

Solutions

Alignment upload, and clipping of restriction overhang

```
library(ShortRead)
# infile<-'C:/Users/daniel/Documents/science/teaching/cesky
# krumlov 2016/course.materials/R.files/CGATA.bam'
param <- ScanBamParam(flag = scanBamFlag(isUnmappedQuery = FALSE),
  what = c("rname", "pos", "seq", "qual"), reverseComplement = TRUE)
f <- scanBam(infile, param = param)[[1]]
chrom <- f$rname # extract the chromosome name
pos <- f$pos # ... the alignment position
seqs <- f$seq # ... and the sequence
seqs[1]

## A DNAStringSet instance of length 1
## width seq
## [1] 70 TGCAGGAAGCGGCCGATAAGGC...TGCTGGGCGGTGTTTGATGGG

seqs <- narrow(seqs, start = 7, end = 70)
seqs[1]

## A DNAStringSet instance of length 1
## width seq
## [1] 64 AAGCGGCCGATAAGGCGTTTTT...TGCTGGGCGGTGTTTGATGGG
```

Sequences of RAD locus with position 1,571,469

```
loc <- "1571469"  
idx <- which(pos == loc)  
stack <- seqs[idx]  
stack
```

```
## A DNASTringSet instance of length 67  
## width seq  
## [1] 64 AGATGCTGGTGACAGCTCAGC...GTCTCGTGGCGAGTGGAATCA  
## [2] 64 AGATGCTGGTGACAGCTCAGC...GTCTCGTGGTGAGTGGAATCA  
## [3] 64 AGATGCTGGTGACAGCTCAGC...GTCTCGTGGTGAGTGGAATCA  
## [4] 64 AGATGCTGGTGACAGCTCAGC...GTCTCGTGGTGAGTGGAATCA  
## [5] 64 AGATGCTGGTGACAGCTCAGC...GTCTCGTGGTGAGTGGAATCA  
## ... ..  
## [63] 64 AGATGCTGGTGACAGCTCAGC...GTCTCGTGGTGAGTGGAATCA  
## [64] 64 AGATGCTGGTGACAGCTCAGC...GTCTCGTGGTGAGTGGAATCA  
## [65] 64 AGATGCTGGTGACAGCTCAGC...GTCTCGTGGTGAGTGGAATCA  
## [66] 64 AGATGCTGGTGACAGCTCAGC...GTCTCGTGGTGAGTGGAATCA  
## [67] 64 AGATGCTGGTGACAGCTCAGC...GTCTCGTGGTGAGTGGAATCA
```

Inspect coverage and the haplotype distribution for this locus

```
length(stack)  # great coverage for genotyping!

## [1] 67

hplt <- sort(table(stack), decreasing = T)
hplt  # heterozygous!

## stack
## AGATGCTGGTGACAGCTCAGCTCAAAACACTGGCCCTGGTTGAGTCTCGTGGTGAGTGAATCA
##                                                                 34
## AGATGCTGGTGACAGCTCAGCTCAAAACACTGGCCATGGTTGAGTCTCGTGGTGAGTGAATCA
##                                                                 31
## AGATGCTGGTGACAGCTCAGCTCAAAACACTGGCCCTGGTTGAGTCTCGCGGTGAGGGGAATCA
##                                                                 1
## AGATGCTGGTGACAGCTCAGCTCAAAACACTGGCCCTGGTTGAGTCTCGTGGCGAGTGAATCA
##                                                                 1
```

Object X containing position and diploid consensus genotype for RAD locus 1,571,469

```
X <- c("pos", "seq") # empty data container with header
X <- rbind(X, c(loc, names(hplt)[1])) # 1st haplotype
X <- rbind(X, c(loc, names(hplt)[2])) # 2nd haplotype
```

Same for RAD locus 2,298,848

```
loc <- "2298848"
stack <- seqs[which(pos == loc)]
hplt <- sort(table(stack), decreasing = T)
hplt

## stack
## TCCTTCAACGAAACTGCTTTCATAGTAGACCTTTTTCAAAGCAGGCTGGAAAAAATACAGAATT 84
##
## TCCTTCAACGAAAATGCTTTCATAGTAGACCTTTTTCAAAGCAGGCTGGAAAAAATACAGAATT 1
##

X <- rbind(X, c(loc, names(hplt)[1])) # 1st haplotype
X <- rbind(X, c(loc, names(hplt)[1])) # 2nd haplotype
```

View the consensus data

```
X
##      [,1]
## X    "pos"
##      "1571469"
##      "1571469"
##      "2298848"
##      "2298848"
##      [,2]
## X    "seq"
##      "AGATGCTGGTGACAGCTCAGCTCAAAACACTGGCCCTGGTTGAGTCTCGTGGTGGTGGGAATCA"
##      "AGATGCTGGTGACAGCTCAGCTCAAAACACTGGCCATGGTTGAGTCTCGTGGTGGTGGGAATCA"
##      "TCCTTCAACGAAACTGCTTTCATAGTAGACCTTTTTCAAAGCAGGCTGGAAAAAATACAGAATT"
##      "TCCTTCAACGAAACTGCTTTCATAGTAGACCTTTTTCAAAGCAGGCTGGAAAAAATACAGAATT"
```

Write to text file

```
# write.table(X,
# file='C:/Users/daniel/Documents/science/teaching/cesky
# krumlov
# 2016/course/materials/R.files/CGATA.consensus.txt',
# row.names=FALSE, col.names=FALSE, quote=FALSE)
```