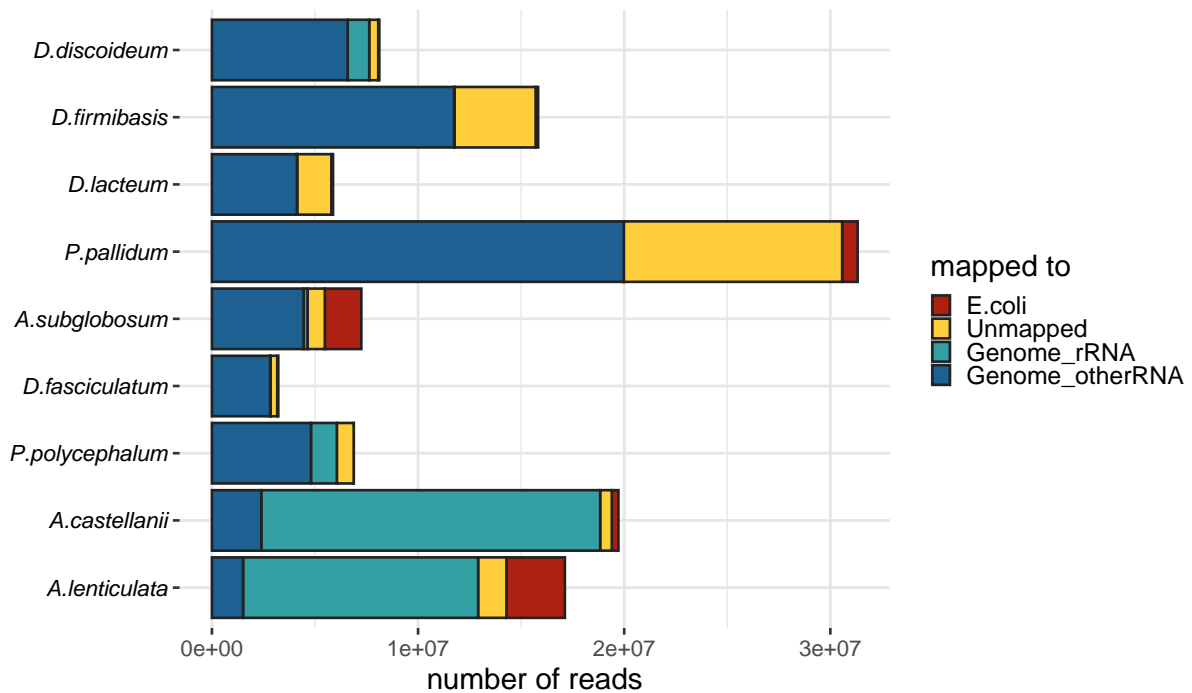


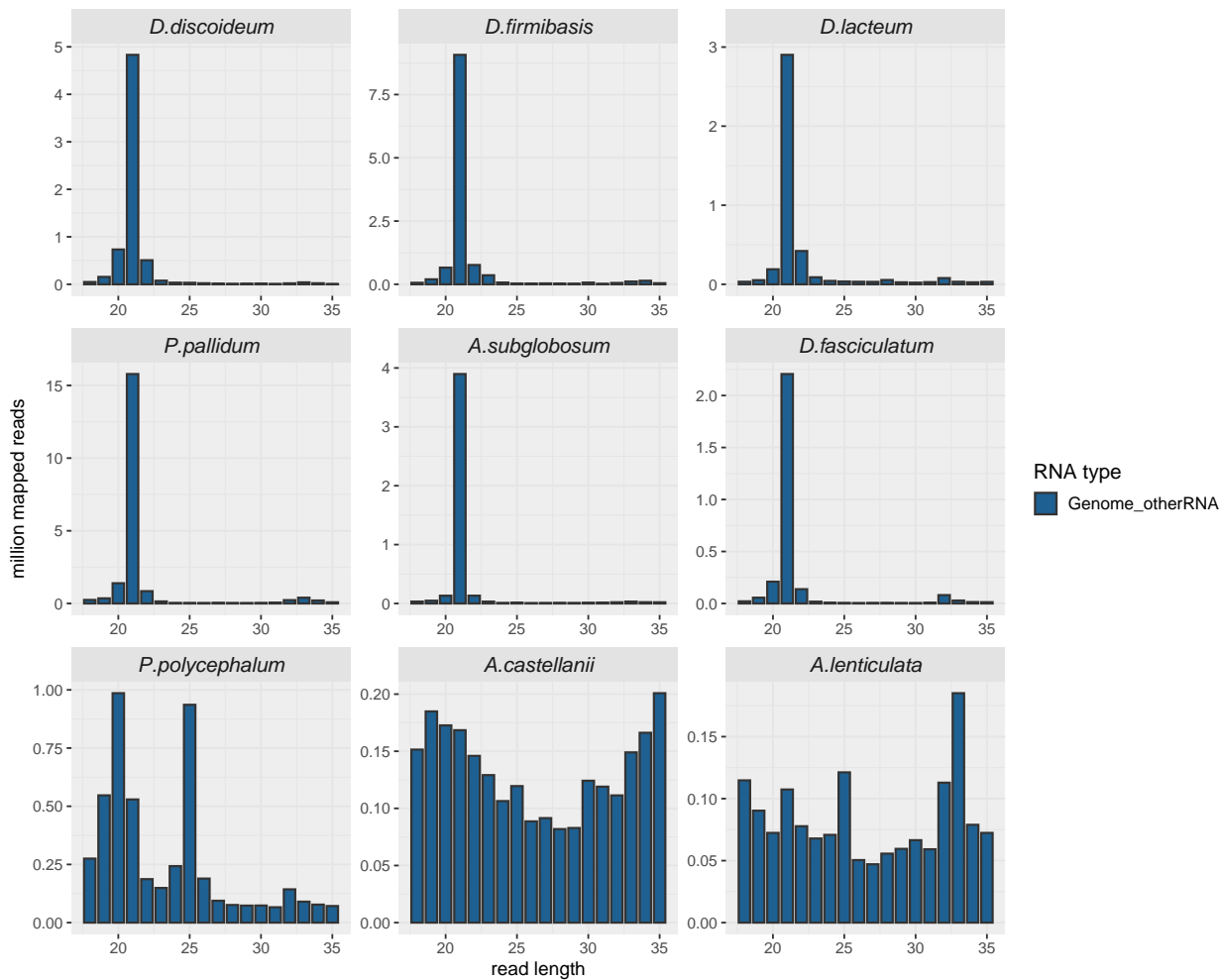
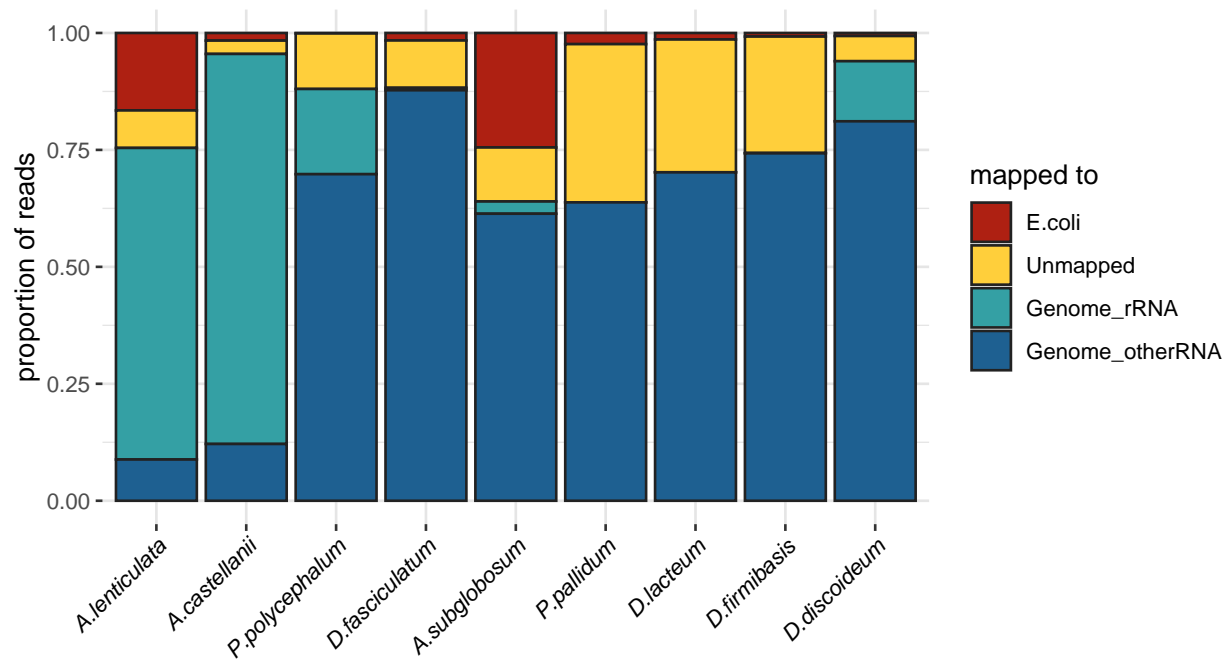
# Analysis of microRNAs identified in Amoebzoa

Bart Edelbroek

11/24/2021

Quantified the number of reads mapping to different parts of the amoeba genome, E. coli, or unmapped. Plotted the length profile of the reads mapping to the genome but not rRNA. The data\_in was generated using mapping\_percentages.sh and extract\_length\_profiles.sh helper scripts.





