Andrew Lutsky
MCB432
Major Assignment #1

Introduction:
　　In the individual section of this major assignment, 16S RNA short reads were assembled into contig sequences and then inputted into the BLAST suite to find a reference genome sequence. These contig sequences and reference sequences were assembled into one long continuous fasta file.

Procedure:
1. Assembled Databases were collected from each member of the group.
    a. For this group the dataset numbers were 3, 11, 12, and 17.
    b. Contig files that were not generated by Cap3 were done by hand using the BLASTN alignment tool.
2. These assemblies were used to construct the metadata table listed below which shows the sample name of the contig files generated from each dataset and the reference sequence collected from BLAST(the top scorer).
    a. The sample information was collected from the BLAST NCBI website using the accession number.
3. Then this metadata table was used to construct an OTU table, which shows the number of each species for each dataset.
    a. This can be found below.
4. The databases were then separated out to respective sample and reference fasta files and then concatenated to each other so that there was one database of samples and one database of references for all of the datasets.
    a. This was done by manually copying and pasting but could be done using cat command in shell.
5. Then the two fasta files containing the contigs and references were piped into a new file and merged and used to generate a phylogenetic tree after alignment. The neighbor joining method was used and done below.

Discussion:
　　This assignment allowed an assembly of contigs to create a database containing 16sRNA data. The process involved using several bioinformatics tools to develop a database that can be used at a later date. This data was inputted only from short reads of the microbial genome and then used to find data later. The tools involved were cap3, BLAST suite, excel, JalView, and MEGA11. This data can be visualized using a phylogenetic tree to examine the phylogenetic relationships between sample and reference data.
　　The OTU and Metadata Table allow for easy access to information of total samples. We can easily find the total number of each species for each dataset and its exact information so that the data is reproducible.

Metadata Table:

Sample_Name	Year	Month	Source	Bacteria_Species	Accession_No.
2016Mar_B06	2016	March	Ben	Streptomyces sp.	KF194352.1
2016Mar_F07	2016	March	Ben	Methylobacterium extorquens	KF572999.1
2016Nov_B06	2016	November	Ben	Bacillus subtilis	MT133339.1
2016Nov_E12	2016	November	Ben	Bacillus sp.	MT071452.1
2017Apr_B06	2017	April	Ben	Bacillus sp.	MG654640.1
2017Apr_E12	2017	April	Ben	Bacillus aryabhattai	MH010171.1
2017Nov_B09	2017	November	Ben	Streptomyces sp.	MN629139.1
2017Nov_F06	2017	December	Ben	Curtobacterium sp	KU560309.1
2016Mar_A03	2016	March	Yishuo	Brevibacillus porteri	OM432153
2016Mar_E02	2016	March	Yishuo	Bacillus subtilis	ON259697
2016Nov_A03	2016	November	Yishuo	Bacillus sp.	KM104683.1
2016Nov_D05	2016	November	Yishuo	Bacillus subtilis	KX822704.1
2017Apr_A03	2017	April	Yishuo	Streptomyces sp.	KX279639.1
2017Apr_D07	2017	April	Yishuo	Streptomyces phaeochromogenes	NR116381.1
2017Nov_A03	2017	November	Yishuo	Bacillus cereus	KY930704.1
2017Nov_E02	2017	November	Yishuo	Bacillus sp.	HQ231223.1
2016Mar_A12	2016	March	Andrew L.	Bacillus aryabhattai	KJ721212.1
2016Mar_F01	2016	April	Andrew L.	Bacillus subtilis	MK267098.1
2016Nov_E05	2016	November	Andrew L.	Bacillus subtilis	KC250200.1
2016Nov_A12	2016	November	Andrew L.	Bacillus aryabhattai	KJ721212.1
2017Apr_A12	2017	April	Andrew L.	Bacillus sp.	LM655320.1
2017Apr_E03	2017	April	Andrew L.	Bacillus sp.	KM104683.1
2017Nov_B02	2017	November	Andrew L.	Streptomyces sp.	KX279646.1
2017Nov_E12	2017	November	Andrew L.	Bacillus sp.	LN849693.1
2016Mar_B01	2016	March	Andrew H.	Bacillus megaterium	KR085848.1
2016Mar_F02	2016	March	Andrew H.	Bacillus megaterium	MN826585.1
2016Nov_B01	2016	November	Andrew H.	Priestia megaterium	MH127589.1
2016Nov_B01	2016	November	Andrew H.	Bacillus wiedmannii	NR_152692.1
2017Apr_B01	2017	April	Andrew H.	Bacillus megaterium	KU605234.1
2017Apr_E04	2017	April	Andrew H.	Lentzea chajnantorensis	NR_149214.1
2017Nov_B03	2017	November	Andrew H.	Arthrobacter pokkalii	MW487810.1
2017Nov_F01	2017	November	Andrew H.	Streptomyces sp.	LC033903.1

Metadata Table Containing all Samples, their species name, and their accession No.

OTU Table:

Species	Group #11	Group#3	Group#17	Group#12
Streptomyces sp.	1	1	2	1
Methylobacterium extorquens			1	
Bacillus subtilis	2	2	1	
Bacillus sp.	3	2	2	
Bacillus aryabhattai	2		1	
Curtobacterium sp.			1	
Brevibacillus porteri		1		
Streptomyces phaeochromogenes		1		
Bacillus cereus		1		
Bacillus megaterium				3
Priestia megaterium				1
Bacillus wiedmannii				1
Lentzea chajnantorensis				1
Arthrobacter pokkalii				1

OTU table

Phylogenetic Tree: