Viral Genome/Metagenome Project

Michael Kassa

05/08/2009

This project begins with my self thinking journey that makes me to think of the origins of life. I always think that life (any living things) begins with a creator which is God. However, science has different explanation for the origin of life. I decided to take a part on science, to explore the relationship between two things. A great opportunity was in front of me, taking Bioinformatics 301, with the help of my instructor I decided to make an exploration that was not done before analyzing Octopus Hot Springs metagenome or Bear paw Hot Springs metagenome for the first time.

1. Purpose:

Acquire a set of reads, segments of nucleotides, which are selected randomly from Octopus Hot Springs metagenome or Bear paw Hot Springs metagenome. Using a ViroBike, Gene Mark and NCBI as a major tool construct large enough extended contigs which can be compared to all known viruses and organism under Gen Bank. To determine for the first time if whether the sequences contain genes? To predict to the extent possible the function of any possible encoded protein and relation to other known organisms.

1. Obtaining a sequence from a metagenome.

From ViroBike :

Step: 1 claiming a read

(ENTER: VIRAL-METAGENOME)

(CLAIM-READ-FOR: MIKE) OCTHS.APNO3699-B2 AND OCTH.APNO3699-G2

[SEQUENCE-SIMILAR-TO](http://biobike.csbc.vcu.edu:9003/help/function?PKG=MIKE69767&PACKAGE=BBI&SYMBOL=SEQUENCE-SIMILAR-TO) ([SEQUENCE-OF](http://biobike.csbc.vcu.edu:9003/help/function?PKG=MIKE69767&PACKAGE=BBI&SYMBOL=SEQUENCE-OF) OCTHSE.APNO3699-B2) IN OCTOPUS-E  
 : DNA-VS-DNA)

No matches were found except itself.

QUERY Q-START Q-END TARGET T-START T-END E-VALUE %ID

1. [OctHSe.APNO3699-b2](http://biobike.csbc.vcu.edu:9003/frame?PKG=MIKE93663&NAME=OctHSe.APNO3699-b2) 1 943 [OctHSe.APNO3699-b2](http://biobike.csbc.vcu.edu:9003/frame?PKG=MIKE93663&NAME=OctHSe.APNO3699-b2) 1 943 0.0 100.0

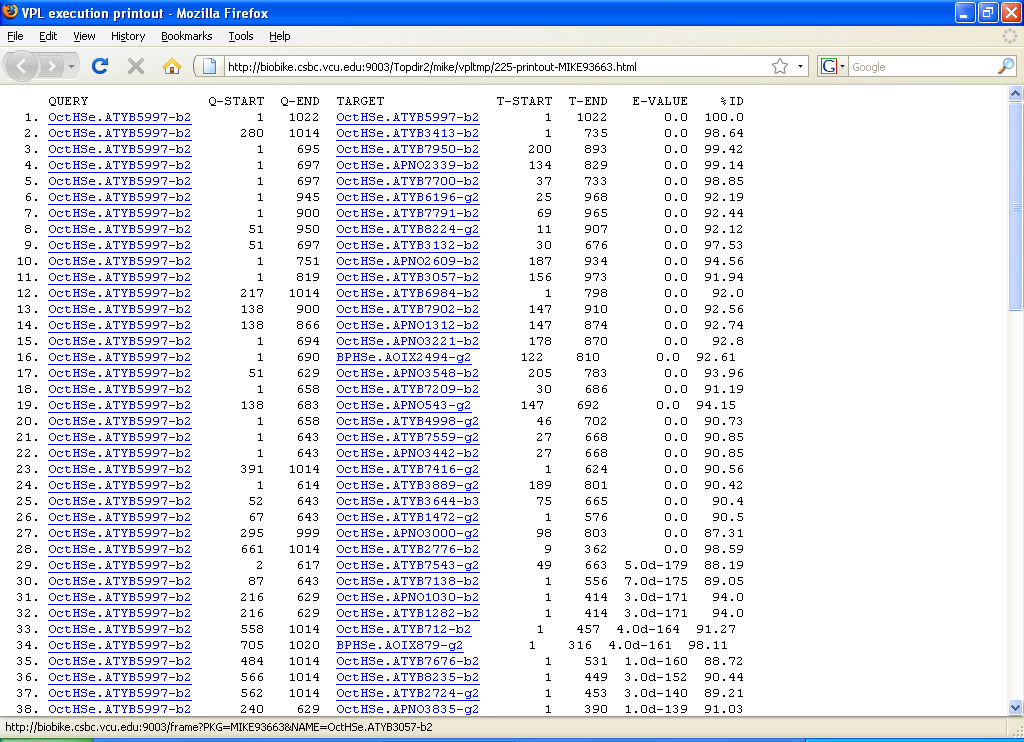
Therefore, I had to claim another read to get match and able to extend my read.

