Against to dm6 by pbmm2

2024年5月13日星期一　　下午2:36

2024.5.13

multiple reads supported snv: (including supplementary mapping):

因为嵌合reads在mapping过程中会采用supplementary mapping

1. mapping to dm6: (fwd + rev deepconsensus bam):
    a. mapping to dm6 by pbmm2
    b. filter the bam: fwd rev reads overlap > 90%
    c. sort & index the overlap90 bam

```
# pbmm2 align --preset CCS --log-level INFO --sort -j 40 /rd/caiya/dm6.fa.mmi $AA_deepconsensus_bam AA.deepconsensus.cx3.pbmm2.dm6.bam
# samtools index -@40 AA.deepconsensus.cx3.pbmm2.dm6.bam

python3 pbmm2filter.chr.py AA.deepconsensus.cx3.pbmm2.dm6.bam # output file: AA.deepconsensus.cx3.pbmm2.dm6.overlap90.bam

samtools sort -@40 AA.deepconsensus.cx3.pbmm2.dm6.overlap90.bam -o AA.deepconsensus.cx3.pbmm2.dm6.overlap90.s.bam
samtools index -@40 AA.deepconsensus.cx3.pbmm2.dm6.overlap90.s.bam
```

2. get mismatch from the filtered bam:
    a. get mismatch from the bam
    b. combine the snv by position and strand
    c. get dsDNA and ssDNA
    d. filter by read quality
    e. get position in the read for the further subreads evaluation

```
inbam=$i.deepconsensus.cx3.pbmm2.dm6.chr.overlap90.s.bam
snv0=$inbam.snv
fwd_snv0=$inbam.fwd.snv
rev_snv0=$inbam.rev.snv

python3 get_mismatch_bam.sup.py /rd/caiya/dm6.major.chr.fasta $inbam
python3 combine_snv_support_reads.py $snv0
python3 filter0.ss_ds.multi_reads.py $snv0"0"

#### qv filter
for j in ds ss
do
	snv=$inbam.$j.out
	for k in 20 30
	do
		python3 filter1.qv.py $snv $j $k

		out_snv=$inbam.$j.q$k.out
		pos_out=$inbam.$j.q$k.readpos
		cat $out_snv | awk '{if($5!="NA"){split($5, a, ","); for(i in a){split(a[i], b, ":"); print b[1]"\t"b[2]}}} {if($6!="NA"){split($6, a, ","); for(i in a){split(a[i], b, ":"); print b[1]"\t"b[2]}}}' > $pos_out

	done
done
```

3. subreads mapping to deepconsensus reads:
    a. gain 4 bam:
        i. fwd_fwd: mapping to + strand of fwd deepconsensus strand
        ii. fwd_rev: mapping to - strand of fwd deepconsensus strand
        iii. rev_rev: mapping to + strand of rev deepconsensus strand
        iv. rev_fwd: mapping to - strand of rev deepconsensus strand

```
AA_subreads_bam=/rd/jiahx/ltrdata/ANNO_XS01KF2020060224_PM-XS01KF2020060224-03_2021-12-14/AA/r64054_20211209_113508/P05DY21533473-1_r64054_20211209_113508_1_A02.subreads.bam
AA_deepconsensus_bam=/rd/caiya/LTR/duplex/AA/mismatch/all_reads/AA.deepconsensus.cx3.bam

samtools view -@40 -h $AA_deepconsensus_bam | awk '{if($0~/^@/){print $0} else{split($1, a, "/"); if(a[4]=="fwd"){print $0}}}' | samtools view -@40 -b > AA.deepconsensus.fwd.bam
pbindex -j 40 AA.deepconsensus.fwd.bam
samtools view -@40 -h $AA_deepconsensus_bam | awk '{if($0~/^@/){print $0} else{split($1, a, "/"); if(a[4]=="rev"){print $0}}}' | samtools view -@40 -b > AA.deepconsensus.rev.bam
pbindex -j 40 AA.deepconsensus.rev.bam


### fwd
actc -j 80 \
	$AA_subreads_bam \
	AA.deepconsensus.fwd.bam \
	AA.subreads_to_deepconsensus.cx3.fwd.bam

samtools view -@40 -b -F 16 AA.subreads_to_deepconsensus.cx3.fwd.bam | samtools sort -@40 -o AA.subreads_to_deepconsensus.cx3.fwd_fwd.s.bam
samtools index -@40 AA.subreads_to_deepconsensus.cx3.fwd_fwd.s.bam

samtools view -@40 -b -f 16 AA.subreads_to_deepconsensus.cx3.fwd.bam | samtools sort -@40 -o AA.subreads_to_deepconsensus.cx3.fwd_rev.s.bam
samtools index -@40 AA.subreads_to_deepconsensus.cx3.fwd_rev.s.bam

rm AA.subreads_to_deepconsensus.cx3.fwd.bam AA.subreads_to_deepconsensus.cx3.fwd.fasta


#### rev
actc -j 80 \
	$AA_subreads_bam \
	AA.deepconsensus.rev.bam \
	AA.subreads_to_deepconsensus.cx3.rev.bam

samtools view -@40 -b -F 16 AA.subreads_to_deepconsensus.cx3.rev.bam | samtools sort -@40 -o AA.subreads_to_deepconsensus.cx3.rev_rev.s.bam
samtools index -@40 AA.subreads_to_deepconsensus.cx3.rev_rev.s.bam

samtools view -@40 -b -f 16 AA.subreads_to_deepconsensus.cx3.rev.bam | samtools sort -@40 -o AA.subreads_to_deepconsensus.cx3.rev_fwd.s.bam
```

```
samtools index -@40 AA.subreads_to_deepconsensus.cx3.rev_fwd.s.bam

rm AA.subreads_to_deepconsensus.cx3.rev.bam AA.subreads_to_deepconsensus.cx3.rev.fasta
```

