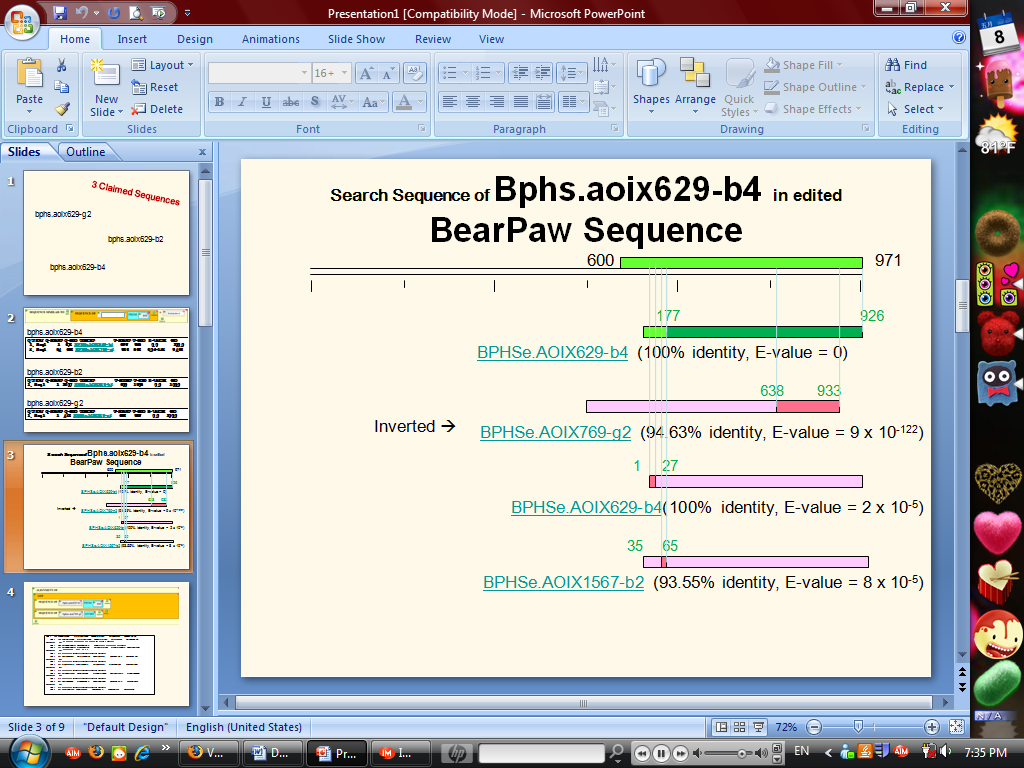
In this study, I have three claimed sequence, they are bphs.aoix629-g2, bphs.aoix629-b2, and bphs.aoix629-b4. Since they are on the same sequence, I compared each one of them with BearPaw sequence in order to find out the one that is most similar to BearPaw sequence. I set the search starting from 600 nucleotides. Both bphs.aoix629-g2 and bphs.aoix629-b2 have only one match that is 100% identically matched with the E- value of 0.0. Bphs.aoix629-b4 has two matches. Other than one with 100% identity, the target, Bphse.aoix769-g2 starting from sequence 933 to 638 matches with the query starting from sequence 64 to 358. It has 94.63% of identity with an E- value of 3.0d-122. The matched target sequence showed that Bphse.aoix769-g2 is inverted.  
 To continue with the analysis, Bphse.aoix769-b4 is selected to compare with the BearPaw Sequence. However, to get rid of the linking in the BearPaw sequence, edited BearPaw sequence is used in this case to higher the probability of getting the right matches of sequence. The edited BearPaw sequence was set starting off from sequence 600 to 971. The narrowing down of the sequence can make sure which of the target can be used for extension and get rid of the unnecessary results. ViroBike shows that the query, Bphse.aoix769-b4 was found to have four sequences matched with the edited BearPaw sequence. Other than one with 100% identity, bphse.aoix769-g2 still has a longer target sequence of 295 long. The drawing below shows the matched sequences of Bphs.aoix629-b4 from Sequence 600 to 971 with edited BearPaw Sequence. The colored parts show the matches accordingly.