A second claim results:

(CLAIM-READ-FOR: MIKE) OCTHS.ATYB5997-B2 AND OCTH.ATYB5997-G2

Comparing my read with octopus-e:

([SEQUENCE-SIMILAR-TO](http://biobike.csbc.vcu.edu:9003/help/function?PKG=MIKE93663&PACKAGE=BBI&SYMBOL=SEQUENCE-SIMILAR-TO) OCTHSE.ATYB5997-B2 IN \*HOT-SPRINGS-EDITED\*)



After three more blast these are the best matches to extend my read:

Whole Extended Read   
Brown = [OctHSe.ATY 5997-b2](http://biobike.csbc.vcu.edu:9003/frame?PKG=MIKE67707&NAME=OctHSe.ATYB4301-b2) (my Read)  
Green =[OctHSe.APN4301-b2](http://biobike.csbc.vcu.edu:9003/frame?PKG=MIKE67707&NAME=OctHSe.APNO659-b2); Blue = [OctHSe.APN-b2](http://biobike.csbc.vcu.edu:9003/frame?PKG=MIKE67707&NAME=OctHSe.APNO659-b2); orange =[OctHSe.APNO3102-b2](http://biobike.csbc.vcu.edu:9003/frame?PKG=MIKE67707&NAME=OctHSe.APNO3102-b2)

1 905 1014 1606 1872 2398 2525 3326

The black region shows how my best selected matches overlap to each other. Connecting these four matches had shows the extended contig containing 3326 nucleotides. Match number 3 and 4 had extra nucleotides that are not overlap completely and these nucleotides were removed to extend the whole contig. I used Gene Mark if there are any genes can be found in my extended contig and the result was seven predictions of gene.

From Gene Mark I had found seven predicted genes:

**Gene Predictions in Text Format**

**Information on input sequence**

Sequence title: Fri May 8 14:13:37 EDT 2009

Length: 3326 bp

G+C Content: 48.14 %

**Parse predicted by GeneMark.hmm 2.0**

GeneMark.hmm PROKARYOTIC (Version 2.6r)