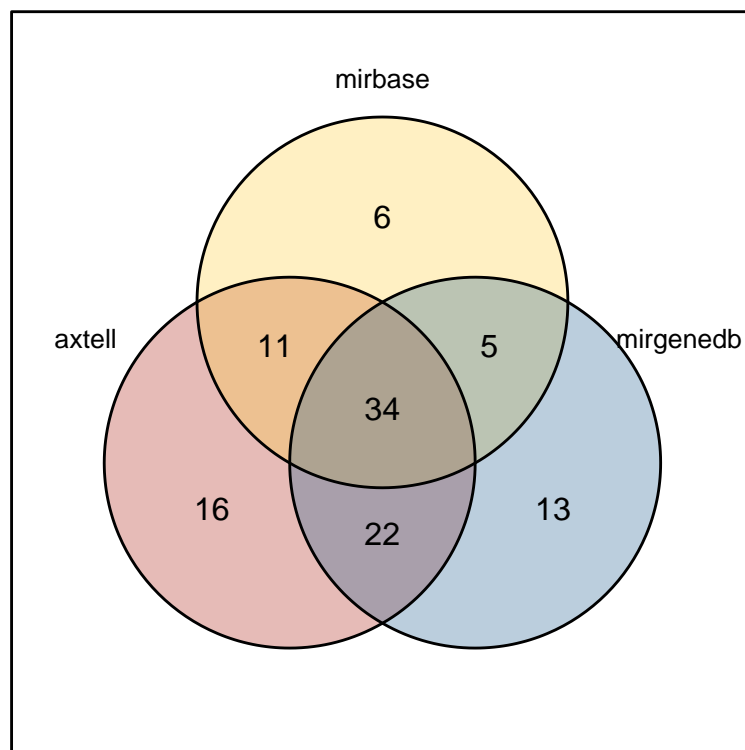
After running the miRNA curation python script, the characteristics of each of the miRNA candidates is summarized in data\_in/species/combined\_analysis.tsv. Each of the miRNA candidates is analyzed by the criteria put forth in Axtell & Meyers (2018), Fromm et al. (2022), and Kozomara et al. (2014 & 2019), as summarized in figure S3. Following manual curation, a final list of high-confidence miRNAs is identified.

```
## # A tibble: 70 x 5
##   miRNA.ID      cluster      axtell_pass mirbase_pass mirgenedb_pass
##   <chr>         <chr>         <chr>         <chr>         <chr>
## 1 acas_Cluster_5138(+) acas_Cluster_51~ TRUE          TRUE          TRUE
## 2 alen_Cluster_556(+) alen_Cluster_55~ 50% precis~ TRUE          TRUE
## 3 alen_Cluster_1355(+) alen_Cluster_13~ TRUE          3 miR* reads TRUE
## 4 alen_Cluster_2130(+) alen_Cluster_21~ 50% precis~ TRUE          TRUE
## 5 alen_Cluster_2233(-) alen_Cluster_22~ TRUE          TRUE          TRUE
## 6 alen_Cluster_4026(-) alen_Cluster_40~ TRUE          TRUE          TRUE
## 7 asub_Cluster_2821(-) asub_Cluster_28~ TRUE          4 miR* reads TRUE
## 8 asub_Cluster_3219(+) asub_Cluster_32~ TRUE          6 miR* reads TRUE
## 9 asub_Cluster_3317(-) asub_Cluster_33~ TRUE          5 miR* reads TRUE
## 10 asub_Cluster_3339(-) asub_Cluster_33~ TRUE          7 miR* reads TRUE
## # ... with 60 more rows
```

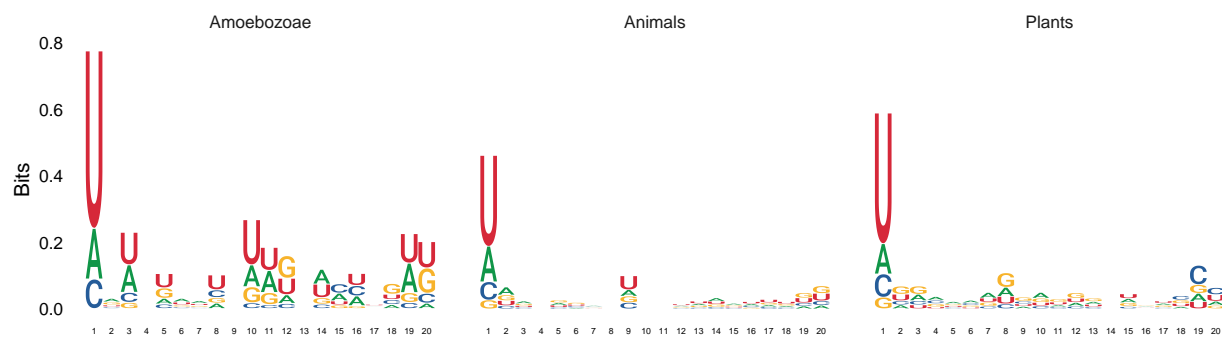
The miRNA sequences of the confirmed candidates are output in fasta files, as well as in a table. The genomic locations of the miRNAs are added and used to generate a gff file to annotated the genomes.

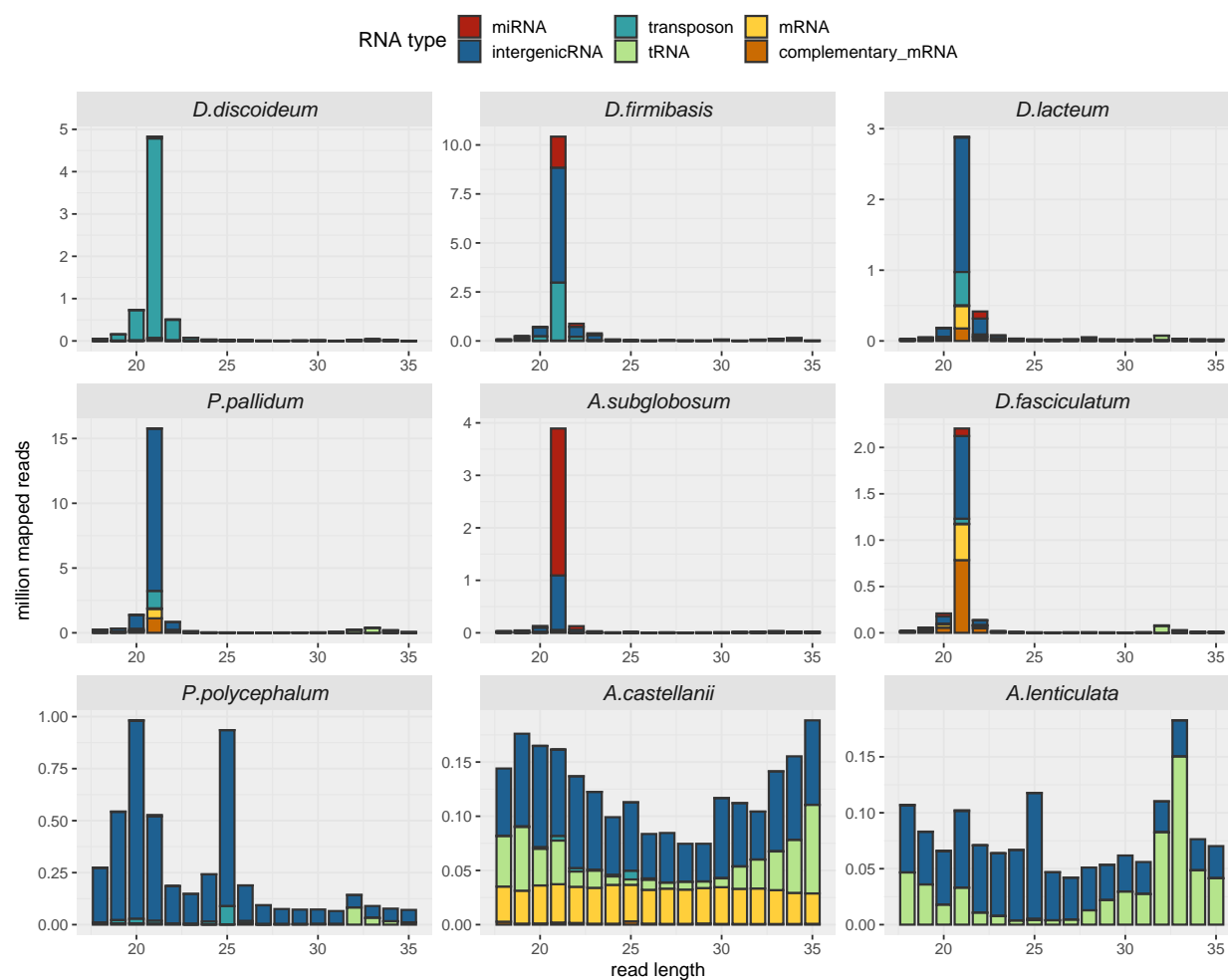
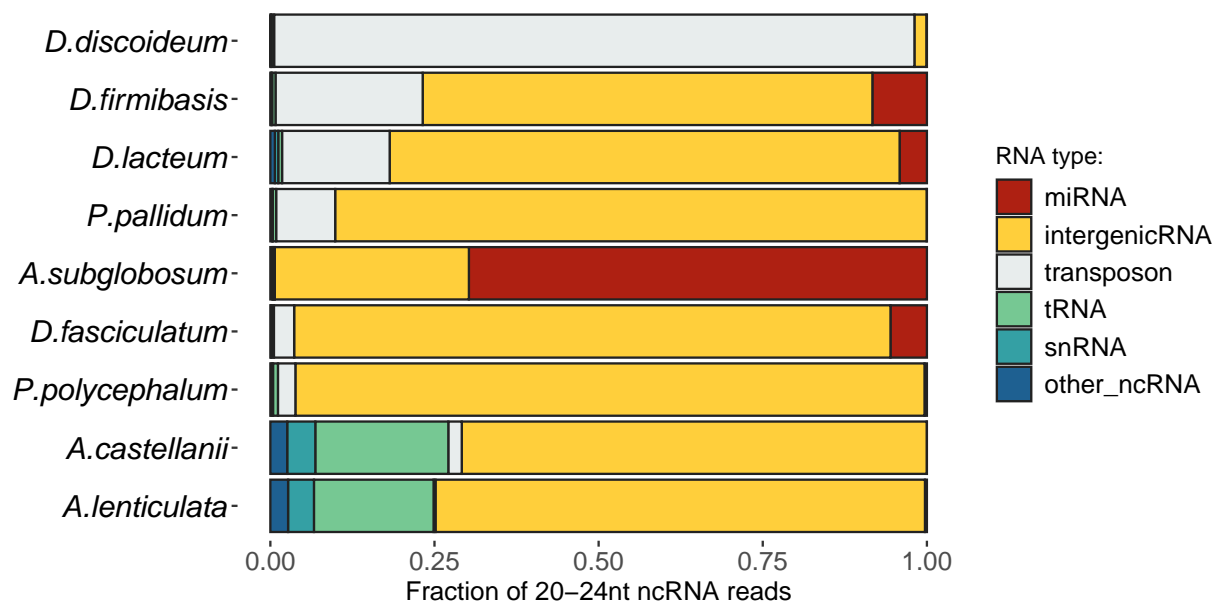
```
## # A tibble: 10 x 5
##   species  combined axtell_pass mirbase_pass mirgenedb_pass
##   <chr>    <int>    <int>    <int>    <int>
## 1 acas      1        