

Analysis

- **Stats-output.tsv:** a tsv file containing a high level statistic(number of contigs, number of contigs with pseudo_labels) for different thresholds and all different samples
- **All_Contigs_Analysis:** Includes metrics and Venn diagrams for contigs analyzed using tools (plasforest_pls, plasgraph2_pls, platon_pls, rfplasmid_pls) at thresholds 0, 100, and 1000.
 - **Venn_Diagram:**Shows Venn diagrams depicting contigs predicted as plasmids by different tools at thresholds 0, 100, and 1000.
 - **Metrics:** Provides metrics such as Precision, Recall, F1 score, and Confusion Matrix computed for all tools at thresholds 0, 100, and 1000.
- **PseudoLabel_Analysis:**Contains Analysis for contigs that have a pseudo label(train_mask is 1)
 - **Venn_Diagram:** Venn_diagram of contigs predicted as Plasmids for all the tools for 3 different thresholds(0,100,1000) only considering contigs that have a pseudo label
 - **Metrics:** Contains different metrics(Precision,recall,F1 score, confusion matrix) computed for all available tools for different thresholds(0,100,1000) only considering contigs that have a pseudo label
 - **Features for pseudo labels:** a data frame containing the features(different classification method results) for contigs that have a corresponding pseudo label.In this document “Plasmid” column represents if a contig is a plasmid(1) or chromosome(0), “contig” column represents the modifies contig name and “sample” column shows which sample the contig belongs to.

Note: For tools that produce a non-binary result threshold of 0.5 is used to produce a binary result for metric computation.