**Documentation for BFB Simulations**

Lena Krockenberger 8/17/22

**BFBSimulator Parameters:**

usage: BFBSimulator.py num\_cycles chr\_str start\_pos end\_pos genome

[-h] [-l L] [-ld {expo,simple}] [-df DF] [-dl DL]

[-dd {expo,simple}] [-xf XF] [-xl XL]

[-xd {expo,simple}] [-o O] [-rev] [-p P] [-b B]

[-bl BL]

**positional arguments:**

num\_cycles Number of BFB cycles to simulate

chr\_str Name of the chromosome to simulate BFB on

start\_pos Start position of the genome section to simulate BFB on (0-indexed, inclusive)

end\_pos End position of the genome section to simulate BFB on (0-indexed, exclusive)

genome FASTA file of the genome

**optional arguments:**

-h, --help show this help message and exit

-l L Mean length of the BFB segments (default 600000)

-ld {expo,simple} Distribution of the BFB segment lengths. "expo"(default): exponential; "simple": all use the same value as the mean length

-df DF Probability for each BFB segment to have one deletion (default 0.1)

-dl DL Mean deletion length of the BFB segments (default 50000)

-dd {expo,simple} Distribution of the BFB segment deletion lengths. "expo"(default): exponential; "simple": all use the same value as the mean deletion length

-xf XF Probability for each BFB segment to have one duplication (default 0.1)

-xl XL Mean duplication length of the BFB segments (default 10000)

-xd {expo,simple} Distribution of the BFB segment duplication lengths. "expo"(default): exponential; "simple": all use the same value as the mean duplication length

-o O Prefix of the output files

-rev If this option is specified, BFB simulation will append string to the 5' side of the region

-p P For each segment, the probability that deletion is simulated before duplication (default 0.5)

-b B Probability that there is a deletion when appending duplicated segment. (default 0.1)

-bl BL Length of deletion when appending duplicated segment. (default 50000)

**Simple Case** - (low segment count, low copy number, no indels, no end deletions):

**Ex with ERBB2:**

num\_cycles 5

chr\_str 17

start\_pos 33600000

end\_pos 36600000 (add 3,000,000 to start\_pos)

genome new\_hg38.fa (fasta file name)

-l L default

-ld {expo,simple} default

-df DF 0 (don’t want deletions)

-dl DL 0 (don’t want deletions)

-dd {expo,simple} simple

-xf XF 0 (don’t want duplications)

-xl XL 0 (don’t want duplications)

-xd {expo,simple} simple

-o O outputs/simple/ERBB2 (can add file paths to output prefixes)

-rev defualt

-p P 0 (no deletions)

-b B 0 (no end deletions)

-bl BL 0 (no end deletions)

**Final command:**

./BFBSimulator.py 5 17 33600000 36600000 new\_hg38.fa -df 0 -dl 0 -dd simple -xf 0 -xl 0 -xd simple -o outputs/simple/ERBB2 -p 0 -b 0 -bl 0

**Intermediate Case** - (regular segment number and copy number, end deletions, no indels):

**Ex with ERBB2:**

num\_cycles 10 (larger number of cycles increases copy number)

chr\_str 17

start\_pos 33600000

end\_pos 39600000 (add 6,000,000 to start\_pos)

genome new\_hg38.fa (fasta file name)

-l L default

-ld {expo,simple} default

-df DF 0 (don’t want deletions)

-dl DL 0 (don’t want deletions)

-dd {expo,simple} simple

-xf XF 0 (don’t want duplications)

-xl XL 0 (don’t want duplications)

-xd {expo,simple} simple

-o O outputs/end\_Deletions/ERBB2 (can add file paths to output prefixes)

-rev default

-p P 0 (no deletions)

-b B 1 (always simulate end deletion when appending new segment)

-bl BL 50000

**Final Command:**

./BFBSimulator.py 10 17 33600000 39600000 new\_hg38.fa -df 0 -dl 0 -dd simple -xf 0 -xl 0 -xd simple -o outputs/end\_Deletions/ERBB2 -p 0 -b 1 -bl 50000

**Complex Case** - (regular segment number and copy number, end deletions, indels):

**Ex with ERBB2:**

num\_cycles 10

chr\_str 17

start\_pos 33600000

end\_pos 39600000 (add 6,000,000 to start\_pos)

genome new\_hg38.fa (fasta file name)

-l L default

-ld {expo,simple} default

-df DF default

-dl DL default

-dd {expo,simple} default

-xf XF default

-xl XL default

-xd {expo,simple} default

-o O outputs/complex/ERBB2 (can add file paths to output prefixes)

-rev default

-p P default

-b B default

-bl BL default

**Final Command:**

./BFBSimulator.py 10 17 33600000 39600000 new\_hg38.fa -o outputs/complex/ERBB2

**Script for Running Simulations** – simulations.sh

Usage: ./simulations.sh

Parameters: none

When running simulations.sh, 5 different oncogenes are simulated through each of the 3 cases (simple, intermediate, and complex). The 5 genes include: ERBB2, MYC, EGFR, BCL2, and PIK3CA.

Simple case: 1 simulation per gene with the same parameters for “simple case” detailed above (5 simulations total)

Intermediate case: 3 simulations per gene with the same parameters for “intermediate case” detailed above (15 simulations total)

Complex case: 2 simulations per gene with the same parameters for “complex case” detailed above (10 simulations total)

In total, simulations.sh runs 30 simulations of varying complexity.

To modify details of the simulations, the script must be edited.