**Repeat Masker and RNA finder document**

This document provides the steps followed for masking Wyeomyia smithii\_ sequences using Repeat masker and finding the ncRNA secondary structure using SturctRNAFinder package. This document is prepared for Dr.Liz Cooper's Lab, North Carolina Research Campus.

1. Repeat Masker

For Masking process three packages are executed repeatmodeler, repeatmasker and processRepeat.

output file (wyeomyia-families.fa) created from repeatmodeler command is used for RepeatMasker.

Process repeat uses wyeomyia.polished\_small.fasta.cat.all.gz file from RepeatMasker.

repeatMasker's wyeomyia.polished.fasta.tbl output gives the below repeat elements.

SINEs: 0.09 %

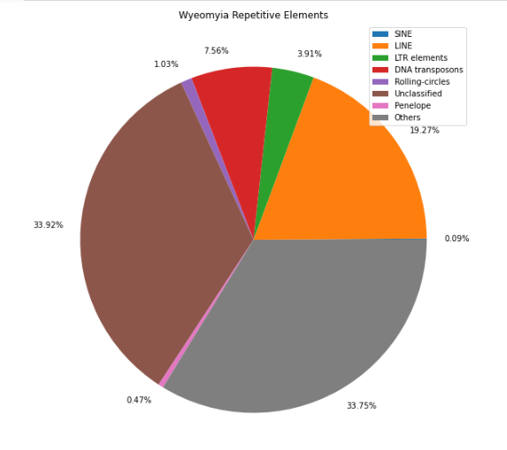
LINEs: 19.27 %

LTR elements: 3.91 %

DNA transposons 7.56 %

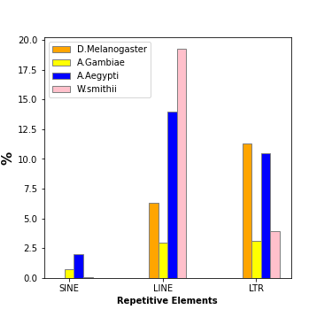
Based on processRepeats's output file wyeomyia.polished.fasta.cat.all.gz.tbl we found that, bases masked: 352626842 bp ( 37.81 %).

The repetitive elements are plotted based on these two files wyeomyia.polished.fasta.tbl.



Also, a comparative plot (Drosophila melanogaster,

Anopheles gambiae, Aedes aegypti, Wyeomyia smithii) for the repetieve elements is generated using python script (Jupyter NB).



2. Find RNA secondary structure using structRNAFinder

2.1. execute the structRNAFinder script

2.2. From the new\_wyeomyia.polished.tab output get the unique family ids and query name to get the unique Htmls/structures.(allQueryNames.txt, copyqueryFiles.ipynb)