4. subreads correction:
    a. generated pileup file for ds/ss DNA mutation readpos file
    b. combine them by filter2.pileup.py
    c. remove the snv within the reads boundary 15/20/25 bp region.

```
for i in AA AAC C01 E01
do
    bampath=/rd1/caiya/LTR/duplex/sup_test

    deepconsensus_fwd_fq=$i.deepconsensus.cx3.fwd.fastq1
    deepconsensus_rev_fq=$i.deepconsensus.cx3.rev.fastq1
    fwd_fwd_bam=$bampath/$i.subreads_to_deepconsensus.cx3.snv.fwd_fwd.s.bam
    rev_rev_bam=$bampath/$i.subreads_to_deepconsensus.cx3.snv.rev_rev.s.bam
    fwd_rev_bam=$bampath/$i.subreads_to_deepconsensus.cx3.snv.fwd_rev.s.bam
    rev_fwd_bam=$bampath/$i.subreads_to_deepconsensus.cx3.snv.rev_fwd.s.bam

    for j in ds ss
    do
        for k in q20 q30
        do
            snv=$i.deepconsensus.cx3.pbmm2.dm6.chr.overlap90.s.bam.$j.$k.out
            snv_pileup=$i.deepconsensus.cx3.pbmm2.dm6.chr.overlap90.s.bam.$j.$k.plus_minus.out
            readpos=$i.deepconsensus.cx3.pbmm2.dm6.chr.overlap90.s.bam.$j.$k.readpos
            fwd_readpos=`basename $readpos ".readpos"`".fwd.readpos"
            rev_readpos=`basename $readpos ".readpos"`".rev.readpos"
            pileup_file=$i.deepconsensus.cx3.pbmm2.dm6.chr.overlap90.s.bam.$j.$k.plus_minus.out
            cat $readpos | awk '{split($1, a, "/"); strand=a[4]; if(strand=="fwd"){print $0}}' > $fwd_readpos
            cat $readpos | awk '{split($1, a, "/"); strand=a[4]; if(strand=="rev"){print $0}}' > $rev_readpos

    if [ $j == "ds" ]; then
        ds_fwd_pileup=`basename $fwd_readpos ".readpos"`".pileup"
        ds_rev_pileup=`basename $rev_readpos ".readpos"`".pileup"
        samtools mpileup --no-output-ins --no-output-ins --no-output-del --no-output-del --no-output-ends --min-BQ 0 -f $deepconsensus_fwd_fq -l $fwd_readpos $fwd_fwd_bam > $ds_fwd_pileup
        samtools mpileup --no-output-ins --no-output-ins --no-output-del --no-output-del --no-output-ends --min-BQ 0 -f $deepconsensus_rev_fq -l $rev_readpos $rev_rev_bam > $ds_rev_pileup
        python3 combine.pileup.ds.py $snv $ds_fwd_pileup $ds_rev_pileup

    else
        ss_fwd_fwd_pileup=`basename $fwd_readpos ".readpos"`"_fwd.pileup"
        ss_fwd_rev_pileup=`basename $fwd_readpos ".readpos"`"_rev.pileup"
        ss_rev_rev_pileup=`basename $rev_readpos ".readpos"`"_rev.pileup"
        ss_rev_fwd_pileup=`basename $rev_readpos ".readpos"`"_fwd.pileup"

    samtools mpileup --no-output-ins --no-output-ins --no-output-del --no-output-del --no-output-ends --min-BQ 0 -f $deepconsensus_fwd_fq -l $fwd_readpos $fwd_fwd_bam > $ss_fwd_fwd_pileup
    samtools mpileup --no-output-ins --no-output-ins --no-output-del --no-output-del --no-output-ends --min-BQ 0 -f $deepconsensus_fwd_fq -l $fwd_readpos $fwd_rev_bam > $ss_fwd_rev_pileup
    samtools mpileup --no-output-ins --no-output-ins --no-output-del --no-output-del --no-output-ends --min-BQ 0 -f $deepconsensus_rev_fq -l $rev_readpos $rev_rev_bam > $ss_rev_rev_pileup
    samtools mpileup --no-output-ins --no-output-ins --no-output-del --no-output-del --no-output-ends --min-BQ 0 -f $deepconsensus_rev_fq -l $rev_readpos $rev_fwd_bam > $ss_rev_fwd_pileup
    python3 combine.pileup.ss.py $snv $ss_fwd_fwd_pileup $ss_fwd_rev_pileup $ss_rev_rev_pileup $ss_rev_fwd_pileup

    fi

    python3 filter2.pileup.py $pileup_file $j 5 0.8


        done
    done
done


python3 rm_boundary_snv.multireads.py AA.deepconsensus.cx3.fwd.fastq1 AA.deepconsensus.cx3.rev.fastq1 25 AA.snv.list
python3 rm_boundary_snv.multireads.py AAC.deepconsensus.cx3.fwd.fastq1 AAC.deepconsensus.cx3.rev.fastq1 25 AAC.snv.list
python3 rm_boundary_snv.multireads.py C01.deepconsensus.cx3.fwd.fastq1 C01.deepconsensus.cx3.rev.fastq1 25 C01.snv.list
python3 rm_boundary_snv.multireads.py E01.deepconsensus.cx3.fwd.fastq1 E01.deepconsensus.cx3.rev.fastq1 25 E01.snv.list
```

Against to hap1/hap2 or pri+alt assembly

2024年5月13日星期一　下午3:23

2024.5.13

Pri+alt and hap1/hap2 assembly mapping:

之所以选择使用AAC组装作为参考基因组，为了消除germline和TE 之间gene conversion的影响，
得到的snv理论上包括de novo mutation，AA中特异的somatic mutation，以及 AA中LTR复制扩增带来的mutation

- pri + alt assembly is similar as against to dm6.fa
- hap1/hap2 mapping, 只保留在hap1和hap2中同一read同一位置出现snv的mutation：
  比如m64054_211209_114633/26411840/deepconsensus/fwd 55位置在
  hap1中：h1tg000001l　16644　G　C
  hap2中：h2tg000018l　17444　G　C