Sequence file name: sequence, RBS: N

Model file name: heuristic\_no\_rbs.mat

Model organism: Heuristic\_model\_for\_genetic\_code\_11\_and\_GC\_30

Fri May 8 14:13:37 2009

Predicted genes

Gene Strand LeftEnd RightEnd Gene Class

# Length

1 - 382 555 174 1

2 - 688 1167 480 1

3 - 1142 1768 627 1

4 - 1823 2065 243 1

5 - 2062 2481 420 1

6 - 2483 2590 108 1

7 - 2878 3234 357 1

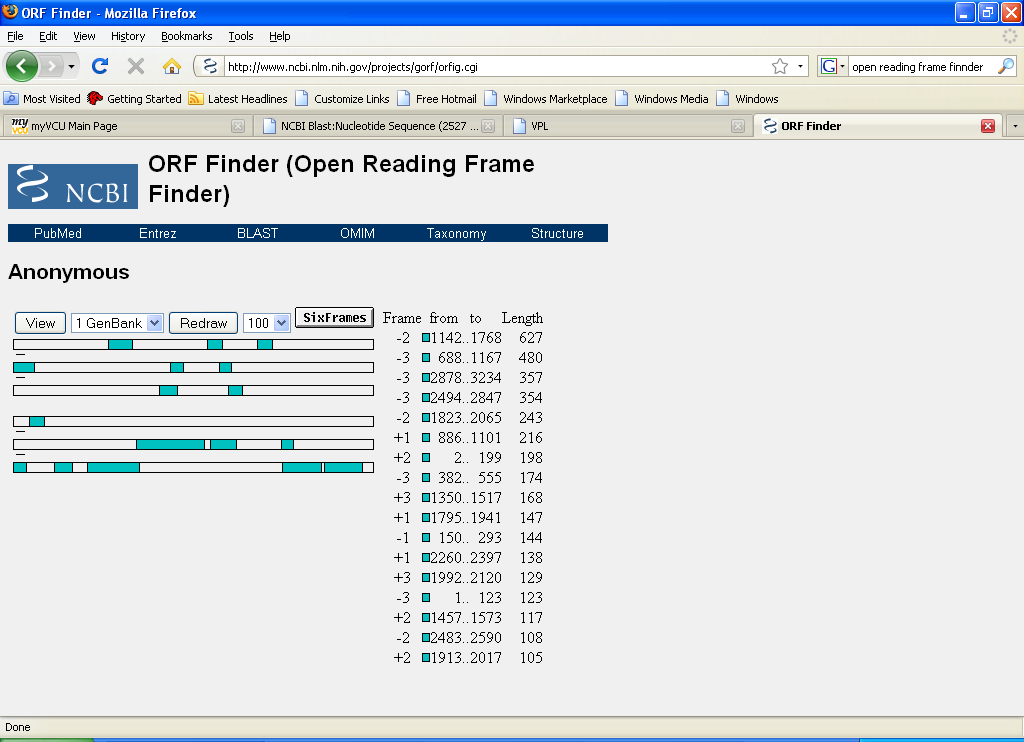
1 3326

Gene 1 Gene 2

Gene 3 Gene 4

Gene Gene6 Gene7

**Reading- Frames-Of the extended sequence:**

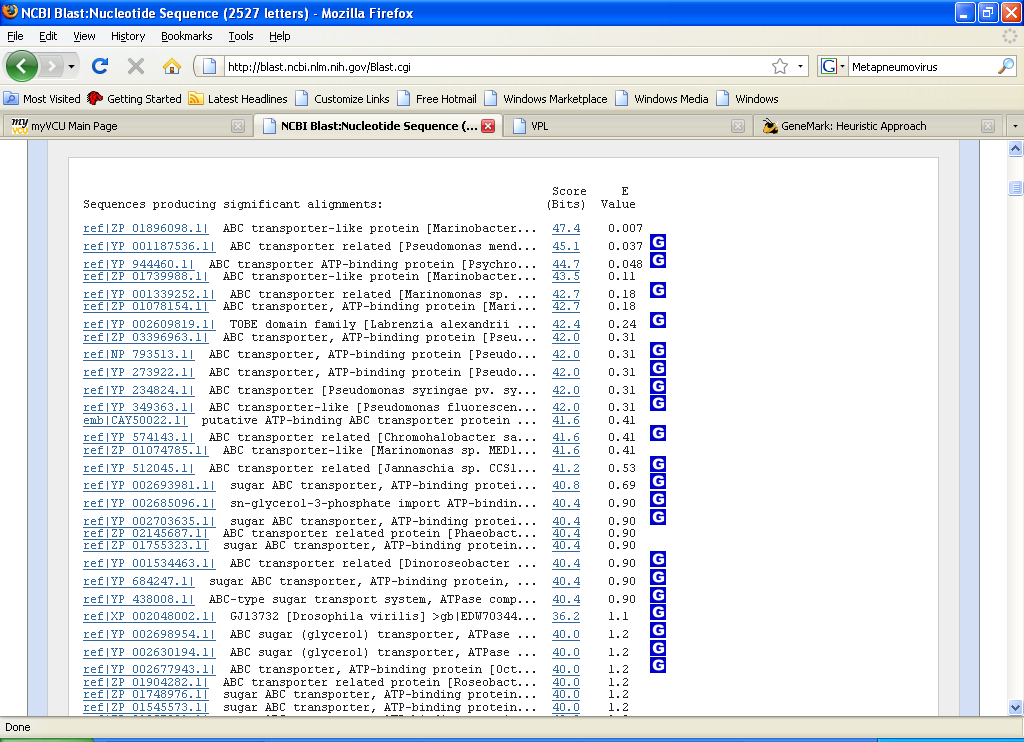
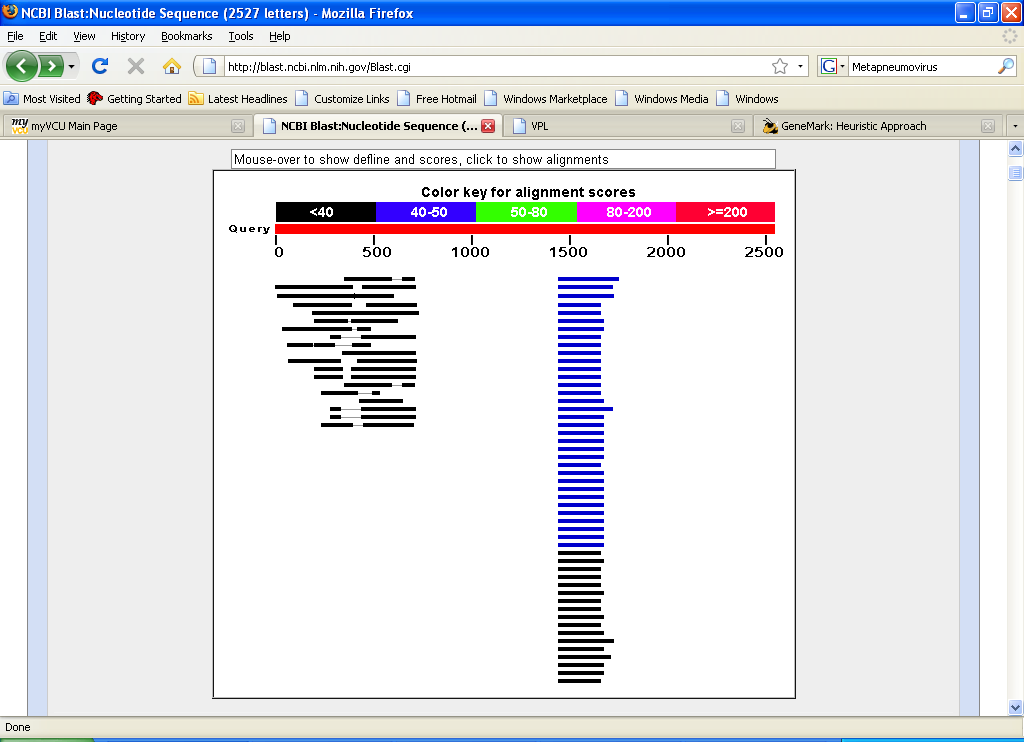