1        1        1
## 2 alen      5        4        5        5
## 3 asub     22       23       17       24
## 4 ddis      8       10        8        6
## 5 dfas      4        9        4        4
## 6 dfir      5        9        5        7
## 7 dfir_new  9       13        9       11
## 8 dlac      2        2        1        2
## 9 ppal      4        7        3        5
## 10 ppol      7       10        6       11

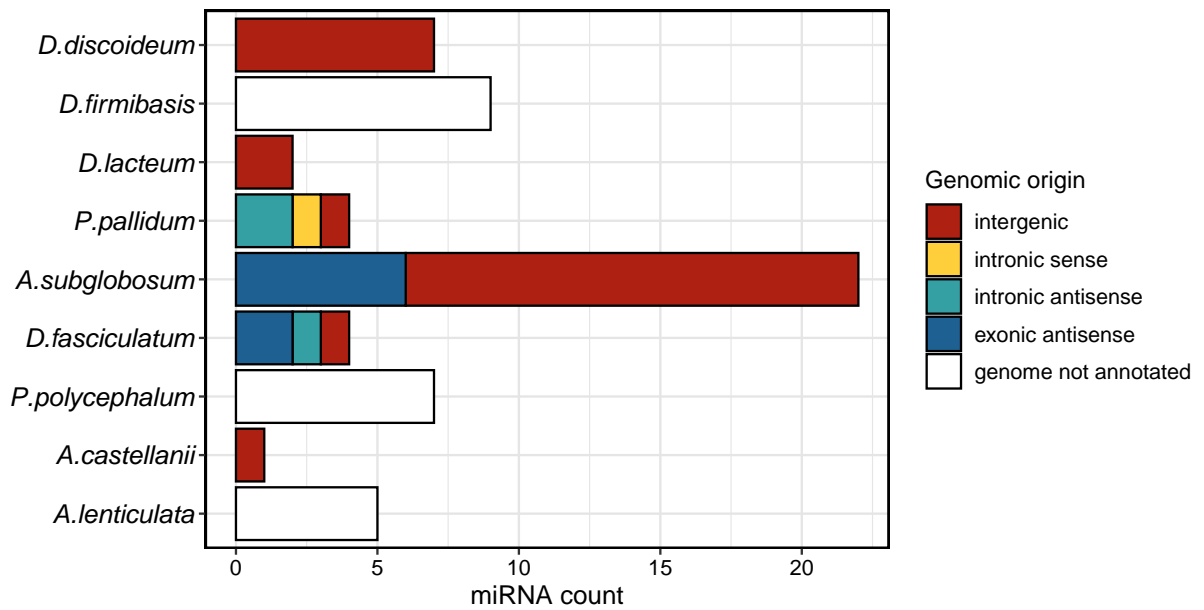
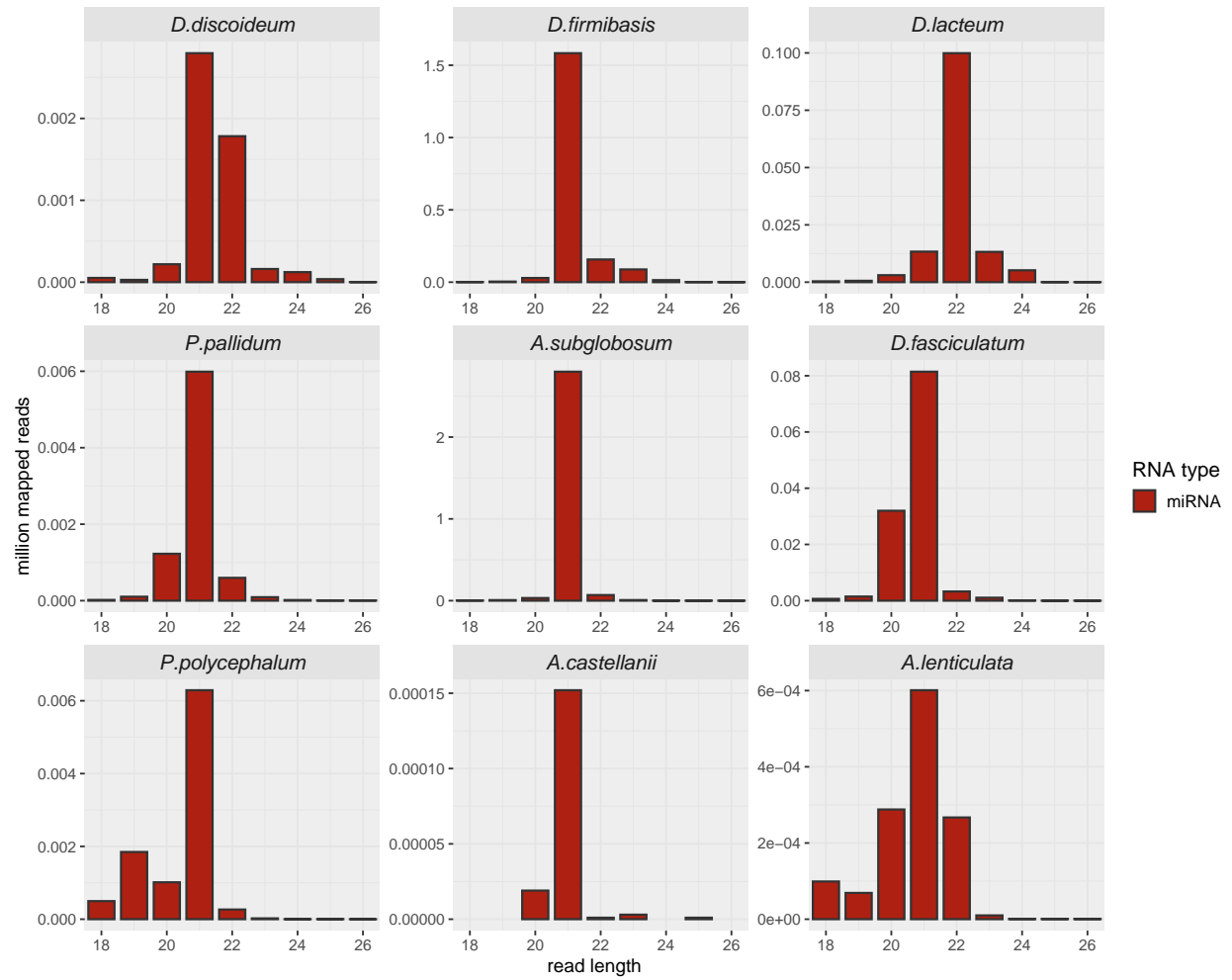
