

Evolution and genomics, Cesky Krumlov

Daniel Berner

2. February 2016

Genomic analyses using RADseq:

2. Consensus genotyping

Demo

Upload a stickleback alignment to R, and inspect the elements (Illumina SE76, from Roesti et al. 2012 Mol. Ecol., limited to ChrIII only)

```
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
```

```
chrom[1]
```

```
## [1] chrIII
## Levels: chrIII
```

```
pos[1]
```

```
## [1] 10735613
```

```
seqs[1]
```

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

Sequences of RAD locus with position 68,564

```
loc <- "68564"  
idx <- which(pos == loc) # very fast pattern search  
stack <- seqs[idx]  
stack
```

```
## A DNASTringSet instance of length 32  
## width seq  
## [1] 64 AAGACTGTGGCCGGAGGCGCC...TTTTTGCTTTGCTGCATTTCT  
## [2] 64 AAGACTGTGGCCGGAGGCGCC...TTTTTGCTTTGCTGCATTTCT  
## [3] 64 AAGACTGTGGCCGGAGGCGCC...TTTTTGCTTTGCTGCATTTCT  
## [4] 64 AAGACTGTGGCCGGAGGCGCC...TTTTTGCTTTGCTGCATTTCT  
## [5] 64 AAGACTGTGGCCGGAGGCGCC...TTTTTGCTTTGCTGCATTTCT  
## ... ..  
## [28] 64 AAGACTGTGGCCGGAGGCGCC...TTTTTGCTTTGCTGCATTTCT  
## [29] 64 AAGACTGTGGCCGGAGGCGCC...TTTTTGCTTTGCTGCATTTCT  
## [30] 64 AAGACTGTGGCCGGAGGCGCC...TTTTTGCTTTGCTGCATTTCT  
## [31] 64 AAGACTGTGGCCGGAGGCGCC...TTTTTGCTTTGCTGCATTTCT  
## [32] 64 AAGACTGTGGCCGGAGGCGCC...TTTTTGCTTTGCTGCATTTCT
```

For this RAD locus, inspect coverage and the haplotype distribution

```
hplt <- sort(table(stack), decreasing = T)
hplt # seems a homozygous locus

## stack
## AAGACTGTGGCCGGAGGCGCCTGGACTTACAAGTAGGCTACTTTTTTTGCTTTGCTGCATTTCT
## 29
## AAGACTGTGGCCGGAGGCGCCTGGACTTACAAGTAGGCTACTTTTTTTGCTTTGCTGCCTTTCT
## 2
## AAGACTGTGGCCGGAGGCGCCTGGACTTACAAGTAGGCTAATTTTTTTGCTTTGCTGCATTTCT
## 1
```

Same for RAD locus with position 107,127

```
hplt <- sort(table(seqs[which(pos == 107127)]), decreasing = T)
hplt # homozygous too

##
## TTGAACGCTTCCATATCGTGTTCCTTCTTGACATACACACAAAATGCCACCTTGATGAGGACAC
## 44
## TTGAACGCTTCCATATCGTGTTCCTTCTTGACATACACACAAAATGCCACATTGATGAGGTCAC
## 1
## TTGAACGCTTCCATATCGTGTTCCTTCTTGACATACACACAAAATGCCACCTTGATGAGTACAC
## 1
```

A minimal consensus genotyping loop, compiling chromosome ID, position, and diploid consensus genotype in object A

```
loci <- c(68564, 107127)
A <- c("chrom", "pos", "seq") # empty data container with header
for (i in 1:length(loci)) {
  loc <- loci[i] # loop through the elements of 'loci'
  stack <- seqs[which(pos == loc)]
  hplt <- sort(table(stack), decreasing = T)
  A <- rbind(A, c("chrIII", loc, names(hplt)[1])) # attach haplotypes
  A <- rbind(A, c("chrIII", loc, names(hplt)[1]))
}
A
```

```
##      [,1]      [,2]
## A "chrom"    "pos"
##   "chrIII"  "68564"
##   "chrIII"  "68564"
##   "chrIII"  "107127"
##   "chrIII"  "107127"
##      [,3]
## A "seq"
##   "AAGACTGTGGCCGGAGGCGCCTGGACTTACAAGTAGGCTACTTTTTTTGCTTTTGCTGCATTTCT"
##   "AAGACTGTGGCCGGAGGCGCCTGGACTTACAAGTAGGCTACTTTTTTTGCTTTTGCTGCATTTCT"
##   "TTGAACGCTTCCATATCGTGTTCCTTCTTGACATACACACAAAATGCCACCTTGATGAGGACAC"
##   "TTGAACGCTTCCATATCGTGTTCCTTCTTGACATACACACAAAATGCCACCTTGATGAGGACAC"
```

Write object A out to text file

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

Tasks

- ▶ Upload the stickleback alignment *CGATA.bam*
- ▶ Clip the restriction overhang from the sequences
- ▶ Access the sequences for the RAD locus with position 1,571,469
- ▶ What is the sequence coverage for this RAD locus - sufficient to call a diploid genotype with confidence?
- ▶ Explore the haplotype distribution at this locus - homo- or heterozygous?
- ▶ Make an object *X* containing the diploid consensus genotype for this locus, along with its position
- ▶ Analogously, add the diploid genotype and position data from the RAD locus with position 2,298,848 to *X*
- ▶ Write the consensus genotype object into a text file