Assignment 2

Bacterial Phylogenetic Trees

V.Pragatheeswaran 150476J

This is submitted as a partial fulfillment for the module

BM4321: Genomic Signal Processing

Department of Electronic and Telecommunication Engineering

University of Moratuwa

10th of June 2019

Introduction of the analyzed bacterial species

Acetobacter falls under the genus of acetic acid bacteria. These types of bacteria are known for their conversion of ethanol to acetic acid in the presence of oxygen. The earliest interaction of acetobacter was in 1998, when two species were extracted from red wine and cider vinegar. The following are the scientific classification for acetobacter,

Domain : Bacteria

Phylum : Proteobacteria

Class : Alphaproteobacteria

Order : Rhodospirillales

Family : Acetobacteraceae

In this assignment a total of 11 organisms (22 accessions) under the acetobacter genus will be taken to analyses the taxonomy between them. Adding the mutation distances will be focused on this report too.

This document will specifically analyses all these 11 organisms based on the protein – succinate dehydrogenase.

Similar Work

A lot of study have been conducted in finding the phylogenetic relationship between many sub genus in the acetic acid bacteria. Pitiwittayakul *et al* 2015 [1], isolated samples of **Acetobacter thailandicus** from Thailand , using 16S rRNA gene sequences as well as 16S-23S rRNA ITS sequences the classification was compared. The paper goes on to mention about various methodologies of extraction and the analysis methodologies used. Figure 1 , was the neighbor joining (NJ) tree presented in the results which we will take into account for this assignment for verification purposes.

The results show that Acetobacter can be classified in two major phylogenetic groups, the first

group corresponds to **acetobacter aceti** and the second one corresponds to **acetobacter pasteurianus.** These results were obtained at a bootstrap value of 100%.

Tanasupawat S, *et al* 2009 [2] did an identification of Acetobacter , Gluconobacter and Asaia strains isolated in thailand based on 16S-23S rRNA gene. As usual the paper discussed deeper into isolation and positions as well. If we analyze the NJ tree presented in the discussion, where they have created the phylogenetic tree based on 16s rRNA gene requences for thai isolates assigned to the genus Acetobacter. The tree is shown in figure 2 and the bootstrap values derived are from 1000 replications.

A very similar work was done by Matsutani M *et al* 2010 [3] too. This NJ tree is presented in figure 3.

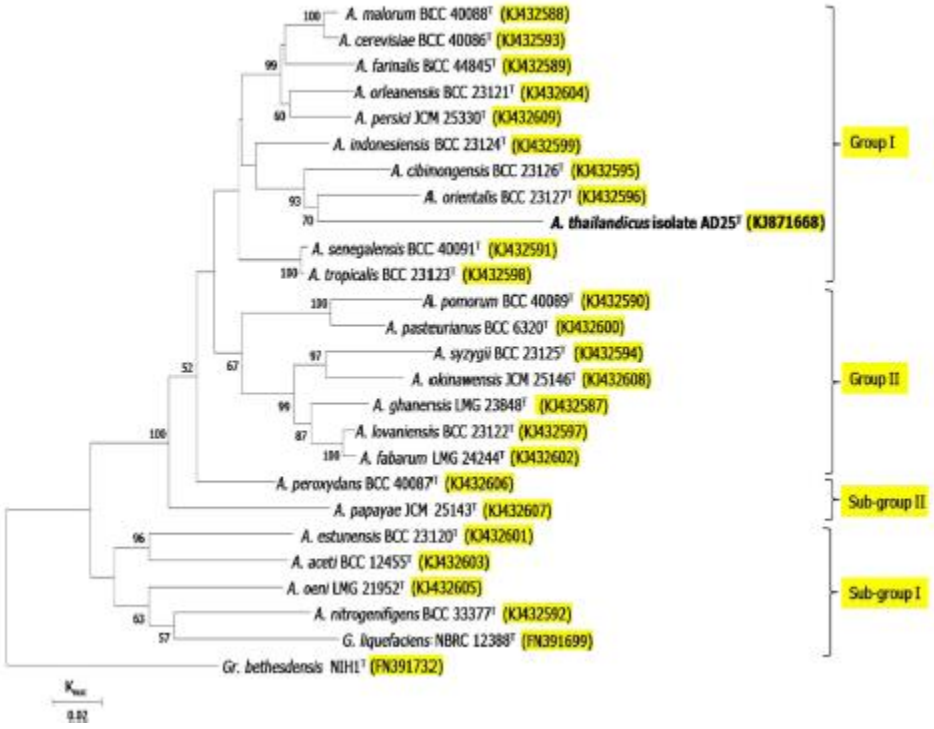


Figure ***Error! No text of specified style in document.***

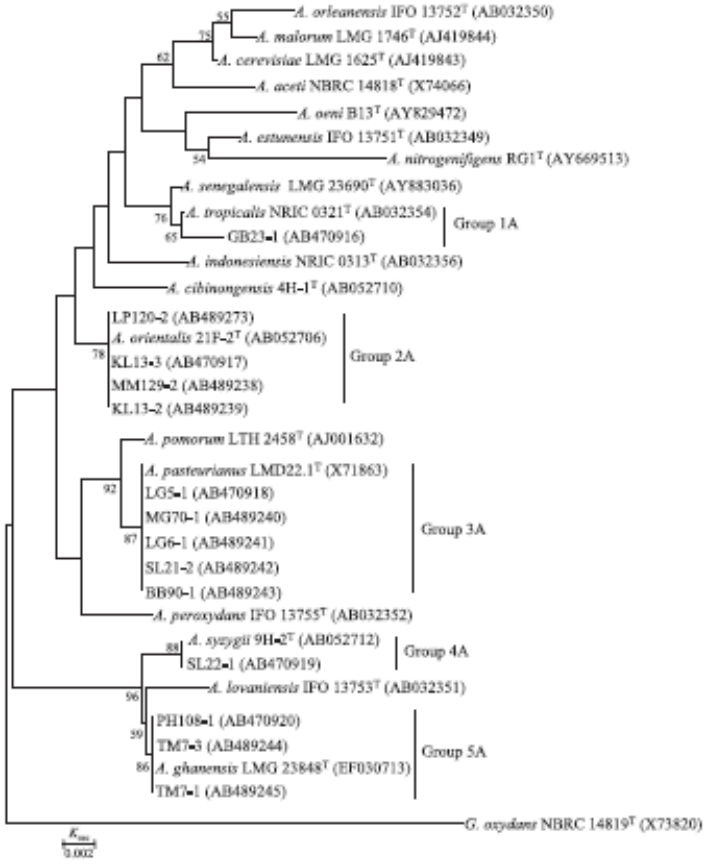


Figure 2

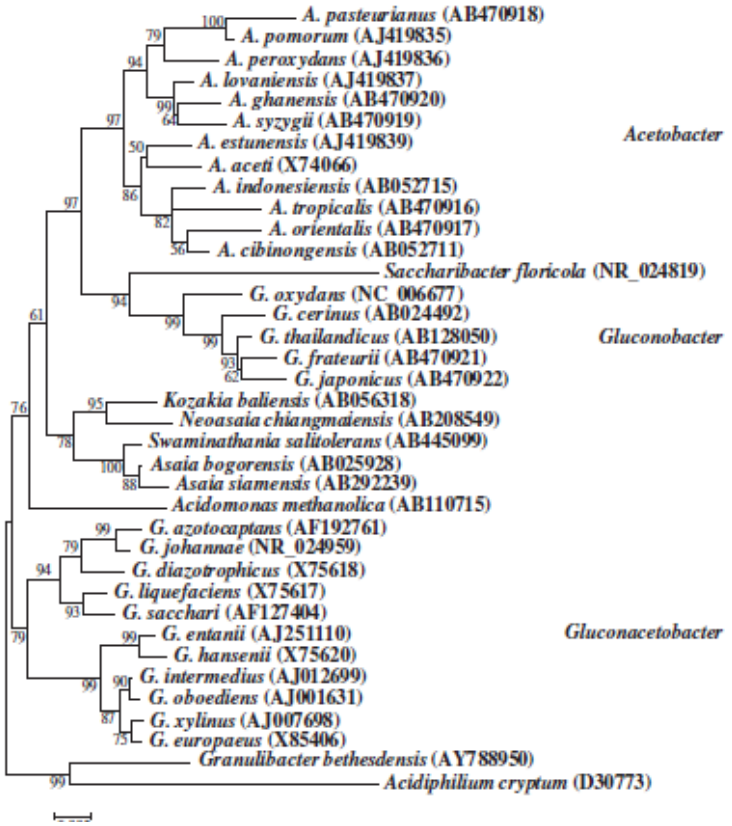


Figure 3

Significance of the selected protein

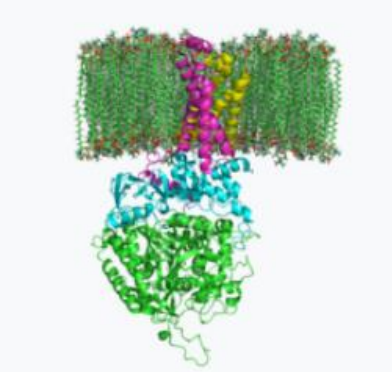
"succinate dehydrogenase" is the protein which was selected in order to analyze the taxonomy. This enzyme complex is found mainly in bacterial cells and in the inner mitochondrial membrane of the eukaryotes. Additionally this is also the sole enzyme which participates in both citric acid cycle and electron transport chain. It is also said that, high amounts of succinate dehydrogenase in muscle demonstrates high mitochondrial content and high oxidative potential [3].

Ruiz *et al* 1972 [4],studied regulation of succinate dehydrogenase the 'Escherichia coli'. Their results mentioned that the bacteria used the glucose first than oxidizing the succinate dehydrogenase. The respiration activity difference was also visualized in bacteria which were grown in glucose and succinate grown ones. It is also discussed that the formation of succinate dehydrogenase by the bacteria is controlled by the environmental conditions.

Pre processing sequences

a. Protein selection

The protein tables for all 22 species of interest were obtained from the GenBank. After some analysis the "succinate dehydrogenase" was selected for the analysis purposes as it was present in all 22 species after manual inspection. The selected protein starting and ending positions were obtained from the protein table. In situation of protein appearing in multiple locations one of them were selected at random.



Succinate Dehydrogenase

b. Extract genes

The sequences with the above extracted start and end points were extracted as FASTA format from GenBank.

c. Gene alignment

The downloaded sequences were loaded individually in clustal X, and aligned. The output of the alignment was saved for future purposes.

d. Similarity matric calculation

Post alignment, the similarity matrix was calculated with the inbuilt function in Clustal X.