## pdf
## 2
```

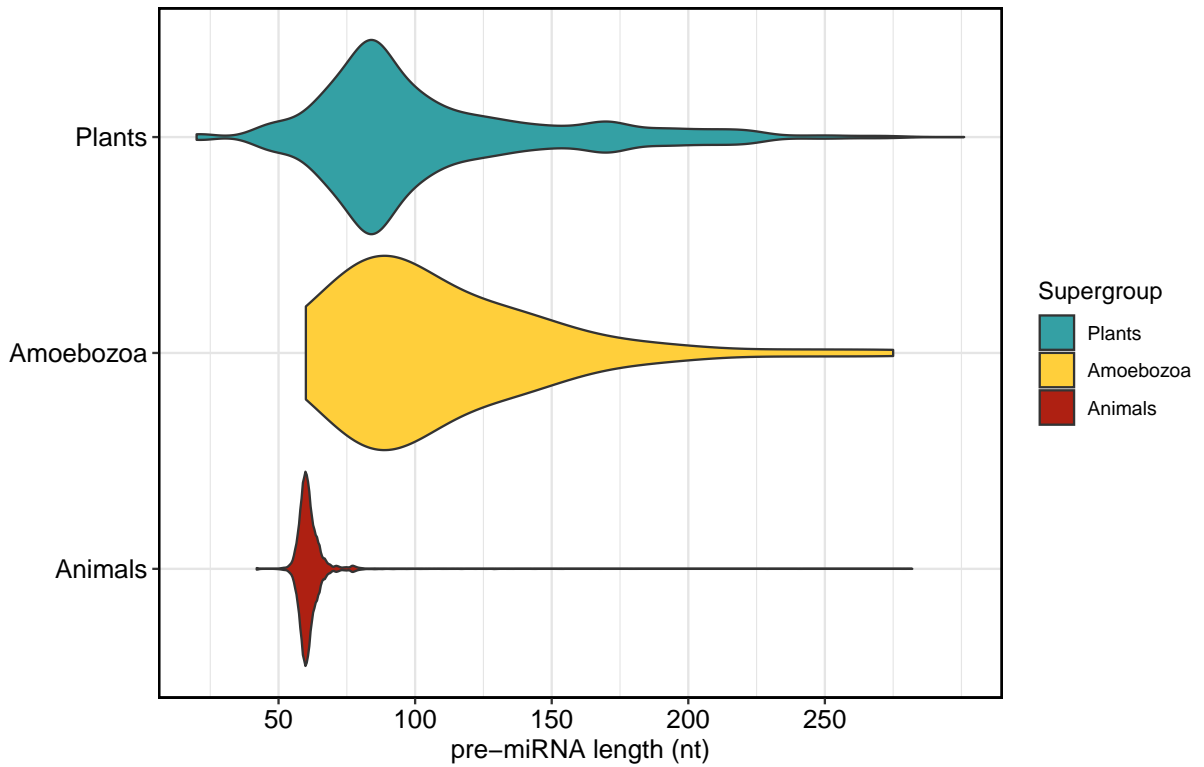
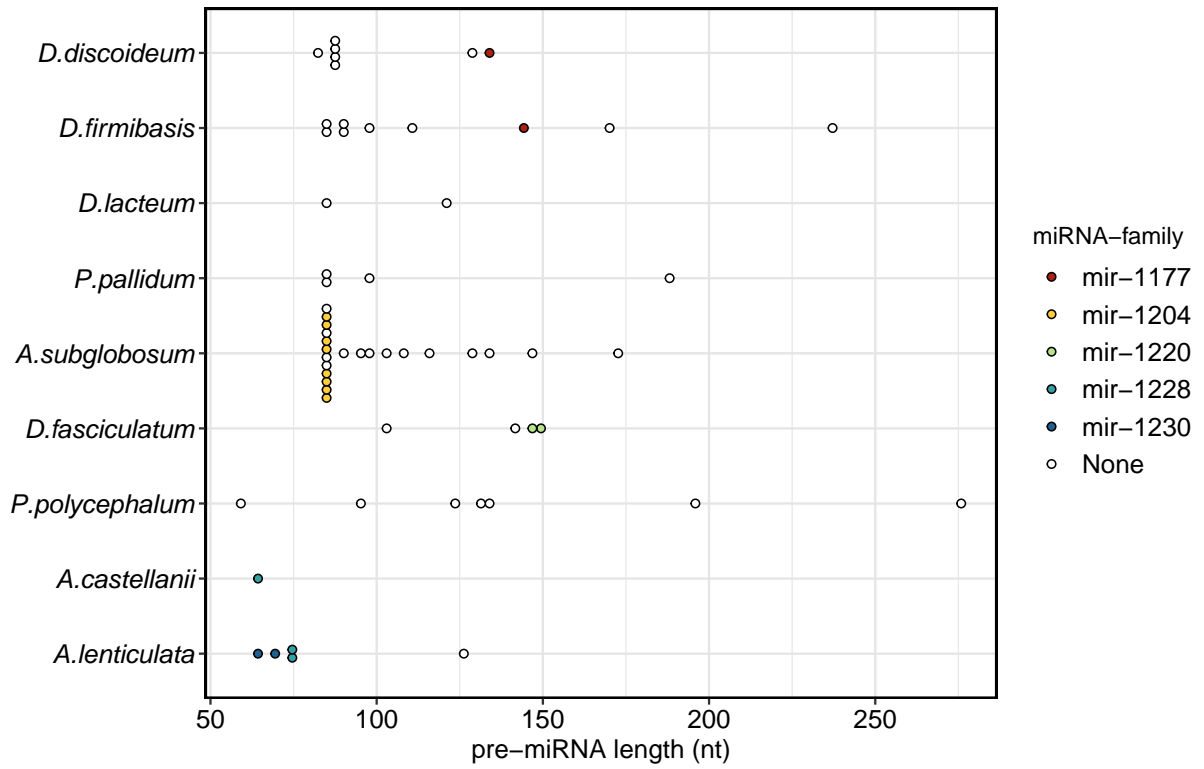


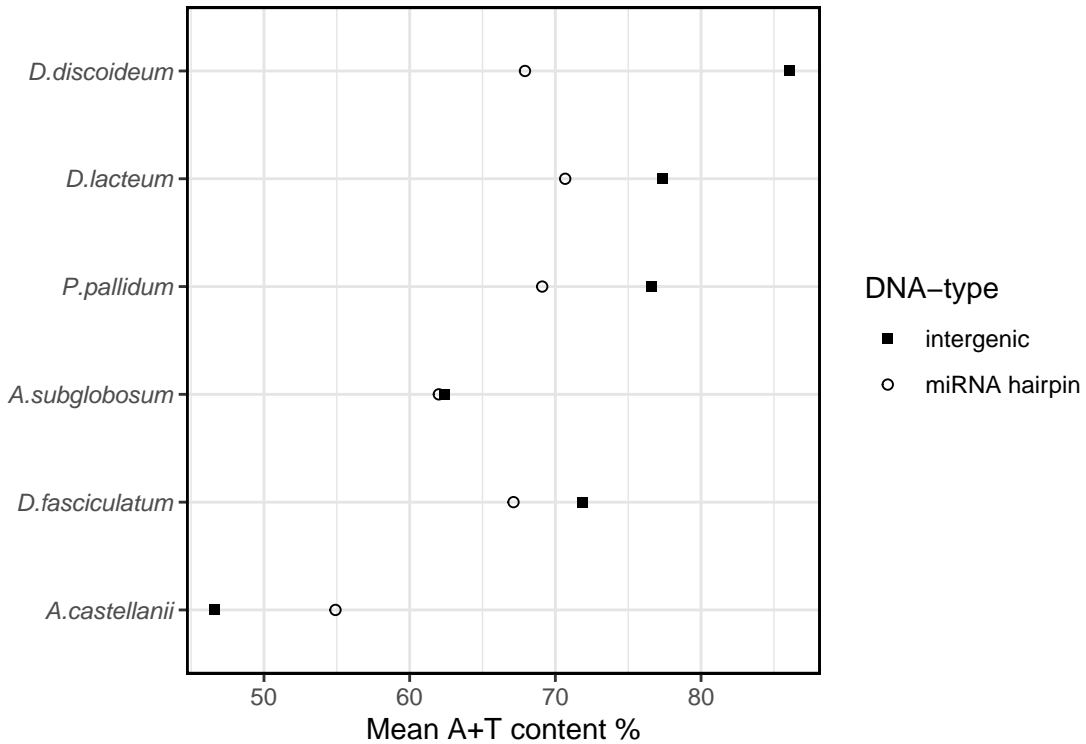
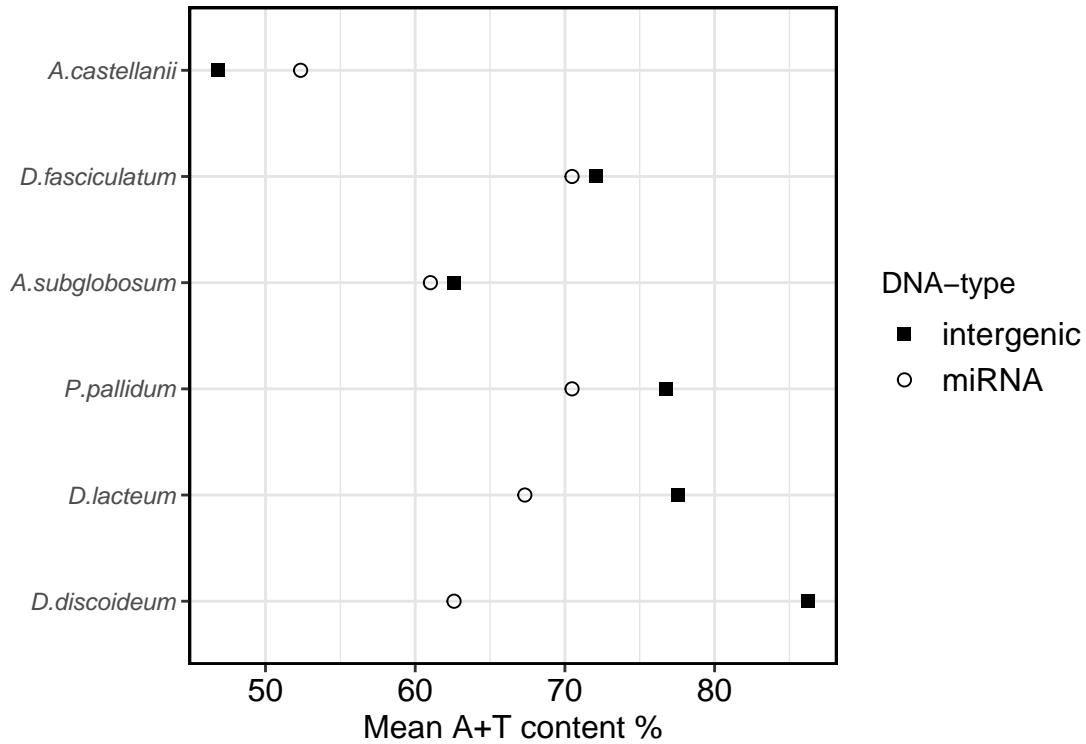
Following identification of the miRNAs, the characteristics of the miRNAs are analyzed. For those characteristics where it is relevant, the comparison is made with Plants and Animals. Plant and Animal miRNA sequences were accessed from PmiREN 2.0 and MirGeneDB 2.1 respectively.