1. filter the fastq by read qv: output file: AA.deepconsensus.cx3.q20.fastq

```
AA_fwd_fq=/rd/caiya/LTR/duplex/AA.deepconsensus.cx3.fwd.fastq
AA_rev_fq=/rd/caiya/LTR/duplex/AA.deepconsensus.cx3.rev.fastq

python3 filter_rq_fwd_rev_fastq.py 20 $AA_fwd_fq $AA_rev_fq
python3 filter_rq_fwd_rev_fastq.py 30 $AA_fwd_fq $AA_rev_fq
```

1. mapping to hap1 and hap2 assembly by winnowmap2:
   a. winnowmap2 mapping q20/q30 fastq file to reference genome
   b. filter 90% overlap and sort and index bam
   c. generate snv file (primary mismatch file)
   primary mismatch file:
   h1tg000001l　14596　C　T　m64054_211209_114633/95814109/deepconsensus/rev:1547:-
   h1tg000001l　14597　T　C　m64054_211209_114633/95814109/deepconsensus/rev:1546:-
   h1tg000001l　16644　G　C　m64054_211209_114633/26411840/deepconsensus/fwd:55:+

```
for i in p_a_ctg hap1 hap2
do
	for j in 20 30
	do
		fq=/rd/caiya/LTR/duplex/AAC_ref/AA.deepconsensus.cx3.$j.fastq.gz
		ref=AAC.$i.fa

		winnowmap -W AAC.$i.repetitive_k15.txt -ax map-pb --MD --eqx $ref $fq | samtools view -@40 -F0x104 -bS > AA.deepconsensus.cx3.q$j.AAC.$i.F104.bam
		samtools sort -@40 AA.deepconsensus.cx3.q$j.AAC.$i.F104.bam -o AA.deepconsensus.cx3.q$j.AAC.$i.F104.s.bam
		samtools index -@40 AA.deepconsensus.cx3.q$j.AAC.$i.F104.s.bam
		rm AA.deepconsensus.cx3.q$j.AAC.$i.F104.bam

		python3 /rd/caiya/LTR/duplex/AAC_ref/pbmm2filter.p_ctg.py AA.deepconsensus.cx3.q$j.AAC.$i.F104.s.bam
		samtools sort -@ 40 AA.deepconsensus.cx3.q$j.AAC.$i.F104.s.overlap90.bam -o AA.deepconsensus.cx3.q$j.AAC.$i.F104.overlap90.s.bam
		samtools index -@ 40 AA.deepconsensus.cx3.q$j.AAC.$i.F104.overlap90.s.bam
		rm AA.deepconsensus.cx3.q$j.AAC.$i.F104.s.overlap90.bam

		inbam=AA.deepconsensus.cx3.q$j.AAC.$i.F104.overlap90.s.bam
		snv0=$inbam.snv
		fwd_snv0=$inbam.fwd.snv
		rev_snv0=$inbam.rev.snv

		python3 /rd/caiya/LTR/duplex/AAC_ref/get_mismatch_bam.winnowmap2.py $ref $inbam

	done
done
```

2. combine hap1 hap2 primary mismatch file by readpos located in the reads:
   a. combine snv by filter_hap12_shared_snv.multireads0.py
   b. get the shared snv of hap1 and hap2, filter the mutation by ds/ss by combine_snv_support_reads.winnowmap2.py and filter0.ss_ds.multi_reads.py

```
for j in q20 q30
do
	python3 /rd/caiya/LTR/duplex/AAC_ref/hap1_2/filter_hap12_shared_snv.multireads0.py AA.deepconsensus.cx3.$j.AAC.hap1.F104.over lap90.s.bam.snv AA.deepconsensus.cx3.$j.AAC.hap2.F104.overlap90.s.bam.snv AA.deepconsensus.cx3.$j.AAC.F104.overlap90

	for i in hap1 hap2
	do
		snv0=AA.deepconsensus.cx3.$j.AAC.F104.overlap90.shared.$i.snv

		python3 /rd/caiya/LTR/duplex/AAC_ref/combine_snv_support_reads.winnowmap2.py $snv0
		python3 /rd/caiya/LTR/duplex/AAC_ref/filter0.ss_ds.multi_reads.py $snv0"0"

	done
done
```