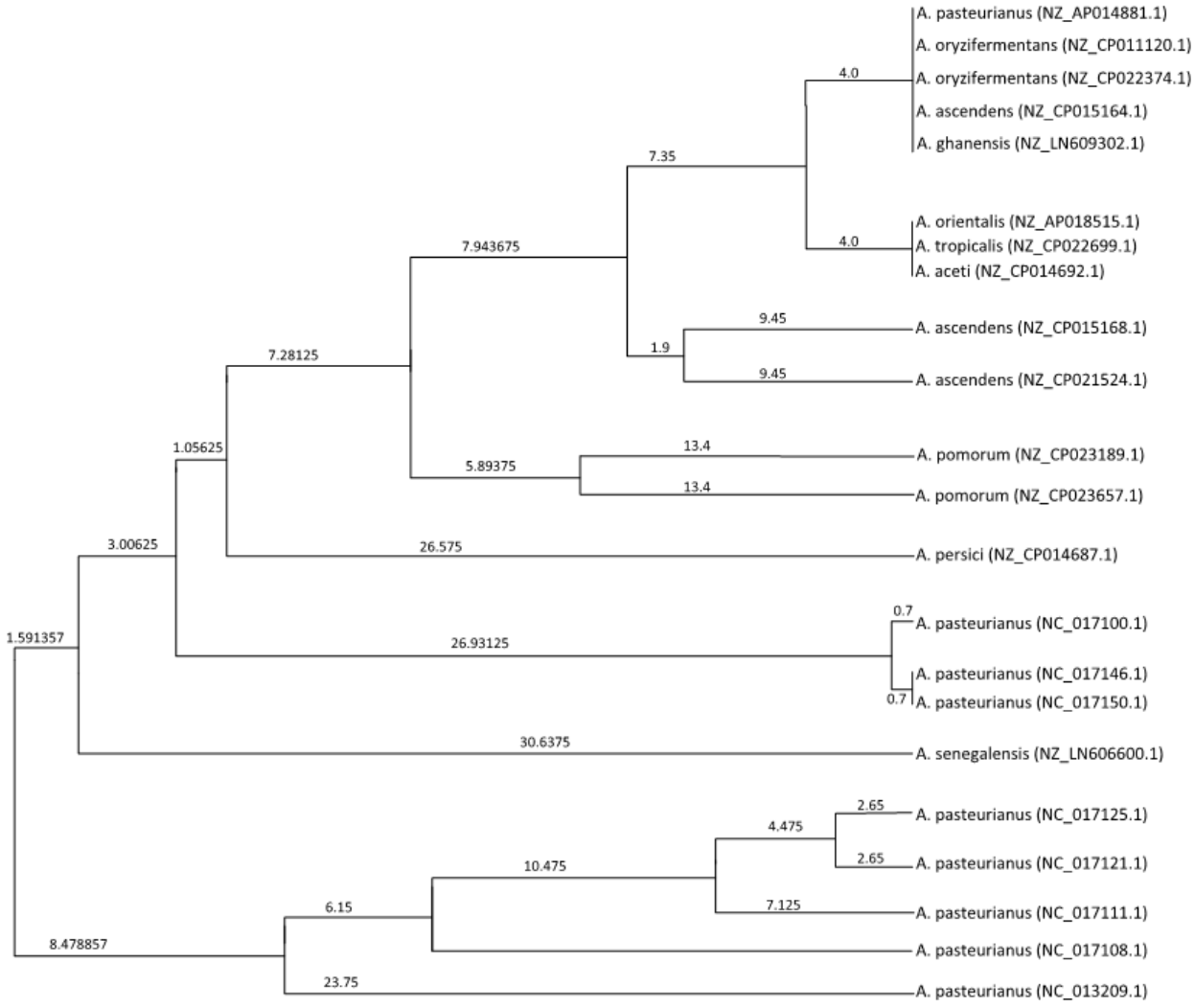
$$S_{i,j} = S_{j,i} = \frac{100}{N}\sum_{k=1}^{N} p \quad where \quad p = \begin{cases} 1, & A_{i,k} = A_{j,k} \\ 0, & otherwise \end{cases}$$