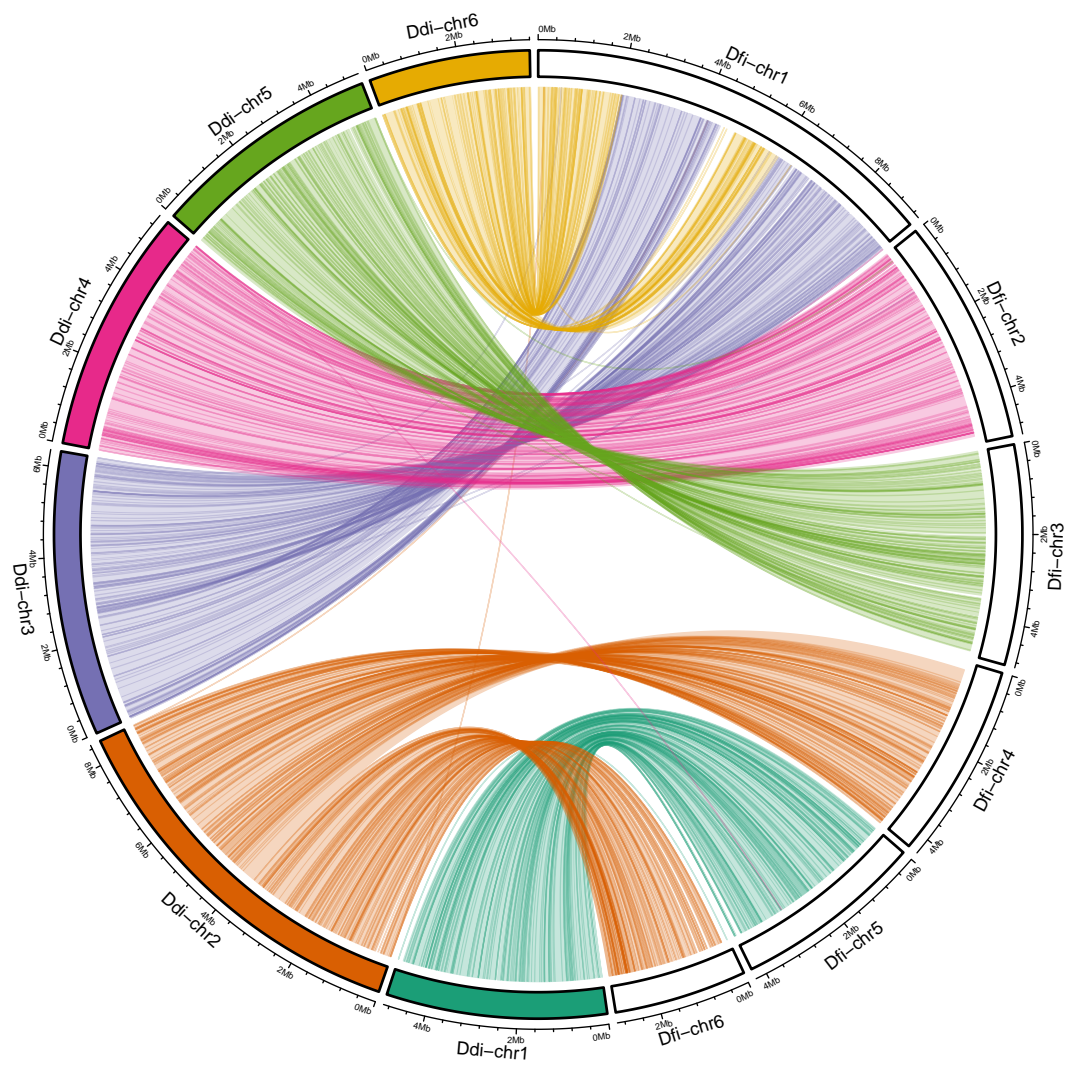






Using Satsuma2, the synteny blocks between *D. firmibasis* and *D. discoideum* were identified, with the results summarised in `data_in/satsuma_synteny.out`. Here, the synteny blocks are linked together if they are within 5000nt of each other. The first circos plot shows all the synteny that was identified between the two genomes; the second plot shows only those synteny blocks that contain a miRNA on either genome, with the label identifying which miRNA is on the region.





```
## null device
##      1
```

```
## pdf
##      2
```