3. subreads pileup correction:
   a. get all the mutation position together
   b. samtools pileup
   c. integrate the subreads to deepconsnesus pileup information into snv file
   d. remove the boundary 15/20/25 bp snv for each reads

```
bampath=/nfs119/rd1/yacai/LTR/duplex/AA
deepconsensus_fwd_fq=/rd/caiya/LTR/duplex/AA/mismatch/pileup/AA.deepconsensus.cx3.fwd.fastq1
deepconsensus_rev_fq=/rd/caiya/LTR/duplex/AA/mismatch/pileup/AA.deepconsensus.cx3.rev.fastq1
fwd_fwd_bam=$bampath/AA.subreads_to_deepconsensus.cx3.fwd_fwd.s.bam
rev_rev_bam=$bampath/AA.subreads_to_deepconsensus.cx3.rev_rev.s.bam
fwd_rev_bam=$bampath/AA.subreads_to_deepconsensus.cx3.fwd_rev.s.bam
rev_fwd_bam=$bampath/AA.subreads_to_deepconsensus.cx3.rev_fwd.s.bam

fwd_readpos=AA.deepconsensus.cx3.AAC.F104.pri.overlap90.shared.fwd.readpos
rev_readpos=AA.deepconsensus.cx3.AAC.F104.pri.overlap90.shared.rev.readpos
fwd_fwd_pileup=AA.deepconsensus.cx3.AAC.F104.pri.overlap90.shared.fwd_fwd.pileup
fwd_rev_pileup=AA.deepconsensus.cx3.AAC.F104.pri.overlap90.shared.fwd_rev.pileup
rev_rev_pileup=AA.deepconsensus.cx3.AAC.F104.pri.overlap90.shared.rev_rev.pileup
rev_fwd_pileup=AA.deepconsensus.cx3.AAC.F104.pri.overlap90.shared.rev_fwd.pileup

for i in hap1 hap2
do
	for j in ds ss
	do
		for k in q20 q30
		do
			snv=AA.deepconsensus.cx3.$k.AAC.F104.overlap90.shared.$i.$j.out
			prefix=`basename $snv ".out"`
			readpos=$prefix.readpos
			fwd_readpos0=$prefix.fwd.readpos
			rev_readpos0=$prefix.rev.readpos
			pileup_file=$prefix.plus_minus.out

			cat $snv | awk '{if($5!="NA"){split($5, a, ","); for(i in a){split(a[i], b, ":"); print b[1]"\t"b[2]}}} {if($6!="NA"){split($6, a, ","); for(i in a){split(a[i], b, ":"); print b[1]"\t"b[2]}}}' > $readpos
			cat $readpos | awk '{split($1, a, "/"); strand=a[4]; if(strand=="fwd"){print $0}}' > $fwd_readpos0
			cat $readpos | awk '{split($1, a, "/"); strand=a[4]; if(strand=="rev"){print $0}}' > $rev_readpos0

	done
	done
done

cat *.fwd.readpos | awk '{all[$0]++}END{for(i in all){print i}}' > $fwd_readpos
cat *.rev.readpos | awk '{all[$0]++}END{for(i in all){print i}}' > $rev_readpos

samtools mpileup --no-output-ins --no-output-ins --no-output-del --no-output-del --no-output-ends --min-BQ 0 -f $deepconsensus_fwd_fq -l $fwd_readpos $fwd_fwd_bam > $fwd_fwd_pileup
samtools mpileup --no-output-ins --no-output-ins --no-output-del --no-output-del --no-output-ends --min-BQ 0 -f $deepconsensus_fwd_fq -l $fwd_readpos $fwd_rev_bam > $fwd_rev_pileup
samtools mpileup --no-output-ins --no-output-ins --no-output-del --no-output-del --no-output-ends --min-BQ 0 -f $deepconsensus_rev_fq -l $rev_readpos $rev_rev_bam > $rev_rev_pileup
```

```
samtools mpileup --no-output-ins --no-output-ins --no-output-del --no-output-del --no-output-ends --min-BQ 0 -f $deepconsensus_rev_fq -l $rev_readpos $rev_fwd_bam > $rev_fwd_pileup


for i in hap1 hap2 p_a_ctg
do
	for j in ds ss
	do
		for k in q20 q30
		do
			if [ $i == "p_a_ctg" ]; then
				snv=/rd/caiya/LTR/duplex/AAC_ref/supplementary/AA.deepconsensus.cx3.$k.AAC.$i.F104.overlap90.s.bam.$j.out
			else
				snv=/rd/caiya/LTR/duplex/AAC_ref/supplementary/AA.deepconsensus.cx3.$k.AAC.F104.overlap90.shared.$i.$j.out
			fi

			prefix=`basename $snv ".out"`
			readpos=$prefix.readpos
			fwd_readpos=$prefix.fwd.readpos
			rev_readpos=$prefix.rev.readpos
			pileup_file=$prefix.plus_minus.out

			if [ $j == "ds" ]; then
				python3 /rd/caiya/LTR/duplex/AAC_ref/combine.pileup.ds.win.py $snv $fwd_fwd_pileup $rev_rev_pileup
			else
				python3 /rd/caiya/LTR/duplex/AAC_ref/combine.pileup.ss.win.py $snv $fwd_fwd_pileup $fwd_rev_pileup $rev_rev_pileup $rev_fwd_pileup
			fi

			python3 /rd/caiya/LTR/duplex/AAC_ref/filter2.pileup.py $pileup_file $j 5 0.8
			vaf_file=$prefix.plus_minus.subpass_5.vaf_0.8.out
			echo $vaf_file >> AA.deepconsensus.AAC_p_a_ctg_hap1_2.ref.sup.snv.list

		done
	done
done

python3 /rd/caiya/LTR/duplex/AAC_ref/rm_boundary_snv.multireads.py $deepconsensus_fwd_fq $deepconsensus_rev_fq 20 AA.deepconsensus.AAC_p_a_ctg_hap1_2.ref.sup.snv.list
python3 /rd/caiya/LTR/duplex/AAC_ref/rm_boundary_snv.multireads.py $deepconsensus_fwd_fq $deepconsensus_rev_fq 25 AA.deepconsensus.AAC_p_a_ctg_hap1_2.ref.sup.snv.list
```

