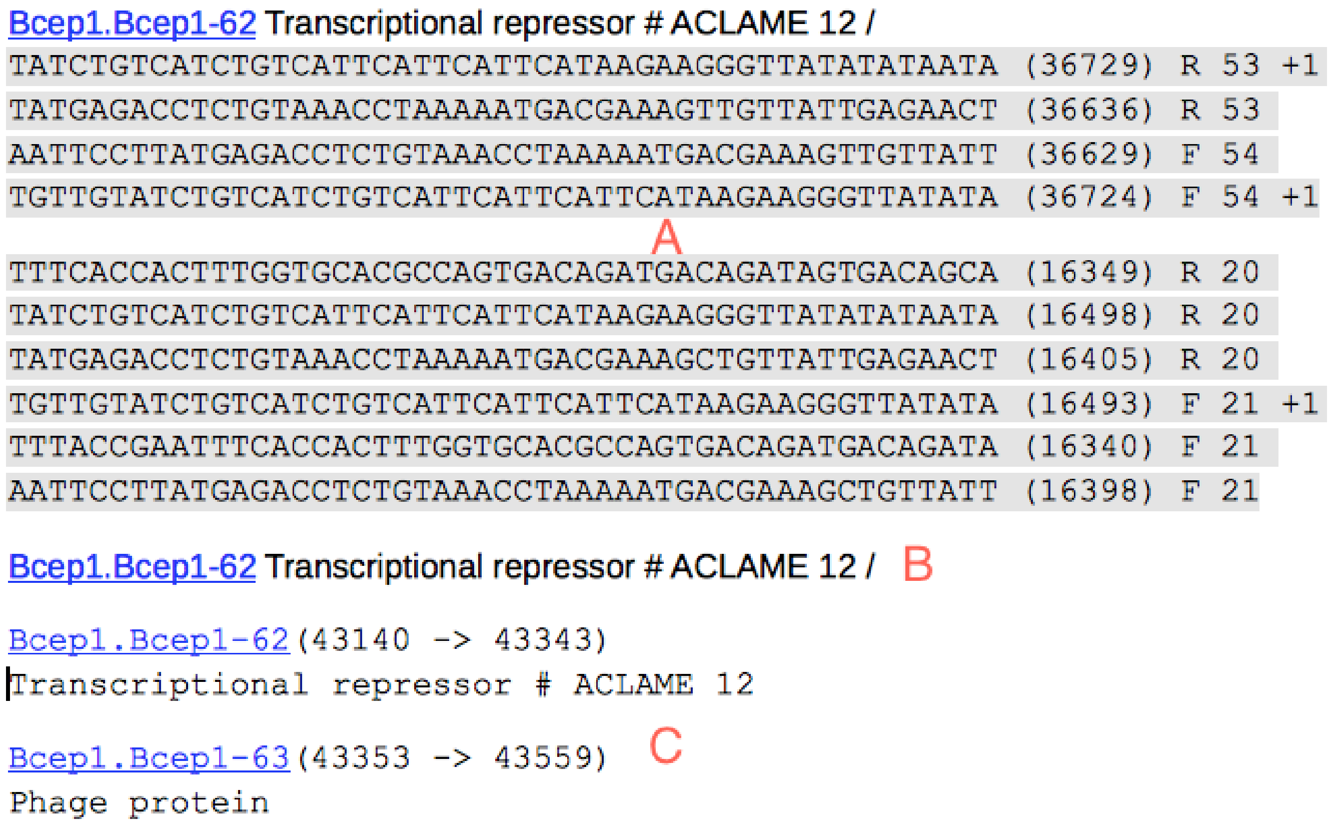
As I gave two examples for the operators of two different phages that I believe they are similar to the operators of phage lambda, I will give and show two other phages that I don't believe they have any generality with phage lambda. The first phage is Vibrio-phage-K139. So I picked the most repeated sequence of the function motifs-in. And they all lie on the gene K139.K139p21 and K139.K139p22 as you can see in figure 9 A. Then I went to biobike to find what are the genes that have repressors, I found one gene which is K139.K139p06 and you can see that in figure 9 B. There is more than one problem in here. One of which, is that none of the repeated sequences is in gene K139.K139p06 which means that those repeated sequences are not operators that a repressor can bind to. One other problem is when you take a look at figure 9 C, you see that the gene K139.K139p06 and the surrounding genes of it, are moving or going in the same direction and that made me confused because when I saw the map of Vibrio-phage-K139, I didn't see any blue genes either front or behind K139.K139p06. In contrast, I saw the K139.K139p21 and K139.K139p22 are going to the opposite direction of each other. So I believe that those sequences act and seem like a operators. But, they are not mentioned any where as a repressor. If I want to conclude about this genes and sequences, I would say that this exact phage Vibrio-phage-K139 and its following genes K139.K139p21, K139.K139p22, and K139.K139p06 need to be studied more and more.



```
CGCATCATACGCCGCGAAAACACCCTGATTCGCCTACAAAAATTCCGCC (15437) F 22
CCCGAACACTGCTTGCTGAAAGCCAACTTTTTACGGCGTATGGTGGCGA (15517) F 22
AGCGCATCATACGCCGCGAAAACACCCTGATTCGCCTACAAAAATTCCG (15435) R 21
ACCCGAACACTGCTTGCTGAAAGCCAACTTTTTACGGCGTATGGTGGCG (15516) R 21

K139.K139p06 Phage repressor protein CII
interesting!

K139.K139p06 (4241 -> 4780)
hage repressor protein CII
 K139.K139p07 (4793 -> 5227)
  GpB
```

Figure9A.The most repeated sequences in motifs-in of Burkholderia-phage-Bcep1, figure 9B is the genes of Burkholderia-phage-Bcep1 that have repressors, figure 9 C is the genes that have repressors and their coordinates in sequence-viewer of biobike.

The second phage is Burkholderia-phage-Bcep1. So I picked the most repeated sequence of the function motifs-in. And they all lie on two different places in the genome, the first place is between the gene Bcep.Bcep1-53 and Bcep.Bcep1-54, and the second place is between the gene Bcep.Bcep1-20 and the gene Bcep.Bcep1-21 as you can see in figure 10 A. Then I went to biobike to find what are the genes that have repressors, I found one gene which is Bcep.Bcep1-62 and you can see that in figure 10 B. There is more than one problem in here. One of which, is that none of the repeated sequences is in gene Bcep.Bcep1-62 which means that those repeated sequences are not operators that a repressor can bind to. One other problem is when you take a look at figure 10 C, you see that the gene Bcep.Bcep1-62 and the surrounding genes of it, are moving or going in the same direction and that made me confused because when I saw the map of Burkholderia-phage-Bcep1, I didn't see any blue genes either front or behind Bcep.Bcep1-62. In contrast, I saw the Bcep.Bcep1-20 and Bcep.Bcep1-21 in one place and Bcep.Bcep1-20 and Bcep.Bcep1-21 in another place are going to the opposite direction of each other. So I believe that those sequences act and seem like a operators. But, they are not mentioned any where as a repressor. If I want to conclude about this genes and sequences, I would say that this exact phage Burkholderia-phage-Bcep1 and its following genes , Bcep.Bcep1-53, Bcep.Bcep1-54, Bcep.Bcep1-20, and Bcep.Bcep1-21 need to be studied more and more.