NZ_CP014692.1	Acetobacter aceti
NC_013209.1/AP011121.1	Acetobacter pasteurianus
NZ_CP023657.1/CP023657.1	Acetobacter pomorum
NZ_CP022699.1/CP022699.1	Acetobacter tropicalis
NZ_CP014687.1/CP014687.1	Acetobacter persici
NZ_LN606600.1/LN606600.1	Acetobacter senegalensis
NZ_CP011120.1/CP011120.1	Acetobacter oryzifermentans
NZ_CP015164.1	Acetobacter ascendens
NZ_CP015168.1	Acetobacter ascendens
NZ_CP021524.1	Acetobacter ascendens
NZ_CP022374.1/CP022374.1	Acetobacter oryzifermentans
NZ_AP018515.1/AP018515.1	Acetobacter orientalis
NZ_CP023189.1/CP023189.1	Acetobacter pomorum
NC_017100.1/AP011128.1	Acetobacter pasteurianus
NC_017121.1/AP011135.1	Acetobacter pasteurianus
NC_017125.1/AP011142.1	Acetobacter pasteurianus
NC_017146.1/AP011149.1	Acetobacter pasteurianus
NZ_LN609302.1	Acetobacter ghanensis
NC_017111.1/AP011156.1	Acetobacter pasteurianus
NC_017150.1/AP011163.1	Acetobacter pasteurianus
NC_017108.1/AP011170.1	Acetobacter pasteurianus
NZ_AP014881.1/AP014881.1	Acetobacter pasteurianus

Tree construction

The created similarity matrix was added as the input for the algorithm which was testing in the lab. The outputs of the weighted pair group method with arithmetic mean (WPGMA) and the NJ tree were created and drawn on a 3rd party application. The following is the accession number and the organism chat.

