Sergio Mares

Phylogenetic Tree of 2701 Sequences

Introduction
A total of 2699 *Pseudomonas* sequences were analyzed in this study, comprising 22 isolation sources. Those 4156 sequences were obtained through a basic local alignment search tool to query a nucleotide sequence against a database of nucleotide sequences (BLASTN). The nucleotide sequence used in the alignment was a protein whose absence causes an alteration in swarming motility, and more importantly an alteration in the tolerance of high levels of Ca^{2+} ; *carP.* A set of 2699 sequences were selected based on a identity value percentage greater than 90%.

Multiple Sequence Alignment Analysis
In order to visualize the evolution of *carP*, the isolation source was obtain. In addition, a unique identification name was given to the sequences to ease the identification and evolutionary patterns in a phylogenetic tree.

The multi sequence alignment software MEGA 7.0, aligned all the sequences and a FASTA file was export from this alignment.
Then MOTHUR software was utilized to create distance matrix. A lower triangle matrix calculated the similarities among the sequences.

Then, a cluster command assigned the sequences to operational taxonomic units (OTUS) at 0.001% distant from similar sequence in the OTU with a precision of 1000. The command resulted in a total of 144 OTUS. Then, a representative file used one alignment from each OTU.
After the addition of *carP* (PAO1) and *PA0327* (PAO1), MEGA 7.0 constructed a phylogenetic tree of maximum likelihood with *carP* rooted.

Observations

1. After the commanding MOTHUR to assign
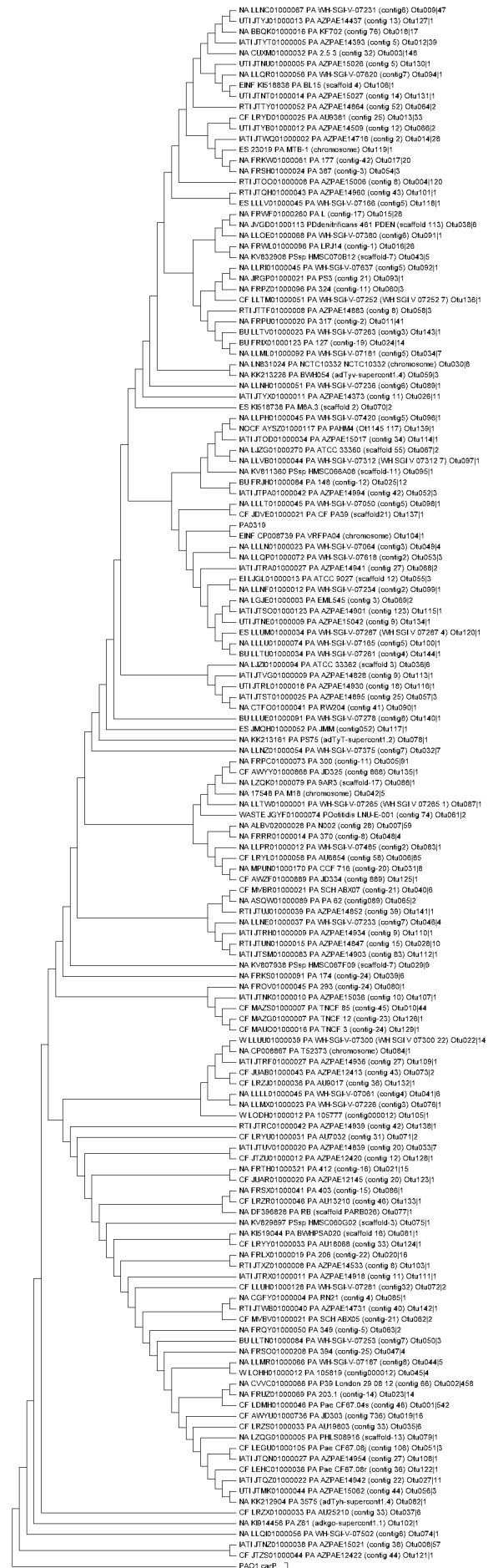
sequences to OTUS, 70 OTUS contained one sequence. This can be interpret as an increase in the distance after 2629 sequences, resulting in 70 sequences being highly distant from the rest. **My question is** should I remove the rest of the 70 sequences? If not, what does having this 70 sequences mean?

2. The point of commanding OTUS was to better visualize the various isolation source. However, 45% of the representing sequences have a NA label. Should I construct a tree without the NA sequences?

3. Based on the observation above, we can visualize that 4 out of the 144 representative files are non-*Pseudomonas aeruginosa* species. For reference 52 sequences out of the 2699 from the original file are not *P. aeruginosa.*

4. The OTU with more sequences and closer distance within sequences is located with the red rectangle.