1. Vcf文件处理：提取contig，Start，end列

Contig1 1 3487

less -S E-fontanierii-4.gvcf.gz | grep -v '^#' | awk '{print $1,$2,$8}' | awk '{split($3,x,"=");name=x[2];print $1,$2,name}' | awk '{split($3,x,";");name=x[1];print $1,$2,name}' > vcf.txt

Contig1 3488 3627

Contig1 3628 3689

Contig1 3690 3695

Contig1 3696 3706

Contig1 3707 2

Contig1 3708 3779

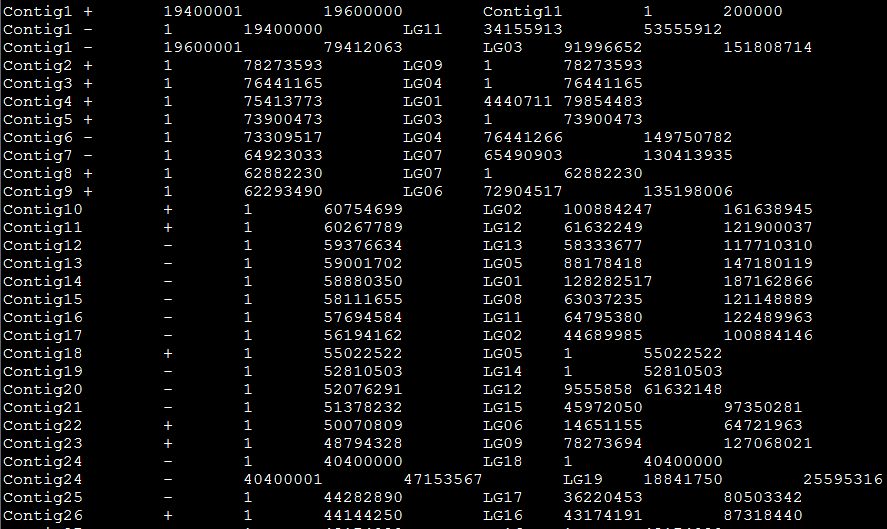
Contig1 3780 2

Contig1 3781 3828

Contig1 3829 3834

2. bed文件处理：

awk '{print $5,$6,$7,$8,$1,$2,$3}' cishu.LG.bed | tr ' ' '\t' | sort -Vk 1 > bed.txt



3. match:

setwd("/Users/lxl/Desktop/")

vcf <- read.table("vcf.txt")

bed <- read.table("bed.txt")

vcf$V4 <- NA

vcf$V5 <- NA

vcf$V6 <- NA

for (i in 1:nrow(vcf)) {

for (j in 1:nrow(bed)) {

if (vcf[i,1] == bed[j,1]) {

if((vcf[i,2] > bed[j,3]) & (vcf[i,3] < bed[j,4])){

vcf[i,4] <- bed[j,5]

vcf[i,5] <- bed[j,6]

vcf[i,6] <- bed[j,7]

}

}

}

}

4. 合并vcf.txt文件与原始vcf

join vcf.txt E-fontanierii-4.gvcf

最后去掉多余的列