水稻流程

2016年7月13日

11:05

1 下载数据

<ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByStudy/sra/SRP/SRP017/SRP017256>

2 转化数据

使用sra-toolkit软件 将sra转化成fastq

3 查看数据质量

nohup fastqc -o ~/fastqc\_test/no.1 -t 6 ~/original\_data/fastq/SRR618545.fastq &

4 控制数据质量

5 下载参考序列

<ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/all.con>

6 建立索引

nohup build\_all ~/genome/all.con &

7 序列比对

nohup batmeth2-align -g ~/genome/rice/all.con -i ~/trimmomatic/rice/no.3/SRR618547\_tri.fastq -o ~/batmeth/map-1/rice/outPrefix-47 -p 1 -n 2 &

echo ''|qsub -l nodes=comput24:ppn=1,walltime=1200:00:00 -d ./ -e ./err -o ./0-err -N test（提交到计算节点）

8 去甲基化

~/software/batmeth2/bin/split -o ~/batmeth/get\_meth-2/rice/47/Final\_Result-47.sam -g ~/genome/rice/all.con -n 2 -i ~/batmeth/map-1/rice/outPrefix-47 -m ~/batmeth/get\_meth-2/47/Final.methratio-47.txt -p 6

9 使用近似的贝叶斯模型检测BS-seq的单核苷酸多态性

a) Convert Sam files into BAM files-转化格式sam转到bam

samtools view -bS ~/batmeth/get\_meth-2/rice/47/Final\_Result-47.sam > ~/batmeth/bam-3.1/Final\_Result-47.bam

b) Sort the bam file -用samtools sort将bam文件进行排序

samtools sort ~/batmeth/bam-3.1/Final\_Result-47.bam -o ~/batmeth/sort-3.2/Final\_Result-47.sort.bam

rm ~/batmeth/bam-3.1/Final\_Result-47.bam

c) BS-Seq variation detection -BS-Seq变异检测

BS-Snper --fa ~/genome/rice/all.con --input ~/batmeth/sort-3.2/Final\_Result-47.sort.bam --output ~/batmeth/snp-3.3/rice/snp-47\_result\_file

10 下载基因组注释文件

<ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/all.gff3>

11 注释文件中的每种功能元件要提取出来，方便后面做图

awk '$3=="gene"' all.gff3 > gene.gff3

awk '$3=="three\_prime\_UTR" || $3=="five\_prime\_UTR"' all.gff3 > UTR.gff3

12 基因或DNA甲基化水平和密度试验

methyGff -o ~/batmeth/methyGff-4/rice/all.gene/OUT\_PREFIX-47 -G ~/genome/rice/all.con -g ~/genome/rice/all.gff3 -m ~/batmeth/get\_meth-2/rice/47/Final.methratio-47.txt --TSS --TTS -B -P

methyGff -o ~/batmeth/methyGff-4/rice/cds/OUT\_PREFIX-47 -G ~/genome/rice/all.con -g ~/genome/rice/cds.gff3 -m ~/batmeth/get\_meth-2/rice/47/Final.methratio-47.txt --TSS --TTS -B -P

12 安装R和ggplot2，pheatmap

>install.packages("pheatmap")

会跳出选项，选择镜像：选最后一个，再选中国厦门

>install.packages("ggplot2")

会跳出选项，选择镜像：选最后一个，再选中国厦门

13 画图（不太懂？？？）

a) 染色体DNA甲基化密度水平，跨基因DNA甲基化分布，在2kb up of gene，gene body，2kb gene body的DNA甲基化水平图

methyPlot ~/batmeth/get\_meth-2/rice/47/methBins.txt ~/batmeth/methyplot-5.1/rice/all\_gene/methy.distri-47.pdf 0.025 ~/batmeth/methyGff-4/rice/all.gene/47/OUT\_PREFIX-47.Methylevel.1.txt ~/batmeth/methyplot-5.1/rice/all\_gene/methlevel-47.pdf TSS TTS ~/batmeth/methyGff-4/rice/all.gene/47/OUT\_PREFIX-47.AverMethylevel.1.txt ~/batmeth/methyplot-5.1/rice/all\_gene/elements-47.pdf

b) 基因密度，内显子和全基因组DNA甲基化水平

Rscript ~/software/batmeth2/src/density\_plot\_with\_methyl\_oneSample.r ~/batmeth/get\_meth-2/rice/47/methBins.txt ~/batmeth/methyGff-4/rice/exon/OUT\_PREFIX-47.annoDensity.1.txt ~/batmeth/methyGff-4/rice/gene/OUT\_PREFIX-47.annoDensity.1.txt ~/batmeth/rscript-5.2/exon\_gene.test.pdf Oryza\_sativa(一个横纵坐标的标记，一般就是写自己做的物种的编号就行)

c) 全基因组的DNA甲基化热图

GeneMethHeatmap ~/batmeth/methyGff-4/rice/all.gene/47/OUT\_PREFIX-47 ~/batmeth/methyGff-4/rice/all.gene/49/OUT\_PREFIX-49 1.0 0.6 0.2

14 Get DMC or DMR/DMG\*\* 得到DMC，DMR，DMG(DM表示甲基化差异；C指胞嘧啶；R是Region，区域的意思；G指的是gene基因)

1)获得甲基化差异区域

不懂结果的含义

batDMR -g ~/genome/rice/all.con -L -o\_dm rice.gene.CHH.dmr.txt -1 ~/batmeth/get\_meth-2/rice/47/Methyl\_Region.CHH.txt -2 ~/batmeth/get\_meth-2/rice/49/Methyl\_Region.CHH.txt

batDMR -g ~/genome/rice/all.con -L -o\_dm rice.gene.CHG.dmr.txt -1 ~/batmeth/get\_meth-2/rice/47/Methyl\_Region.CHG.txt -2 ~/batmeth/get\_meth-2/rice/49/Methyl\_Region.CHG.txt

batDMR -g ~/genome/rice/all.con -L -o\_dm rice.gene.CG.dmr.txt -1 ~/batmeth/get\_meth-2/rice/47/Methyl\_Region.CG.txt -2 ~/batmeth/get\_meth-2/rice/49/Methyl\_Region.CG.txt

2)

3) awk '$6<0.05 && sqrt($13\*$13)>0.6 ' rice.gene.CG.dmr.txtt > rice.gene.CG.dmr.filter.txt

15[7] DM annotation 画图-甲基化差异

DMCannotationPlot -o abc -G ~/genome/rice/all.con -g ~/genome/rice/all.gff3 -d ~/batmeth/DMR-6.1/47-49/rice.gene.CG.dmr.fliter.txt -c CG