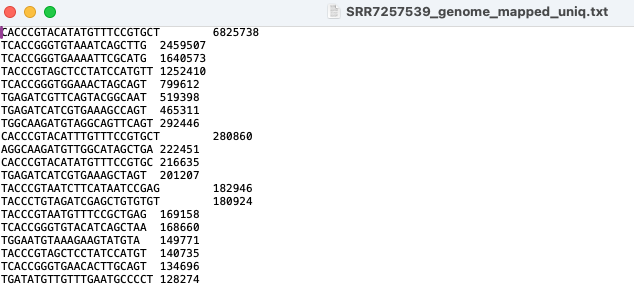
Python scripts for determining if isomiRs are templated or untemplated:

**Background:**

1. We had miRNA sequencing data from the parasite Fasciola hepatica. This miRNA seq data was processed according to the paper (<https://doi.org/10.1080/15476286.2022.2099646>). miRNA seq data as aligned to the parasite genome, then all aligned reads were collapsed into a “tag file”, which looks like this:



It is similar to a count file. However, it has the “tag” or read sequence instead of the gene name.

1. This file was then used as the input file for IsomiRSEA (<https://eda.polito.it/isomir-sea/>). This program identified the isomiRs and classified them.

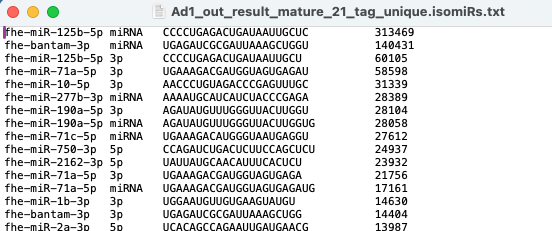
Run script: *isomiR-SEA\_run\_script.py*

1. From this output, we only used the “out\_result\_mature\_22\_tag\_unique.txt” file.
2. We then used the custom perl scripts from the paper above paper (<https://github.com/gppbioinfo/isomiR_celegans>). These perl scripts filtered the output files for type of end variation (3p, 5p, both) and the number of nucleotides difference.

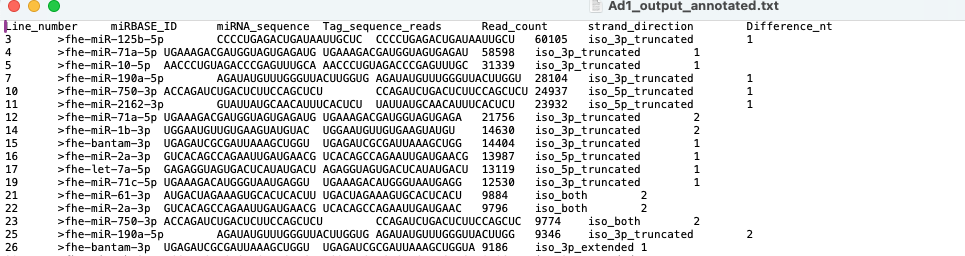
Run script: *getisomiR.pl (On papers Github)*

Run script: *isomiR\_tagging.pl (On papers Github)*

1. The output from this looks like this and was then summarised in excel
   1. Get isomiR



* 1. Tagging (I only really used this file for downstream analysis)

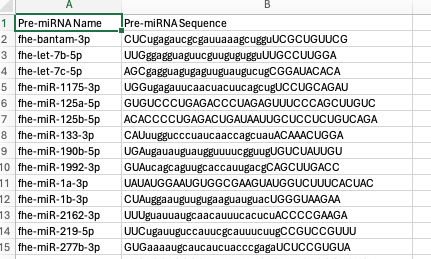


**Templated vs untemplated**

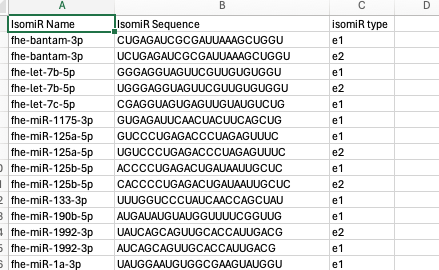
From there, I then determined if these were templated or untemplated at each position in the isomiR sequence using custom python script.

For this you need to know the pre-miRNA sequence, for each miRNA I only used 3nt upstream and 10nt downstream of the mature miRNA sequence to align to.

1. Input files need to be created:
   1. Pre\_miRNA.csv = miRNA name and pre-miRNA sequence (3nt upstream and 10nt downstream canonical mature). These sequences were obtained from miRBase/manually edited.



* 1. isomiR\_3.csv or isomiR\_5.csv = I processed 3 end and 5 end isomiRs separately. These file contains the miRNA name, isomiR sequence and isomiR type. The isomiR type is based on the type of alteration and can be found in “key for isomiR type.xlxs”. This helps determine the start position for the isomiR sequence to start aligning to the pre-miRNA.



1. Run script: *templated.py*
   1. Output file will be csv (output\_5.csv), this was then converted to tab-deiminated txt file (summary\_input\_5.txt).

A screenshot of a table

Description automatically generated

You could leave this hear but I want to summarise further.

1. Run *templated\_summary.py*
   1. Output: output\_templated\_summary\_5.txt
2. Run *nt\_summary.py*
   1. Output: output\_nt\_summary\_5.txt

This data I then manually processed in excel to get count files for templated or untemplated at each position, the used ggplot barplot to produce a figure like this:

