Preliminary figures from plasmid clasification benchmarking project

Lu Wang

2025-04-08

CAMI2 benchmarking calculations:

- Accuracy: total number of correct predictions, both positives and negatives, divided by total number of predictions. This predicts overall correctness.

    - (TP + TN) / (TP + TN + FP + FN)

- Precision: accuracy of positive predictions. How many of the predicted positive instances were actually positive?

    - TP / (TP + FP)

- Recall: how well can it identify True Positives out of all the actual positives? True Positive Rate. What proportion of actual plasmids did it detect?

    - TP / (TP + FN)

- Specificity: how well can it identify negative instances out of all actual negative instances?

    - TN / (TN + FP)

- False Positive Rate: proportion of all actual negatives that were classified incorrectly as positives. Fraction of chromosomes incorrectly classified as plasmids.

    - FP / (FP + TN)

- F1-Score - harmonic mean of precision and recall $\rightarrow$ focusing on positives. This is a balanced measure of a model's performance

    - (2 x Precision x Recall) / (Precision + Recall)

## Performance Metrics by Classifier and Assembler
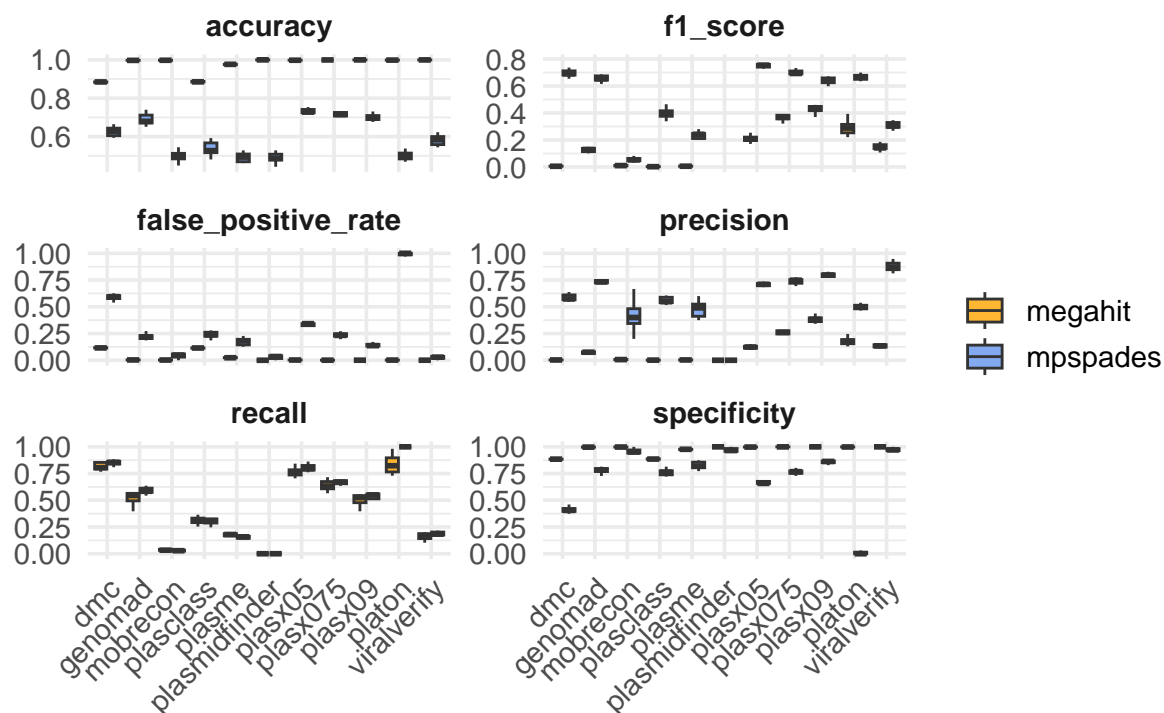
Comparison of plasmid classifier performance across all 9 classification tools. Note: platon has a low specificity/high false positive rate for the metaplasmidSPAdes sample set because platon classified almost all metaplasmidSPAdes contigs (across all 10 samples) as plasmid.

A closer look at the F-1 Score
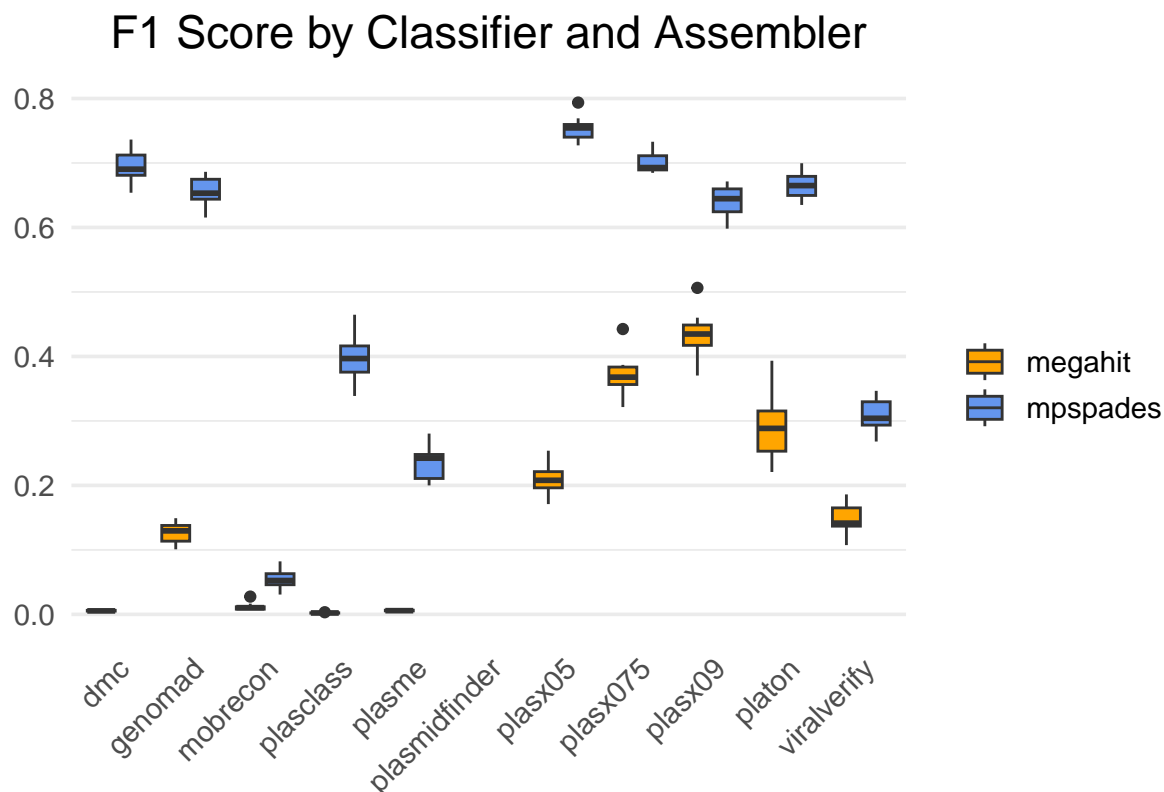
F-1 score for all 9 plasmid classification tools compared against megahit and metaplasmidSPAdes assembly datasets. No output for plasmidfinder because there were no True Positives predicted for either assembly dataset. F-1 score better for all classifiers for metaplasmidSPAdes assembly dataset. Plasx with 0.5 cut off had the highest F-1 score

Precision/Recall Graph for CAMI2 Samples

Precision vs Recall by Assembler and Classifier
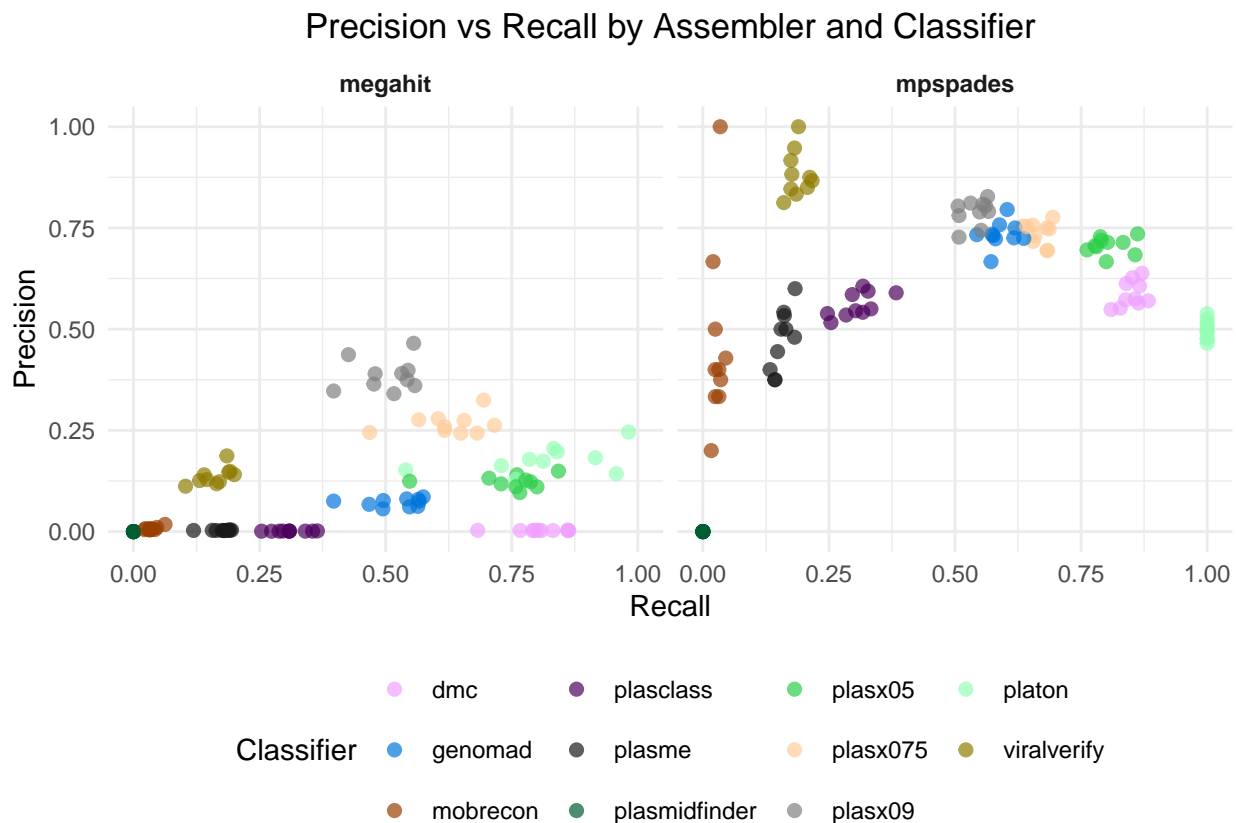
Assembly with metaplasmidSPAdes increases precision for almost all plasmid classifiers (with the exception of plasmidfinder). Caveat: this method includes only circular elements.With megahit assemblies, all tested plasmid classifiers (except plasx) had precision scores under 0.25. This means that < 25% of predicted plasmids were actually plasmids. The best performing cluster are the plasx/genomad cluster in the metaplasmidSPAdes assembly sample set. But would depend on whether you favor higher recall or higher precision.

