

By running Smash++, positions of similar regions in reference and target sequences, relative redundancy and also redundancy (complexity) of the regions is saved in a “.pos” file. This tab-delimited file has a header including:

1. The string “##SMASH++” as a specifier for the Smash++ tool,
2. The “PARAM” line to list the parameters used to generate the position file,
3. The “INFO” line to provide the names and sizes of reference and target sequences,
4. The line with the names of columns of the body,

and a body including:

1. RBeg (Reference Begin): beginning of a reference region,
2. REnd (Reference End): end of a reference region,
3. RRelRdn (Reference Relative Redundancy): relative redundancy obtained by compressing the associated target block considering as reference this reference block,
4. RRdn (Reference Redundancy): redundancy (Complexity) of the detected reference block, calculated by reference-free compression,
5. TBeg (Target Begin): beginning of a target region,
6. TEnd (Target End): end of a target region,
7. TRelRdn (Target Relative Redundancy): relative redundancy obtained by compressing the associated reference block considering as reference this target block,
8. TRdn (Target Redundancy): redundancy (Complexity) of the detected target block, calculated by reference-free compression,
9. Inv (Inverted repeat): if the corresponding line is an inverted repeat. “F” (False) means it is regular and “T” (True) means it is inverted.

As an example, the header of the following “.pos” file shows that Smash++ was run by the parameters “-r dataset/REF -t dataset/TAR -l 0 -m 1000” and also the reference “REF” and the target “TAR” are 50,000 base long. The body shows that there is a reference block, from the position 5000 up to 10000, similar to a target block, from the position 2000 to 7000. Relative redundancy of compressing the TAR block, using the REF block as reference, is 1.2473. This number is 1.0187 when the REF block is compressed based on the model of the TAR block. Also, redundancies (complexities) of the REF and the TAR blocks are 1.9010 and 1.8580, respectively. The “F” in the last column shows that this similarity is not of the form inverted repeat. The second record of the body shows an inverted repeat rearrangement.

```

1 ##SMASH++
2 ##PARAM=<-r dataset/REF -t dataset/TAR -l 0 -m 1000>
3 ##INFO=<Ref=REF,RefSize=50000,Tar=TAR,TarSize=50000>
4 #RBeg  REnd  RRelRdn  RRdn  TBeg  TEnd  TRelRdn  TRdn  Inv
5 5000   10000  1.0187   1.9010  2000  7000  1.2473   1.8580  F
6 20000  30000   1.2367   1.9777  40000 30000  1.2545   1.9888  T

```

References