4. located the snv position in the dm6 reference genome:
   a. 取snv上下游25bp序列（hap1/hap2 assembly seq）
   b. bwa mapping to reference genome
   c. 只保留primary mapping（一般primary mapping都是长的mapping，能够包含住位于short reads中间的snv）
   d. primary mapping position + 25 为snv的位置（粗略定位）
   e. 根据参考基因组注释信息对snv进行定位得到pos1 file （get_snv_type.py）
   f. 合并统计snv的种类，放到.mutationtype文件

```
for r in hap1 hap2 p_a_ctg
do
	for i in 200
	do
		>AA.F104.LR$i.$r.mutationtype
		for j in ds ss
		do
			for k in q20 q30
			do

				if [ $r == "p_a_ctg" ]; then
					prefix=AA.deepconsensus.cx3.$k.AAC.$r.F104.overlap90.s.bam.$j.plus_minus.subpass_5.vaf_0.8.rm_20bp
					ref=/rd/caiya/LTR/AA_ecc/AAC_ref/AAC.deepconsensus.pri.hifiasm0198.asm.p_a_ctg.fa
				else
					prefix=AA.deepconsensus.cx3.$k.AAC.F104.overlap90.shared.$r.$j.plus_minus.subpass_5.vaf_0.8.rm_20bp
					ref=/rd/caiya/LTR/AAC_AUC/deepconsensus/primary_asm/AAC.deepconsensus.primary.asm.bp.$r.p_ctg.fa
				fi

				snv=$prefix.out
				bed=$prefix.LR$i.bed
				fasta=$prefix.LR$i.fasta
				sam=$prefix.LR$i.bwa.dm6major.sam
				pos=$prefix.LR$i.bwa.dm6major.pos

				cat $snv | awk 'BEGIN{OFS="\t"; while((getline < "/rd/caiya/LTR/duplex/AAC_ref/supplementary/mpileup_test/AAC.deepconsensus.primary.asm.bp.hap12_p_a_ctg.p_ctg.length") > 0) len[$1]=$2}{start=$2-"$i"; if (start < 1) start=1; end=$2+"$i"; if (end > len[$1]) end=len[$1]; print $1, start, end, "SNV_"NR}' > $bed
				bedtools getfasta -nameOnly -fi $ref -bed $bed -fo $fasta

				bwa mem -t 8 /rd/caiya/dm6.major.chr.fasta $fasta > $sam

				samtools view -F0x904 $sam | awk '{print $3"\t"$4+"$i"-1"\t"$1}' > $pos
				python3 /rd/caiya/LTR/duplex/AAC_ref/bwa_dm6/get_snv_type.py $pos

				cat $pos"1" | awk '{all[$4]++}END{for(i in all){print "'$j'"\t"'$k'"\t"i"\t"all[i]}}' >> AA.F104.LR$i.$r.mutationtype

			done
		done
	done
done
```