a. WPGMA tree

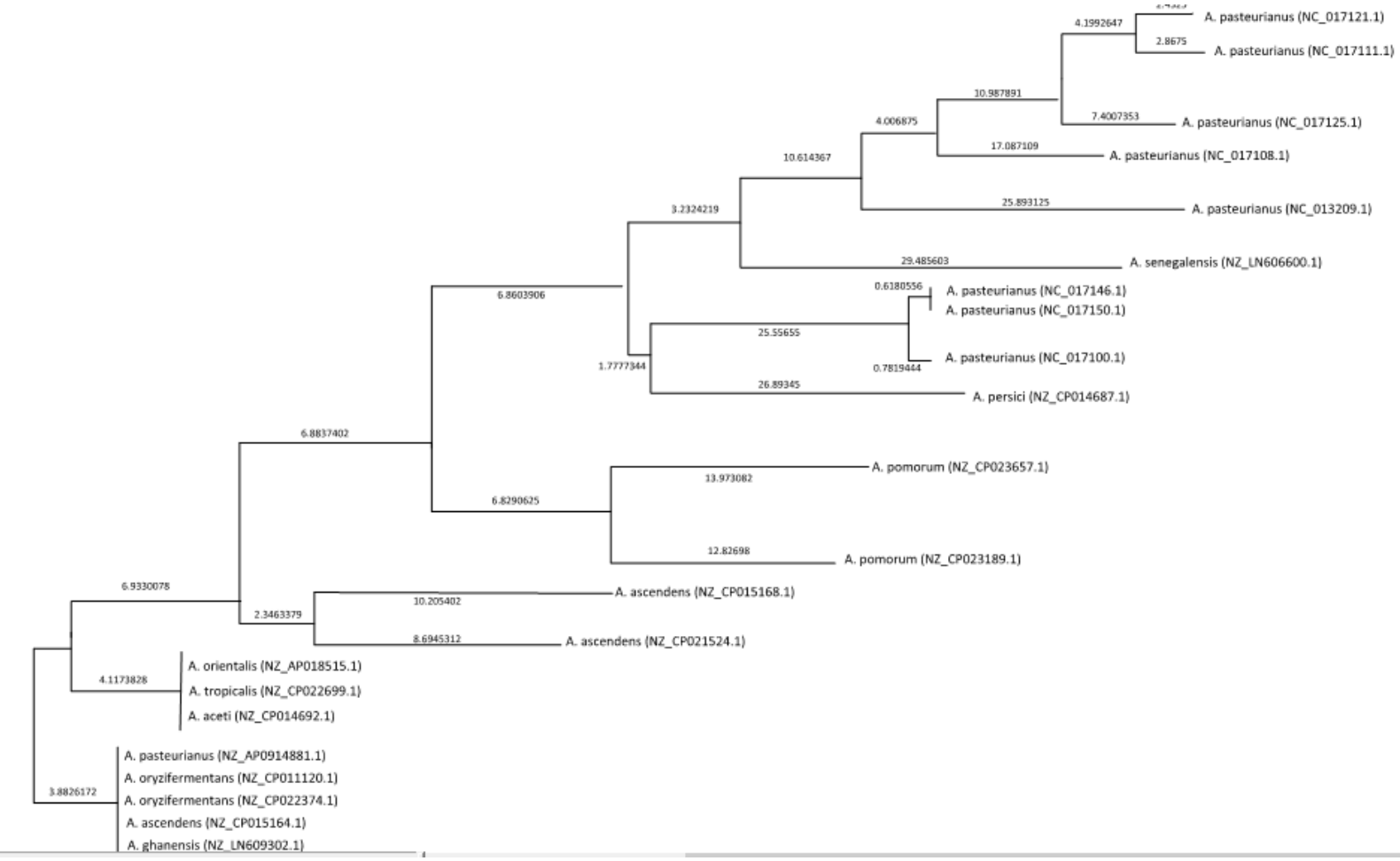


Discussion

As WPGMA's underlying assumption is constant evolution rate, the distances from top to bottom would be equal despite the branch. As shown in the above figure, As expected most of the Acetobacter pasteurianus species extractions have been clustered together or in the same branch itself. The only exception to the above situation is the accession NZ_AP014881.1. Also this can be suspected as a adaptation which happened due to the species area of growth.A similar patter ins visible in between two cluster of Acetobacter pasteurainus itself. Apart from this the Acetobacter pomorum , Acetobacter ascendens,Acetobacter oryzae fermentans extractions have been clustered together. If we focus on the cluster groups on the top branches with different species, namely , Acetobacter ascendens , Acetobacter ghanesis and the second cluster group orientails, Acetobacter tropicalis and Acetobacter aceti suggests that there could have been horizontal gene transfer among them.

Also from the graph we can conclude that senegalensis is at the most distance, meaning not much horizontal transfer has been through this particular specie , based on this protein. Additionally we can see that Acetobacter tropicals , Acetobacter orientalis, Acetobacter ghanensis ,and Acetobacter ascendens have a very close ancestor than any other species which are in the discussion table.