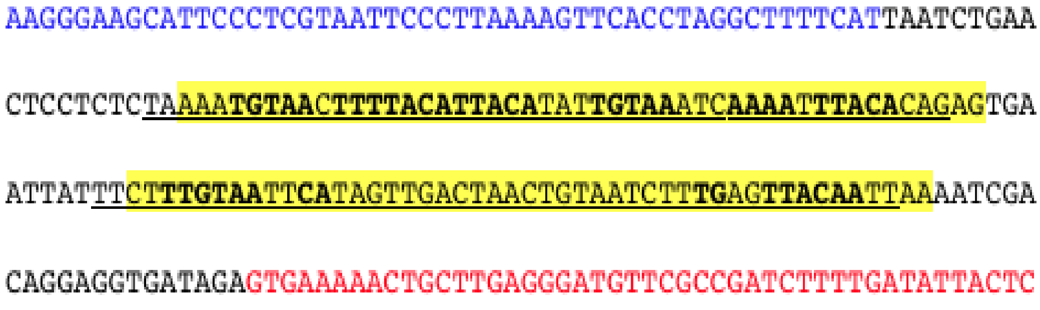


```
Bcep1.Bcep1-62 Transcriptional repressor # ACLAME 12 /
TATCTGTCATCTGTCATTCATTCATTCATAAGAAGGGTTATATATAATA (36729) R 53 +1
TATGAGACCTCTGTAAACCTAAAAATGACGAAAGTTGTTATTGAGAACT (36636) R 53
AATTCCTTATGAGACCTCTGTAAACCTAAAAATGACGAAAGTTGTTATT (36629) F 54
TGTTGTATCTGTCATCTGTCATTCATTCATTCATAAGAAGGGTTATATA (36724) F 54 +1

TTTCACCACTTTGGTGCACGCCAGTGACAGATGACAGATAGTGACAGCA (16349) R 20
TATCTGTCATCTGTCATTCATTCATTCATAAGAAGGGTTATATATAATA (16498) R 20
TATGAGACCTCTGTAAACCTAAAAATGACGAAAGCTGTTATTGAGAACT (16405) R 20
TGTTGTATCTGTCATCTGTCATTCATTCATTCATAAGAAGGGTTATATA (16493) F 21 +1
TTTACCGAATTTCACCACTTTGGTGCACGCCAGTGACAGATGACAGATA (16340) F 21
AATTCCTTATGAGACCTCTGTAAACCTAAAAATGACGAAAGCTGTTATT (16398) F 21

Bcep1.Bcep1-62 Transcriptional repressor # ACLAME 12 /

Bcep1.Bcep1-62(43140 -> 43343)
Transcriptional repressor # ACLAME 12

Bcep1.Bcep1-63(43353 -> 43559)
Phage protein
```

Figure10A.The most repeated sequences in motifs-in of Vibrio-phage-K139, figure 10B is the genes of Vibrio-phage-K139 that have repressors, figure 10 C is the genes that have repressors and their coordinates in sequence-viewer of biobike.

The operators of three of the 11 phages are similar to the operators of phage lambda and that is 27.2%. Phage lambda is not even similar to 50% of these phages. Unfortunately, looking for generality between operators that the operators binds to between different phages from different phage families, reached a dead end. And the best way to conclude this experiment is that phage Lambda, Bacillus-phage-1, Listeria-phage-A006 , and Lactobacillus-johnsonii-prophage-Lj928 have a similarity/generality between their operators that the repressors bind to.

When I was about to conclude my research, I saw some thing interesting in one phage among the phages I did my research with. The phages is Bacillus-phage-1 and the interesting thing is that the operators of phage Bacillus-phage-1 that the repressors binds to, are four as you can see in figure 7 A. But when I went to sequences-viewer on the phage Bacillus-phage-1 in biobike, I found that those four operators are actually two operators because in sequence-viewer, they lay in just two identical sequences and they might be internally palindromic sequences. The following figure, figure 11 shows what I mean. The two highlighted in yellow sequences are where the 4 operators lay. And the bases in bold shows the palindromic part.