Drawing1. Matched sequences of Bphs.aoix629-b4 (Sequence 600 to 971) with edited BearPaw Sequence.

Since bphse.aoix769-g2 is too short, extension is needed for further analysis. Due to previous research, bphse.aoix769-g2 was found to be one that matches bphse.aoix769-b4 the most. So, Bphse.aoix769-b4 was set to align with bphse.aoix769-g2. The result was showed as of below:

Seq 1 1 TCCGATGTTG GTAAGCTGAA CATTGGTTCT GTCGAGATCA ATGGATTTTG

Seq 2 0 ---------- ---------- ---------- ---------- ----------

consensus 1

Seq 1 51 CCAGTCCGTT ATCCACAAGA CTTTTAATCA CATTTCT-TA CTTTTCGGTA

Seq 2 1 ---------- ---CACAAGA CTTTTAATCA CATTTCTACT TTTTTCGGTA

consensus 51 \*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*

Seq 1 100 TTCGCATAGC TTGCTTCTAC GGTTTTAAGA TTTTCTTCCT GAACAAAAA-

Seq 2 38 TTCGCATAGC TTGCTTCTAC GGTTTTAAGA TTTCTTTCCT GAACAAAAAA

consensus 101 \*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\* \*\*\* \*\*\*\*\* \*\*\*\*\*\*\*\*\*

Seq 1 149 CCTGATAGTA AGCGGATGCT ACGTTTTCGA TGATCTGCTC ATTGGTAAGC

Seq 2 88 CCTGATAGTA AGCGGATGCT ACGTTTTCGA TGATCTGCTC ATTGGTAAGC

consensus 151 \*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\*

Seq 1 199 TGTGCATTAA GAACATAAAA ATCTCTTGTA GATTTGGCCG CCTTCAGTCC

Seq 2 138 TGTGCATTAA GAACATAAAA TTCTCTTGTA GATTTTGCCG CCTTCAGTCC

consensus 201 \*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\* \*\*\*\*\* \*\*\*\* \*\*\*\*\*\*\*\*\*\*

Seq 1 249 TGTGAACACC CTCCGAACGA AAATGGCCTG TTGAAGCTGT ACGGAAGCCG

Seq 2 188 TGTGAACACT CTCTGATCGA AAATTGCCTG TTGAAGCTGT ACGGAAGCCG

consensus 251 \*\*\*\*\*\*\*\*\* \*\*\* \*\* \*\*\* \*\*\*\* \*\*\*\*\* \*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\*

Seq 1 299 TAAAGTTCCA GGGCTGTCCA AACCGGGCTC TGAATTTTTT CCCCTCCAAA

Seq 2 238 TAGAGTTCCA GGGCTGTCCA AACTGGGCTC TGA-TTTTTT CGCCTCCAAA

consensus 301 \*\* \*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\* \*\*\* \*\*\*\*\*\* \*\*\* \*\*\*\*\*\* \* \*\*\*\*\*\*\*\*

