

How to configure Y-STR Kit

This document helps to configure the y-STR Kit to have accurate results. The key configuration file is ystr_info.txt.

Comments

All lines with comments begin with '#'.

Titles

If a set of Y-STRs are required to be grouped, it can have a title tag and all Y-STR configs under that title will be grouped.

Y-STR Configs

Each Y-STR configuration is a set of 6 columns separated by 'tab':

1. YSTR Name
2. Begin Position
3. End Position
4. Motif Regular Expression
5. Filter Regular Expression
6. Replace Method

1. Y-STR Name Column

As the name suggests, it is just the name of Y-STR.

2. Begin Position

This is the begin position of the Y-STR. This is not required to be accurate but make sure the entire repeats along with some non-repeats are captured.

3. End Position

This is the end position of the Y-STR. This is not required to be accurate but make sure the entire repeats along with some non-repeats are captured.

4. Motif Regular Expression

The motif regular expression is used to count the actual repeats.

5. Filter Regular Expression

The filter regular expression is used to filter the actual repeats.

6. Replace Method

Replace method is an internal step to decide what to do after the filter step and how to count the motif. Only 'CONCAT' is supported now.

Example

YSTR	BeginPos	EndPos	MotifRegex	FilterRegex	ReplaceMethod
DYS393	3131128	3131246	AGAT	(AGAT){7,18}	CONCAT

Consider DYS393. From all the bases between positions 3131128 to 3131246, it filters/cuts the bases based on FilterRegex. Then, using MotifRegex, it simply counts and that's the value of Y-STR.

E.g.,

ATGCCTGCCAAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATGGTATAC

Based on FilterRegex, it cuts as follows.

AGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT

Then, based on MotifRegex, it counts the number of AGAT, which is 10.