5. remapping fastq reads to dm6 reference:
   a. 按照相同的方法重新将reads mapping到dm6上

```
for i in AA AAC
do
	for j in q20 q30
	do
		fq=$i.deepconsensus.cx3.$j.fastq.gz
		prefix=$i.deepconsensus.cx3.$j

		winnowmap -W /rd/caiya/LTR/E_AU_AUC/OAU/winnowmap2/dm6.repetitive_k15.txt -ax map-pb --MD --eqx /rd/caiya/dm6.fa $fq | samtools view -@40 -F0x104 -bS > $prefix.F104.bam
		samtools sort -@40 $prefix.F104.bam -o $prefix.F104.s.bam
		samtools index -@40 $prefix.F104.s.bam
		rm $prefix.F104.bam

		python3 /rd/caiya/LTR/duplex/AAC_ref/pbmm2filter.p_ctg.py $prefix.F104.s.bam
		samtools sort -@ 40 $prefix.F104.s.overlap90.bam -o $prefix.F104.overlap90.s.bam
		samtools index -@ 40 $prefix.F104.overlap90.s.bam
		rm $prefix.F104.s.overlap90.bam

	done
done
```

6. calculated each part depth:

```
>AA.deepconsensus.cx3.dm6.F104.overlap90.q20_q30.depth
>AAC.deepconsensus.cx3.dm6.F104.overlap90.q20_q30.depth
for i in AA AAC
do
	for k in q20 q30
	do
		bam=$i.deepconsensus.cx3.$k.F104.overlap90.s.bam
		for j in LTR LINE DNA RC RNA VNTR Satellite singleton multicopy UTR small_8_30_intron other_intron intergenic Other Unknown
		do
			bed=/rd/caiya/LTR/duplex/AAC/mismatch_py/mapq60/pileup/subreads_vaf_80/get_input_seq/dm6.repeat_nonrepeat.$j.bed
			samtools depth -@40 $bam -b $bed | awk 'BEGIN{alen=0}{alen=alen+$3}END{print "'$i'"\t"'$k'"\t"'$j'"\t"alen}' >> $i.deepconsensus.cx3.dm6.F104.overlap90.q20_q30.depth
		done
	done
done
```