Seq 1 349 TGCAAGAAGG ATTTCCGAAA TAAG------ ---------- ----------

Seq 2 287 TGCAAGAAGG GATTCCTGAA TGATGGGATT GTAGGTAAGA CCTGCTGTCG

consensus 351 \*\*\*\*\*\*\*\*\*\* \*\*\*\* \*\* \* \*

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 337 CTGAAATCTG AGGTAAAGCT CCTGCTTTTG CCTCATCAAT CTTATATTCG

consensus 401

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 387 GCTTTTTTGA TCTGTAAAGC TGCCTTTTTG GCTTCCGCTT TATTCTGAAG

consensus 451

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 437 TGCCTGTTTG ATCGCTTCCT GTAAGGAAAC CTGCTGTTGT GCAGACACCG

consensus 501

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 487 ATGAAAAACC GAACATCATA AATGCAGCCG CTATTCCTAC TTTTAGCTTT

consensus 551

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 537 TTTGCAGTTA TACGTTTTTT CATAATATTA TACTTCGTTT ATTTTTTTAA

consensus 601

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 587 TGTAATTATT ATGTTAAATG CTTTTATATC TAAATGACGA ATCATTTCTT

consensus 651

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 637 AAAATACTTT AACATTTTTT TAGTTTTTAA ATAGCATATT CAAAATGATC

consensus 701

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 687 TCTTTTCTTT CGGAAATGAT CTTATCAAAC TCTTCCTCAC TGATCATAAG

consensus 751

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 737 ATTTTCCATT AATAAAGGTC TTACCGCACT TGGGAAAACC AGCAGAGAAA

consensus 801

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 787 TCATATTCAG AAGAAACTGG ATAGGTTTCA TCCTTTCTAT ATTTCCACGT

consensus 851

Seq 1 372 ---------- ---------- ---------- ---------- ----------

Seq 2 837 TCCATTTCTG TTTCAATATC TTTAAAAAGC CTTTTCAGTT CATCTTCTTC

consensus 901

Seq 1 373 ---------- ---------- ---------- ---------- -------

Seq 2 887 GATATCCTTT TTATGACAGT TTCCTTTGTT GATCTGTGAA ACGATAT

Table2. Alignment of Bphse.aoix769-b4 with bphse.aoix769-g2

In the table, the stars show the identical nucleotides. To extend the sequence of bphse.aoix769-g2, the sequence of Bphse.aoix769-b4 was copied and connected to the end of the bphse.aoix769-g2 sequence.

Table3. Extended Sequence of bphse.aoix769-g2

The extended sequence of bphse.aoix769-g2 is used to look for species with similar sequence in genebank by searching in translated DNA vs. protein. Top two results happened to show the same target organism of Chryseobacterium gleum ATCC 35910. The first results showed with a low e- value of 1.53181d-35 and a percentage identity of 53.05. The second target organism was showed to be Flavobacterium johnsoniae UW101 with a higher e-value of 4.47762d-19 and an identity percentage of 57.14. These two species are one of the outer membrane efflux proteins.

The tBlastx search was performed again in NCBI. tblastx takes a nucleotide query sequence, translates it in all six frames, and compares those translations to the database sequences dynamically translated in all six frames. NCBI performed the search in a variety of library with including non-redundant GenBank CDS translations, PDB, SwissProt, PIR, and PRF. The top result was showed to be Flavobacterium johnsoniae UW101, with an e – value of 6e- 48. And the second result appears to be Cytophaga hutchinsonii ATCC 33406 with an e – value of 2e- 48. Flavobacterium johnsoniae is an aerobic Gram-negative bacterium that is commonly found in soil and freshwater.   It belongs to the large and diverse group of bacteria known as the bacteroidetes which are also referred to as the 'Cytophaga-Flavobacterium-Bacteroides' group.

Since Flavobacterium johnsoniae and Chryseobacterium gleum are the two top results of the search, more research was done based on these two species. It was showed that Chryseobacterium includes six species that were previously designated members of the genus Flavobacterium. Chryseobacterium gleum previously known as Flavobacterium CDC group, and have been clearly differentiated by DNA-DNA homology and eight phenotypic characteristics. Based on this, I assume that the three claimed sequences, bphs.aoix629-g2, bphs.aoix629-b2, and bphs.aoix629-b4 are obtained from the sequence of Flavobacterium johnsoniae.

Reference

Perlorentzou, Stavroula; Chalkiopoulou, Irene; Athanasiou, Athanasios; Legakis Nikolas;

Christakis, George . "Chryseobacterium indologenes Non-Catheter-Related Bacteremia in a Patient with a Solid Tumor". American Society for Microbiology. December 13, 2004