Precision and recall mean for genomad, plasx 0.75, and plasx 0.9 (metaplasmidSPAdes assemblies)

```
## # A tibble: 8 x 4
## # Groups:   classifier [4]
##   classifier assembler mean_precision mean_recall
##   <chr>      <chr>              <dbl>       <dbl>
## 1 genomad    megahit           0.0720       0.521
## 2 genomad    mpspades          0.734        0.591
## 3 plasx05    megahit           0.123        0.747
## 4 plasx05    mpspades          0.707        0.805
## 5 plasx075   megahit           0.266        0.627
## 6 plasx075   mpspades          0.737        0.667
## 7 plasx09    megahit           0.387        0.503
## 8 plasx09    mpspades          0.789        0.540
```

plasx is the highest performing plasmid classifier for both megahit and metaplasmidSPAdes assembly

datasets. For the CAMI2 marine dataset, the best plasmid classification protocol would be to assemble with metaplasmidSPAdes and then verify results with plasx.

Radar plots just for fun
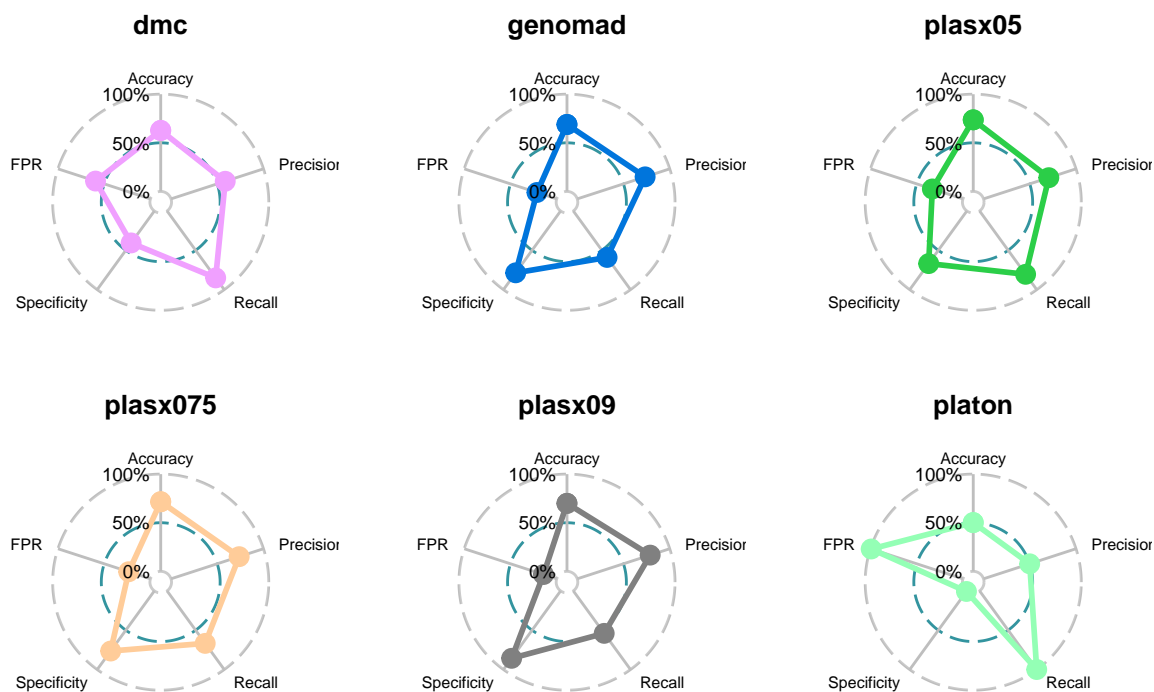
Radar plots of top performing plasmid classifiers (DeepMicroClass (dmc), genomad, plasX (0.5, 0.75, 0.9 confidence score cutoffs), and platon), against metaplasmidSPAdes assemblies