Pratik Patel

BNFO 301

Journey to Brachyspira hyodysenteriae WA1

I found new and strange things during my journey. I started to camp in classes and other places to get some knowledge that would assist me in solving the mystery. Basically I start to learn new tricks in biobike, which is our only support to solve this mystery. This mystery was to find my gene and find whom it belongs to.

First I referred to ‘Introduction to my gene’ and learn the basic stuff about Biobike, and then I moved to the 7 different problem sets to learn the different tricks and steps. I wanted to get the result and surprise everyone. So finally I am telling you real story.

I claimed my ‘read’ in Virobike by using step of claim-read for ‘Pratik’ and my read name was ‘othse.atyb2218-b2’. I looked my sequence using ‘display sequence of my read’. After looking at my sequence, I think I can easily find these numbers which I got in result. I saw number of my sequence was 977 of nucleotides. It was too small that I thought in beginning. I used ‘sequence of’ to see numbers of my sequence. I was using during rest of this project notes of Monday 6th April Initial analysis of reads in calendar. I was so excited to find the new gene. I was in middle of my journey and during the other time, I was so busy with the other stuff, but still every week I went to the camping in Life Science building Room no 335. I din’t only go to room 335 but playing every day with computers in lab with genes and nucleotides and I was learning new things.

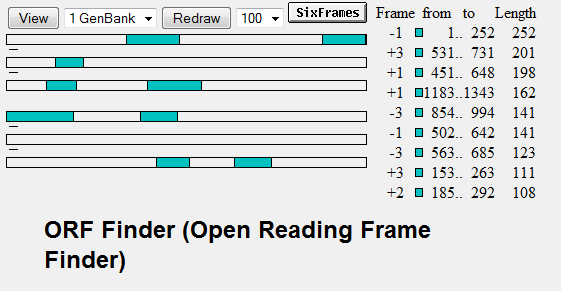
I used ‘sequence similar to’ octopus and I got result that is similar with the beginning and middle part. Next, I defined the octopus100 by using ‘define’ code. I also defined first 100 and contigs with octopus. I got the result like all first 100 sequence with matching some of beginning and middle part in all of them not matching end part. I thought I had done so far but wait, when I was doing that part and my professor buzzed me and told me that we changed the octopus to edited octopus that is called ‘octopus-e’. I was surprised and was lost for some hours; I was worried about new changes in octopus. Later I used the same procedure with the edited octopus and I found better results. A strange thing that I found in the results was that there were no more same sequences. New stuff was not bothering me, so I was kept going by my notes. I turned my journey to different ways, now I went back on results of sequence similar to my reads with the edited octopus. I was searching the query part of my reads for numbers that were matching with the target with other reads. After lot of searching, I found at least some numbers that were closest to my reads. Later, I took on one random read that had closest identity to 100% and was matching my sequence in the middle part. I went through may be 5 to 6 different reads to find overlap between two reads. I drew every single time by sketch and tried to predict the result. In one of them, I found one read that was 97% identified, so I thought to go with that one and move to further steps.

1344

This sketch that I made in my computer, was exactly as I made in my drawing book by hand. I found that there was number that were matching with both the reads, 466 to 960 in my reads and that was matching to 1 to 496 in other read that’s name was Octhse.atyb3673-g2. So I found total 1344 nucleotides in my read and new read. After this all hand drawing I was so tired, I found that in virobike I used ‘aliment’ code to match both sequence. I was so confused and had so many questions, and what was moving in my mind was when will I get my gene, and when I will solve mystery??

I moved further after my tea break and I found one article Appl. Environ. Microbiol wrote by Schoenfeld et al in 6th notes. I read through that and found more clues regarding why sequences matched with one another. Author mentioned that “Viral DNA was physically sheared to 3 to 6 kb using a HydroShear device. In that ends were made blunt using the DNATerminator end repair kit and the fragments were ligated to a double-stranded asymmetrical linker. In one of phosphorylated and other one is nonphosphorylated, they used PCR amplification and used vent DNA polymerase. The amplification products were gel purified again, inserted into the cloning site of the transcription-free pSMART vector and used as transform E.coli.” I read this notes and tried to join my both read’s sequences. There were times when I was lost, but the each weeks camping in the class was very helpful to me. So I used ‘join’ code and I got new sequence and it had 1344 nucleotides. I saw that the sequence and my continued efforts had moved further and further. I went through again same procedure I take that sequence and match with octoupus edited. I found another read that found with my new read. After I went to sketch and aliment and join that sequence all these codes were always helping. I went this procedure four times and moved to numbers of nucleotide near to 2500 but I don’t feel good with that all that. Since I remember that my professor told me that “longer sequence will I get, it was easy to find quickly gene” (Dr. Jeff).

Finally I went further more to I took my old number 1344, and I took that sequence and I copy and paste in NCBI Blast and I found two numbers. I took one of that E-value is 2.7 and also I saw open reading frame of that sequence and I found below table.



Finally I found that my sequence was matching to my gene that was matching to the hypothetical protein. I was very happy. Yes, what I had found was hypothetical protein. It is protein BHW1\_00001. Its full name is Brachyspira hyodysenteriae WA1. It comes with family like Bacteria, spirochaetes. “It colonize the large intestine of piges and cause swine dysentery. The genome of *B. hyodysenteriae* strain WA1 consisted of a single circular 3,000,694 base pair (bp) chromosome, and a 35,940 bp circular plasmid. This is the first confirmed report of the existence of a plasmid in *B. hyodysenteriae*.”

Yes, I am happy but little nervous too, Am I correct this kind of questions was moving around me. Biobike and Virobike was so helpful during my journey. Biobike was so friendly. In beginning I thought it will tough for me. Each and every time you will play with it, you will get better result and also I’ll learn so many thing in the end in Virobike and Biobike.

Sources:-

Biobike and Virobike <http://www.people.vcu.edu/~elhaij/bnfo301-09/Units/Intro-BioBIKE/main.html>  
Assembly of Viral Metagenomes from Yellowstone Hot Springs by Schoenfeld T et al. (2008)

Dr. Jeff Elhai- Professor in Virginia Commonwealth University