Transposition Event Detector "deTEct"

Transposition event detection tool *deTEct* using NGS alignment data and SV calling outputs (VCF files) from PBSV or Sniffles. *deTEct* is part of TransposonUltimate.

- **Input**: Structural variants (VCF file) of PBSV (on PBMM2 alignments) or Sniffles (on NGMLR alignments), transposon annotations (by resonaTE), reference genome (FASTA)
- **Output**: Annotation and classification of detected transposition events (GFF3).

Installation

Installation as CondaPackage:

```
conda install -c derkevinriehl transposition_detector_detect
```

Note: Otherwise you can find all source codes in this Github repository.

Usage

(using demo files of this repository, we use reference genome CB4856 and probe alignments SX3351)

```
transposition_deTEct -help

# demo for sniffles_ngmlr alignments
transposition_deTEct -seqHeadTXT demoFiles/sequence_heads.txt -transpGFF3
demoFiles/FinalAnnotations_Transposons.gff3 -assmFasta
demoFiles/sequence_CB4856.fasta -svTool sniffles -svFile
demoFiles/SX3351_addisababa.sniffles_ngmlr.vcf -outParsedFile
demoFiles/sniffles_ngmlr/SX3351_addisababa.SV.vcf.gff3 -outResultFile
demoFiles/sniffles_ngmlr/SX3351_addisababa.transpositionEvents.gff3

# demo for pbsv_pbsmm2 alignments
transposition_deTEct -seqHeadTXT demoFiles/sequence_heads.txt -transpGFF3
demoFiles/FinalAnnotations_Transposons.gff3 -assmFasta
demoFiles/sequence_CB4856.fasta -svTool pbsv -svFile
demoFiles/SX3351_addisababa.pbsv_pbmm2.vcf -outParsedFile
demoFiles/pbsv_pbmm2/SX3351_addisababa.SV.vcf.gff3 -outResultFile
demoFiles/pbsv_pbmm2/SX3351_addisababa.transpositionEvents.gff3
```

Parameter	Mandatory	Description
seqHeadTXT	(mandatory)	Sequence head names, TXT file (produced by reasonaTE)
transpGFF3	(mandatory)	Transposon annotation file, GFF3 file (produced by reasonaTE)
assmFasta	(mandatory)	Assembly file of reference genome, FASTA file
svTool	(mandatory)	Structural variant detection tool: "pbsv" or "sniffles"
svFile	(mandatory)	Structural variant detection output file, VCF file
outParsedFile	(mandatory)	Target file for VCF parsed outputs

Parameter	Mandatory	Description
outResultFile	(mandatory)	Target file for final results with transposition events

Explanation of output files

SX3351_addisababa.SV.vcf.gff3.matches.gff3

For each filtered structural variant a set of potential transposon annotation candidates (IDs similar to transposon annotation file) is reported:

```
seq1    PBSV    duplication    2909118 2910241 .      +      .      ['23769',
'23770', '23771'];Sseq1TYPE=DUP;END=2910240;Sseq1LEN=1122
seq1    PBSV    deletion       163800 164962 .        +      .      ['1',
'11827 '];Sseq1TYPE=DEL;END=164961;Sseq1LEN=-1161
seq1    PBSV    deletion       290360 290514 .        +      .      ['11843
'];Sseq1TYPE=DEL;END=290513;Sseq1LEN=-153
seq1    PBSV    insertion      343890 344420 .        +      .      ['538
'];merged;Sseq1TYPE=seq4NS;END=344424;Sseq1LEN=533
...
```

SX3351_addisababa.transpositionEvents.gff3

For each final structural variant that is considered to be a transposition event, the given transposon annotation (IDs similar to transposon annotation file) and predicted class are reported:

```
seq1    PBSV    deletion       290360 290514 .        +      .
        Transposon=11843;Class=2/1/2(hAT,TIR,DNATransposon);Sseq1TYPE=DEL;END=
290513;Sseq1LEN=-153
seq1    PBSV    insertion      610241 614786 .        +      .
        Transposon=545;Class=2/1/3(CMC,TIR,DNATransposon);merged;merged;Sseq1T
YPE=seq4NS;END=611763;Sseq1LEN=1521
seq1    PBSV    deletion       879772 884345 .        +      .
        Transposon=556;Class=1/1/2(Gypsy,LTR,Retrotransposon);Sseq1TYPE=DEL;EN
D=884344;Sseq1LEN=-4572
seq1    PBSV    insertion      1126531 1126860 .      +      .
        Transposon=23592;Class=2/1/1(Tc1-
Mariner,TIR,DNATransposon);Sseq1TYPE=seq4NS;END=1126859;Sseq1LEN=327
...
```