AAGGGAAGCATTCCCTCGTAATTCCCTTAAAAGTTCACCTAGGCTTTTCATTAATCTGAA

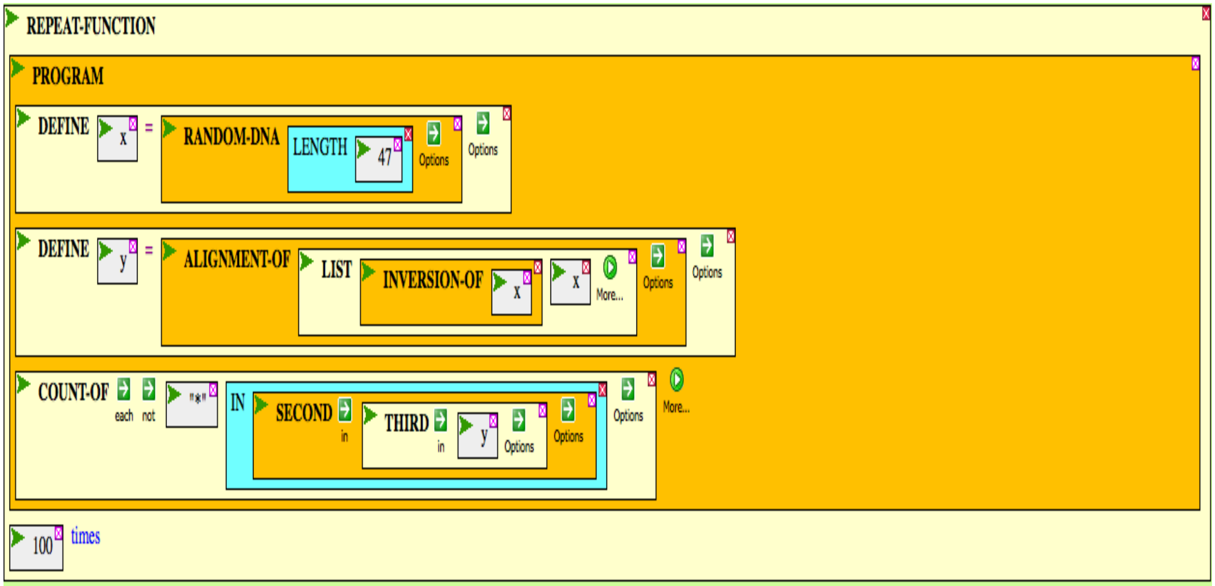
CTCCTCTCTAAAA**TGTAA**C**TTTTACATTACA**TAT**TGTAA**ATC**AAAATTTACA**CAGAGTGA

ATTATTTCT**TTGTAA**TT**CA**TAGTTGACTAACTGTAATCTT**TG**AG**TTACAA**TTAAAATCGA

CAGGAGGTGATAGAGTGAAAAACTGCTTGAGGGATGTTCGCCGATCTTTTGATATTACTC

Figure11. A part of sequence-viewer of the area of interest of Bacillus-phage-1.

What I did next is that I went to biobike to check wither what happening with these sequences is random or not. I did a program in a repeated-function function of biobike to check the palindromic sequence of 100 different random DNA sequences of the length of 47 nucleotides as shown in figure 12. But first I did it with only the sequence of interest which is the following sequence “AAATGTAACTTTTACATTACATATTGTAAATCAAAATTTACACAGAG”. And I got 28 matches of the complementary sequences with the original sequence. Then I did the function with the random 100 DNA sequences that have a length of 47 nucleotide. And the I found 6 sequences that have 26 matches of their complementary sequences to their original sequences. And I found many that have 22, 24 , and 20 matches. And I found one that has 34 matches and another one that has 30. So it seems that the sequence of interest is more random than not random or to be honest the function I did and the program didn't answer or didn't get me what I was looking for and I wasn't satisfied the out come that I got. So I decided to do things by hand.



```
REPEAT-FUNCTION
  PROGRAM
    DEFINE x = RANDOM-DNA LENGTH 47 Options Options
    DEFINE y = ALIGNMENT-OF LIST INVERSION-OF x x More... Options Options
    COUNT-OF each not "*" IN SECOND in THIRD in y Options Options Options More...
  100 times
```

Figure12. Repeat function on biobike and the values and variables is to check the palindromic sequence of 100 different random DNA sequences of the length of 47.

After deciding to do it by hand, I grabbed a piece of paper and a pin and draw whats on figure13 with a help from Jeff Elhai. I discovered that this sequence is partially palindromic. But the way of it being partially palindromic is strange and interesting. As you can see in figure 13, A is palindromic to A in one side and I palindromic to the other side of the sequence. In addition to that, B is palindromic to the B of the other side of the sequence. It is interesting and confusing at the same time.

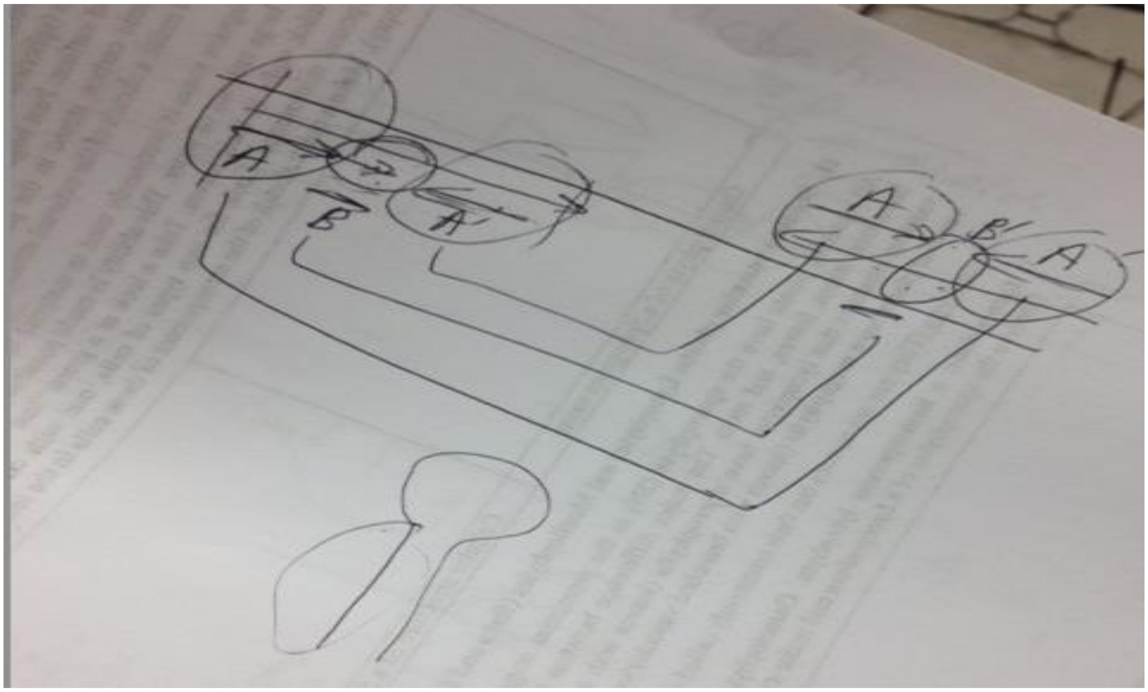


Figure13. A hand written explanation of the area and sequence of interest.

At the end, the only reasonable sense I tried to make of these 2 or 4 sequences is that instead of being 2 long sequences that are mentioned 4 times, you might cut the sequence in half and get 4 short sequences that are similar to each other because of the fact that the two sequences are partially palindromic. In other word and in more details, the sequence instead of being long as the following “AAATGTAACTTTTACATTACATATTGTAAATCAAAATTTACACAGAG”, it can be divided to the following two sequences “TGTAAATCAAAATTTACA” and “TGTAACTTTTACATTACA”.

In conclusion I wish I can keep going with these sequences and try to make sense of them but I cant because of other projects that i'm involved in. These sequences should be studied more and more because they have the potential to be something huge and important.

References:

1- "BioBIKE Portal." *BioBIKE Portal*. N.p., n.d. Web. 09 May 2014.

2- "BNFO 301 Research Group: Clustered Repeats & Regulatory Sequences." *BNFO 301 Research Group: Clustered Repeats & Regulatory Sequences*. N.p., n.d. Web. 09 May 2014.

3- Elhai, J., A. Taton, J. Massar, J. K. Myers, M. Travers, J. Casey, M. Slupesky, and J. Shrager. "BioBIKE: A Web-based, Programmable, Integrated Biological Knowledge Base." *Nucleic Acids Research* 37.Web Server (2009): W28-32. Print.

4- Hardison, R. "Conserved Noncoding Sequences Are Reliable Guides to Regulatory Elements." *Trends in Genetics* 16.9 (2000): 369-72. Print.

5- Helden, J. Van, B. André, and J. Collado-Vides. "Extracting Regulatory Sites from the Upstream Region of Yeast Genes by Computational Analysis of Oligonucleotide Frequencies." *Journal of Molecular Biology* 281.5 (1998): 827-42. Print.

6- J.-J.M. Riethoven "Regulatory regions in DNA: promoters, enhancers, silencers, and insulators" Methods Mol. Biol., 674 (2010), pp. 33–42.

7- Lodish, Harvey F. *Lecture Notebook for Molecular Cell Biology*. New York: W.H. Freeman, 2000. Print.