Sequence Alignment

Tianyi Shi

2020-02-03

Contents

1 Sequence Alignment in R

1.1 Packages

```r
library(Biostrings)
library(seqinr)
```

1.2 DNA sequence

```r
myScoringMat <- nucleotideSubstitutionMatrix(match = 1, mismatch = -1, baseOnly = TRUE)
dnaSeq1 <- "GAATTCGGCTA" # can be a plain string
dnaSeq2 <- DNAString("GATTACCTA") # or an XString
gapOpen <- 2
gapExtend <- 1
dnaAln <- pairwiseAlignment(dnaSeq1, dnaSeq2, substitutionMatrix = myScoringMat, gapOpening = gapOpen, 
dnaAln
```

1.3 Amino acid sequence

Here is a list of built-in amino acid substitution matrices, which can be shown with `data(package="Biostrings")`:

```
BLOSUM100
BLOSUM45
BLOSUM50
BLOSUM62
BLOSUM80
PAM120
PAM250
PAM30
PAM40
PAM70
```

```
aaSeq1 <- 'PAWHEAE'
aaSeq2 <- 'HEAGAWGHE'
aaAln <- pairwiseAlignment(aaSeq1, aaSeq2, substitutionMatrix="BLOSUM62", gapOpening=gapOpen, gapExtens:
aaAln
```

```
ncov <- readDNAStringSet('./src/2019-ncov.fasta', 'fasta')
sars <- readDNAStringSet('./src/sars.fasta', 'fasta')
```

```
virusAln <- pairwiseAlignment(ncov, sars, substitutionMatrix = myScoringMat, gapOpening = gapOpen, gapE:
```

```
ncov1 <- seqinr::read.fasta('./src/2019-ncov.fasta', seqtype = 'DNA')
sars1 <- seqinr::read.fasta('./src/sars.fasta', seqtype = 'DNA')
```

```
library(ape)
myset <- c("U15717", "U15718", "U15719", "U15720", "U15721","U15722", "U15723", "U15724")
myseqs <- read.GenBank(myset)
mydist <- dist.dna(myseqs)
plot(myphylo, type="phylogram", edge.color="red", cex=1, edge.width=1,main="(A) Phylogram")
```