Python Documentation

version

July 11, 2023

Contents

MAFtoolbox User Manual

MAFtoolbox is a Software written in Python that implements a range of operations and transformations on genome alignments in the Multiple Alignment Format (MAF). Examples of use case include the extraction of alignment subblocks based on gene annotations, filtering of sequences based on identity and merging of fragmented neighboring alignment blocks into lnger, coherent blocks.

Note

This project is currently under development.

Indices and tables

- **genindex**
- **modindex**
- **search**

Installation

While a bioconda installation is planned for future release, right now it is heavily suggested to use the precompiled executable built with pyinstaller. Download and unpack the archive, then navigate to the distributed binaries folder with:

```
cd MAFtoolbox/dist/MAFtools
```

You can test the executable with:

```
./MAFtools --help
```

This should produce a short list of programs executable with MAFtools. To get more information on one (here, as an example, for extracting alignment blocks according to genome coordinates), you can type:

```
./MAFtools extract --help
```

As it is obviously annoying to only use MAFtools from inside the download directory, I would suggest setting up an alias for now, like this:

```
alias MAFtools=$(pwd)/MAFtools
```

As mentioned, a full installer automating this process will follow.

Example usage

The MAFtoolbox comes with a few example files, that can be used to play around and get to know basic functionality. One useful application might be highlighting the part of an alignment that includes some annotated gene. The program

```
MAFtools highlight
```

exists for this purpose. In the most simple case, we can simply provide a MAF alignment file and a corresponding annotation file in bed format. Note that the sequence names (for example chrX, chrY…) in the annotation file need to exactly correspond with the sequence names found in the alignment file to be found. As a simple showcase, move to the root directory of the MAFtoolbox archive and type

```
MAFtools highlight --maf Examples/Apoidea_genome_tRNA_blocks_filtered.maf --bed Examples/Ame
```

The output should display the alignments in MAF format, but with the coordinates corresponding to the genes found in the .bed file highlighted in green. What if we want to highlight the annotated genes -and- an additional 5 nucleotides (with respect to the reference sequence) in both directions? We can use the -s (–sense) and -n (–antisense) parameters to add any number of nucleotides to be highlighted:

```
MAFtools highlight --maf Examples/Apoidea_genome_tRNA_blocks_filtered.maf --bed Examples/Ame
```

You will notice that the highlighted regions are now enlarged corresponding to the -s and -n parameters, but these overhang regions will be colored the same way as the annotated sequence. To allow for visual distinction we can give the overhang regions another color, for example red:

```
MAFtools highlight --maf Examples/Apoidea_genome_tRNA_blocks_filtered.maf --bed Examples/Ame
```

Now it should be readily visible what is what.

You can always explore all program options and parameters with the –help function.