7. get mutation rate by mutation type and sequencing depth:

```
python3 /rd/caiya/LTR/duplex/AAC_ref/bwa_dm6/get_mutation_rate.mutationtype_depth.py AA.F104.LR200.hap1.mutationtype /rd/caiya/LTR/duplex/AAC_ref/dm6mapping/sup/AA.deepconsensus.cx3.dm6.F104.overlap90.q20_q30.depth
```

```
python3 /rd/caiya/LTR/duplex/AAC_ref/bwa_dm6/get_mutation_rate.mutationtype_depth.py AA.F104.LR200.hap2.mutationtype /rd/caiya/LTR/duplex/AAC_ref/dm6mapping/sup/AA.deepconsensus.cx3.dm6.F104.overlap90.q20_q30.depth
python3 /rd/caiya/LTR/duplex/AAC_ref/bwa_dm6/get_mutation_rate.mutationtype_depth.py AA.F104.LR200.p_a_ctg.mutationtype /rd/caiya/LTR/duplex/AAC_ref/dm6mapping/sup/AA.deepconsensus.cx3.dm6.F104.overlap90.q20_q30.depth
```

Against to hap1/hap2 or pri+alt assembly (only remain primary mapping, remove all the reads with supplementary)

2024年5月13日星期一　　下午4:31

2024.5.13

去掉AA中所有带有supplementary mapping的reads，剩下的只有primary mapping 的reads：

理论上应该只包括基因组DNA + linear LTR

只有第一步和普通的hap1/hap2 mapping有区别：

```
for i in hap1 hap2
do
	for j in q20 q30
	do
		fq=/rd/caiya/LTR/duplex/AAC_ref/AA.deepconsensus.cx3.$j.fastq.gz
		ref=/rd/caiya/LTR/duplex/AAC_ref/supplementary/AAC.$i.fa

		winnowmap -W AAC.$i.repetitive_k15.txt -ax map-pb --MD --eqx $ref $fq | samtools view -@40 -F0x104 -bS > AA.deepconsensus.cx3.$j.AAC.$i.F104.bam
		samtools sort -@40 AA.deepconsensus.cx3.$j.AAC.$i.F104.bam -o AA.deepconsensus.cx3.$j.AAC.$i.F104.s.bam
		samtools index -@40 AA.deepconsensus.cx3.$j.AAC.$i.F104.s.bam
		rm AA.deepconsensus.cx3.$j.AAC.$i.F104.bam

		samtools view -@40 -f 2048 AA.deepconsensus.cx3.$j.AAC.$i.F104.s.bam | cut -f1 > AA.deepconsensus.cx3.$j.AAC.$i.F104.sup.txt
		samtools view -@40 -h AA.deepconsensus.cx3.$j.AAC.$i.F104.s.bam | grep -v -F -f AA.deepconsensus.cx3.$j.AAC.$i.F104.sup.txt | samtools view -b -@40 > AA.deepconsensus.cx3.$j.AAC.$i.F104.pri.s.bam
		samtools index -@40 AA.deepconsensus.cx3.$j.AAC.$i.F104.pri.s.bam
		rm AA.deepconsensus.cx3.$j.AAC.$i.F104.s.bam AA.deepconsensus.cx3.$j.AAC.$i.F104.s.bam.bai  AA.deepconsensus.cx3.$j.AAC.$i.F1 04.sup.txt

		python3 /rd/caiya/LTR/duplex/AAC_ref/pbmm2filter.p_ctg.py AA.deepconsensus.cx3.$j.AAC.$i.F104.pri.s.bam
		samtools sort -@ 40 AA.deepconsensus.cx3.$j.AAC.$i.F104.pri.s.overlap90.bam -o AA.deepconsensus.cx3.$j.AAC.$i.F104.pri.overlap90.s.bam
		samtools index -@ 40 AA.deepconsensus.cx3.$j.AAC.$i.F104.pri.overlap90.s.bam
		rm AA.deepconsensus.cx3.$j.AAC.$i.F104.pri.s.overlap90.bam

		inbam=AA.deepconsensus.cx3.$j.AAC.$i.F104.pri.overlap90.s.bam
		snv0=$inbam.snv
		fwd_snv0=$inbam.fwd.snv
		rev_snv0=$inbam.rev.snv

		python3 /rd/caiya/LTR/duplex/AAC_ref/get_mismatch_bam.winnowmap2.py $ref $inbam

	done
done
```