Reading- Frames-Of the extended sequence from ViroBike:

* (B 150 458)
* (B 688 1167)
* (B 1142 1768)
* (B 2062 2481)
* (B 2494 2847)
* (B 2878 3234)

And there are clear similarities between NCBI and ViroBike results. There open Reading frame results are the same.

NCBI Blast Result  
**protein** database using a **translated nucleotide**



Using the NCBI I had these proteins that matches with my sequences.

**Protein Result from ViroBike:**

([SEQUENCE-SIMILAR-TO](http://biobike.csbc.vcu.edu:9003/help/function?PKG=MIKE93663&PACKAGE=BBI&SYMBOL=SEQUENCE-SIMILAR-TO) ([JOIN](http://biobike.csbc.vcu.edu:9003/help/function?PKG=MIKE93663&PACKAGE=BBI&SYMBOL=JOIN) MTSEG+SEQ58+60 SEQ114) IN \*KNOWN-VIRUSES\*  
 :TRANSLATED-DNA-VS-PROTEIN)

I did not get any kind of match between my extended sequence and any known- viruses under TRANSLATED-DNA-VS-PROTEIN

However, using less sequence on my extended read, from 1-2527 gives me a known protein below:

QUERY Q-START Q-END TARGET T-START T-END E-VALUE %ID

1. Seq1 34 501 [NC\_007652.p-gene0008](http://biobike.csbc.vcu.edu:9003/frame?PKG=MIKE93663&NAME=NC_007652.p-gene0008) 217 380 8.0d-4 27.44

[NC\_007652.p-gene0008](http://biobike.csbc.vcu.edu:9003/frame?PKG=MIKE93663&NAME=NC_007652.p-gene0008) is

[Sequence-Length](http://biobike.csbc.vcu.edu:9003/frame?pkg=MIKE67707&printlimit=30&printlispy=NIL&hideparents=T&hidechildren=T&hidesuperparts=T&hideparts=T&elementsperrow=5&name=Sequence-Length) =585

Class – Eukaryotic- Virus

Organism- Metapneumovirus

Human **metapneumovirus** (hMPV)

Was isolated for the first time in 2001 in the Netherlands, by using the RAP-PCR (RNA arbitrarily primed [PCR](http://en.wikipedia.org/wiki/PCR)) technique for identification of unknown viruses growing in cultured cells.

Conclusion:

From the result I found I am predicting that the read that I had some kind of similarities with Metapneumovirus that was first isolated in Netherlands.