b. NJ tree

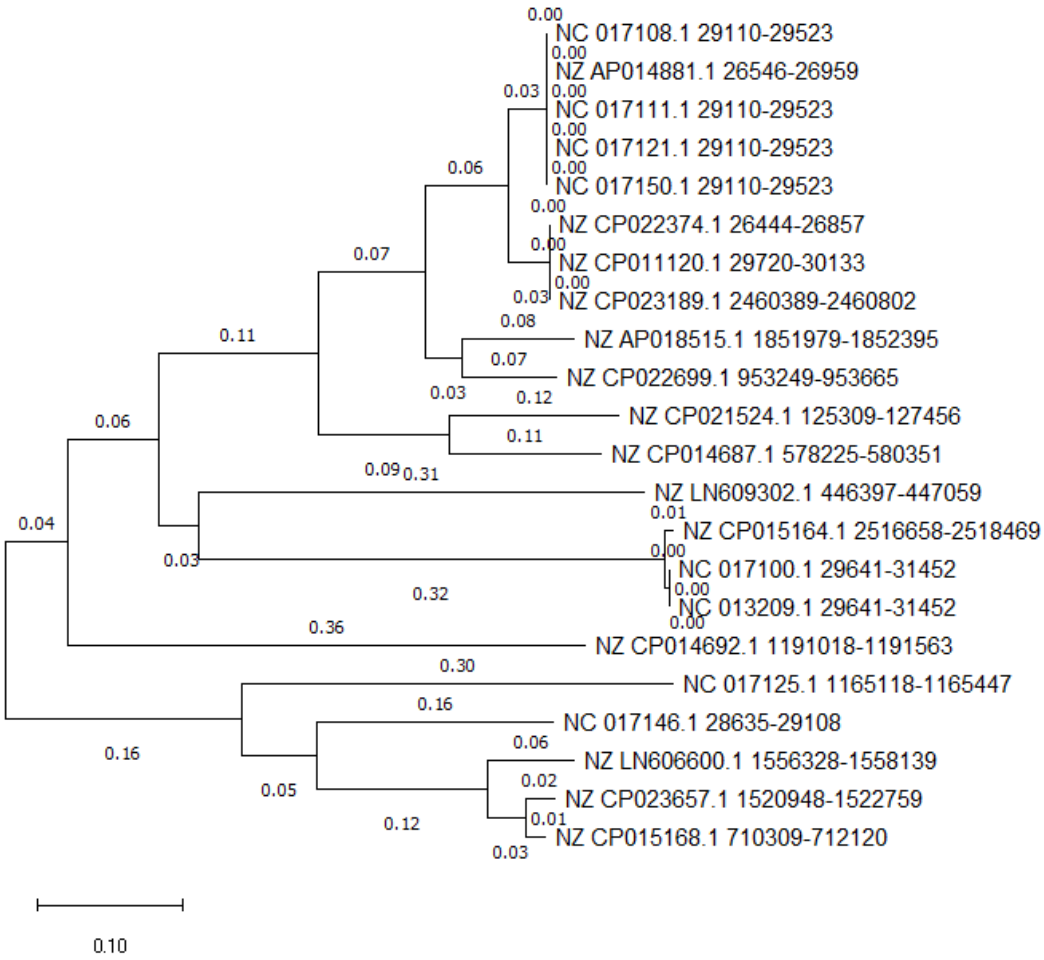


Discussion

According to the NJ tree the acetobacter orzifermentans and acetobacter pasteurianus can be seen to clustered together as expected. From here we can see that the acetobacter pasteurianus has been showing continuous evolution. Some of the later versions of the evolution are from the acetobacter pomorum.

Conclusion

The output which we got from Clustal X and our lab algorithms were bench marked with MEGA-X. In MEGA-X with the FASTA output which we obtained from CLUSTAL-X the tree can be obtained directly. It is very user friendly as the user only has to choose between the tree



category mainly in this case. Even though Clustal -X didn't have the option of a WPGMA, many other tree options were available.

The main reasons for the results being different slightly, mainly due to the pairwise deletion done in the MEGA-X functionality.

References

1. Chaipitakchonlatarn, W., Pitiwittayakul, N. and Yukphan, P. (2015). *Acetobacter thailandicus sp. nov., for a strain isolated in Thailand*. [online] researchgate. Available at:https://www.researchgate.net/publication/270597089_Acetobacter_thailandicus_sp_nov_for_a_strain_isolated_in_Thailand [Accessed 10 Jun. 2019].

2. Tanasupawat S, e. (2009). *Identification of Acetobacter, Gluconobacter, and Asaia strains isolated in Thailand based on 16S-23S rRNA gene internal transcribed spacer restric... - PubMed - NCBI*. [online] Ncbi.nlm.nih.gov. Available at: https://www.ncbi.nlm.nih.gov/pubmed/21566366 [Accessed 10 Jun. 2019].

3. Matsutani M, e. (2010). *Genome-wide phylogenetic analysis of Gluconobacter, Acetobacter, and Gluconacetobacter. - PubMed - NCBI*. [online] Ncbi.nlm.nih.gov. Available at: https://www.ncbi.nlm.nih.gov/pubmed/21182539 [Accessed 10 Jun. 2019].

4. https://en.wikipedia.org/wiki/Succinate_dehydrogenase#Role_in_disease

5. Mic.microbiologyresearch.org. (1972). *Regulation of Succinate Dehydrogenase in Escherichia coli*. [online] Available at: https://mic.microbiologyresearch.org/content/journal/micro/10.1099/00221287-72-1-29?crawler=true&mimetype=application/pdf [Accessed 10 Jun. 2019].