

Background

Other tools like ExpansionHunter, STRetch, etc.

- Focused on known pathogenic TRs
- Only use enclosing reads
- Limited to 6 bp motifs
- Computationally expensive

See Table 1.



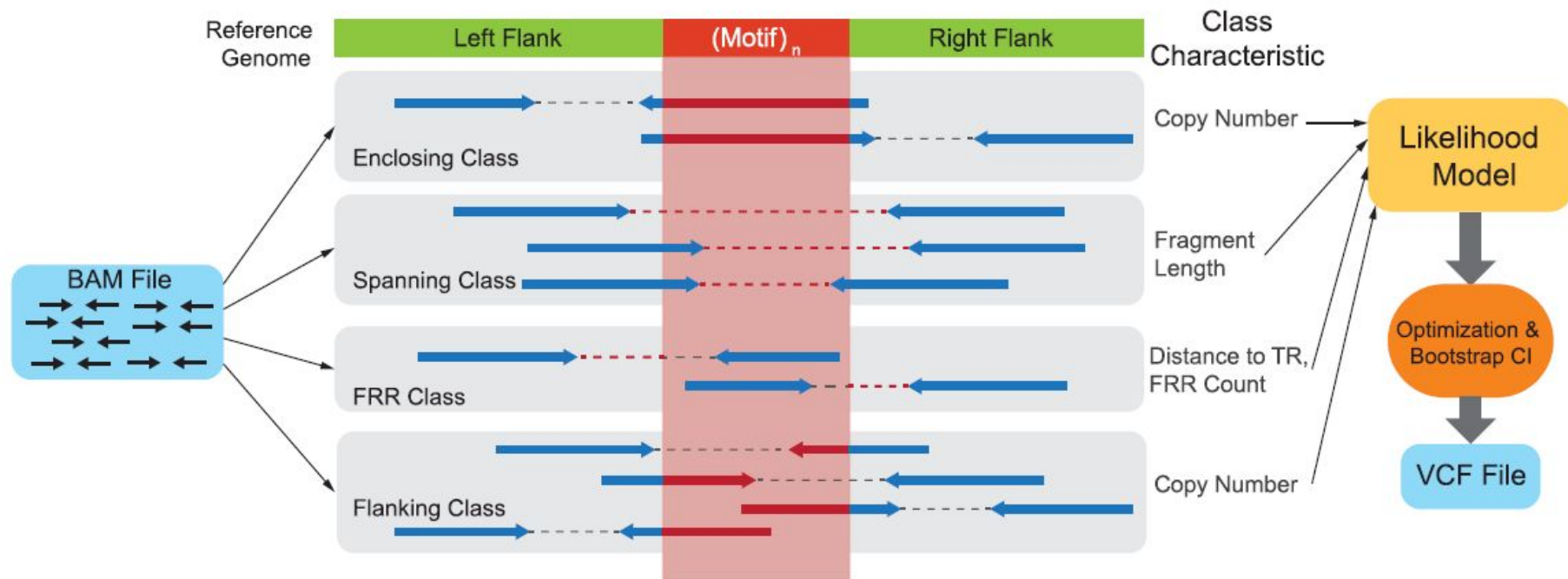
G a n g S T R

Likelihood Model

“general statistical model incorporating multiple properties of paired-end reads into a single maximum likelihood framework capable of genotyping both normal length and expanded repeats”

- Formulas (1), (3), and (5)
- Decompose likelihood into informative contributions for read pairs and Poisson count of observed fully repetitive reads
- Decompose informative aspect into a class probabilities and read (characteristic) probabilities

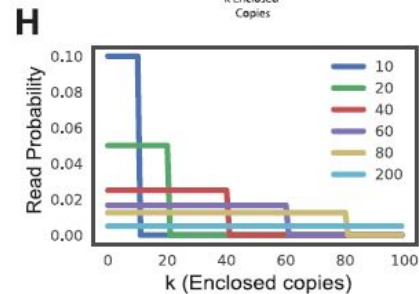
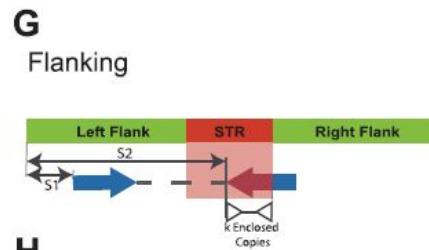
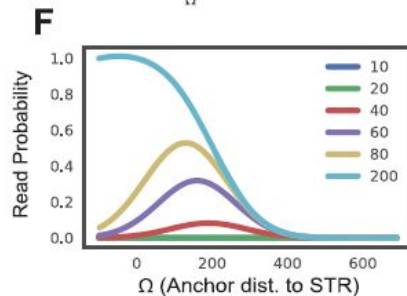
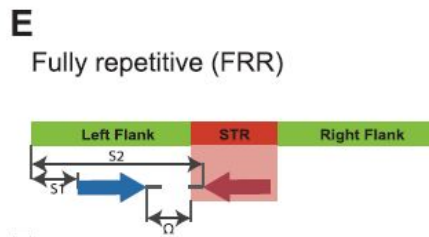
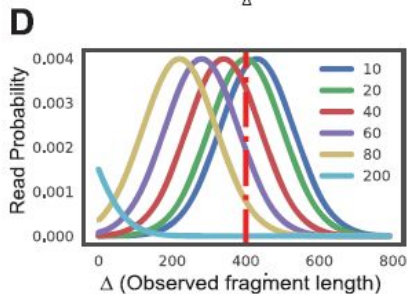
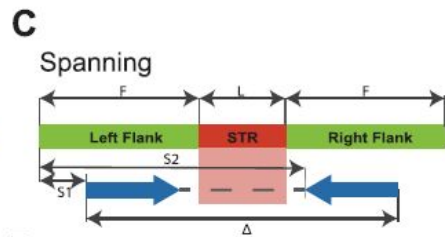
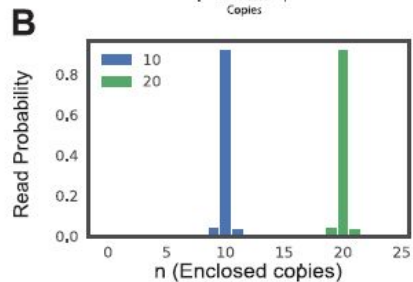
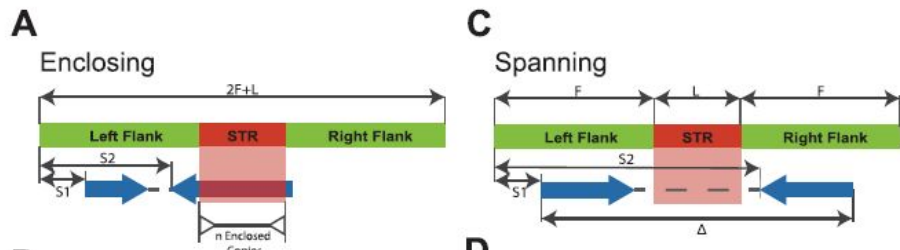
Informative Classes



Brainstorm.

How would you use an
informative class to
est. TR expansions?

Read Probabilities



Discussion.

What advantages
and/or challenges do
you see for this
statistical approach?