

S4 classes - Gene sequence class using slots

```
source("./R/Sequence.R")
```

Function to read sequences

```
readFasta <- function(infile){
  lines <- readLines(infile)
  header <- grep(">", lines)
  cat("Sequences found at lines: ", header, "\n")
  if (length(header) > 1) {
    warning("Reading first sequence only.\n")
    lines <- lines[header[1]:(header[2]-1)]
    header <- header[1]
  }
  .id <- sub("> *", "", lines[header], perl=TRUE)
  .sequence <- toupper(paste(lines[(header+1):length(lines)], collapse=""))
  .alphabet <- toupper(unique(strsplit(.sequence, "")[[1]]))
  if (all(.alphabet %in% c("A", "C", "G", "T"))) {
    newseq <- DnaSeq(.id, .sequence)
  } else if (all(.alphabet %in% c("A", "C", "G", "U"))) {
    newSeq <- RnaSeq(.id, .sequence)
  } else {
    stop("Alphabet ", .alphabet, " is unknown.")
  }
  if (validObject(newseq))
    return(newseq)
}
```

```
fastafilename <- dir(path="./inst/extdata",
                     full.name = TRUE,
                     pattern = "fasta$")
fastafilename
```

```
## [1] "./inst/extdata/aDnaSeq.fasta"      "./inst/extdata/moreDnaSeqs.fasta"
```

Read 1st sequence

```
myseq <- readFasta(fastafilename[1])
```

```
## Sequences found at lines: 1
```

```
myseq
```

```
## Object of class DnaSeq
```

```
## Id: example dna sequence
```

```
## Length: 132
```

```
## Alphabet: A C G T
```

Sequence: AGCATACGACGACTACGACACTACGACATCAGACACTACAGACTACTACGACTACAGACATCAGACACTACATATTTACATCATCAGAG.

```
transcribe(myseq)
```

```

## Object of class RnaSeq
## Id: example dna sequence -- transcribed
## Length: 132
## Alphabet: A C G U
## Sequence: AGCAUACGACGACUACGACACUACGACAUCAGACACUACAGACUACUACGACUACAGACAUCAGACACUACAUUUUACAUCAGAGAG
alphabet(myseq)

## [1] "A" "C" "G" "T"
seq(myseq)

## [1] "AGCATACGACGACTACGACACTACGACATCAGACACTACAGACTACTACGACTACAGACATCAGACACTACATATTTACATCATCAGAGATTATAT
print(myseq)

## > example dna sequence
## 1   AGCATACGA
## 10  CGACTACGAC
## 20  ACTACGACAT
## 30  CAGACACTAC
## 40  AGACTACTAC
## 50  GACTACAGAC
## 60  ATCAGACACT
## 70  ACATATTTAC
## 80  ATCATCAGAG
## 90  ATTATATTAA
## 100 CATCAGACAT
## 110 CGACACATCA
## 120 TCATCAGCAT
## 130 CAT
rev(myseq)

## [1] "TACTACGACTACTACTACACAGCTACAGACTACAATTATATTAGAGACTACTACATTTATACATCACAGACTACAGACATCAGCATCATCAGACA
comp(myseq)

## [1] "TCGTATGCTGCTGATGCTGTGATGCTGTAGTCTGTGATGTCTGATGATGCTGATGTCTGTAGTCTGTGATGTATAAATGTAGTAGTCTCTAATAT
length(myseq)

## [1] 132
myseq[5:10]

## Object of class DnaSeq
## Id: example dna sequence
## Length: 6
## Alphabet: A C G T
## Sequence: TACGAC
myseq[5, 10]

## Object of class DnaSeq
## Id: example dna sequence
## Length: 1
## Alphabet: A C G T
## Sequence: T

```

```
# T
```

```
myseq[1, 10]
```

```
## Object of class DnaSeq
## Id: example dna sequence
## Length: 1
## Alphabet: A C G T
## Sequence: A
```

```
# A
```

Read the 2nd sequence

```
myseq2 <- readFasta(fastafilename[2])
```

```
## Sequences found at lines: 1 4 9
```

```
## Warning in readFasta(fastafilename[2]): Reading first sequence only.
```

```
myseq2
```

```
## Object of class DnaSeq
## Id: sequence 1
## Length: 79
## Alphabet: A C G T
## Sequence: AGTACGTAGTAGCTGCTGCTACGTGCGCTAGCTAGTACGTCAAGGTGAGATACGACGTAGATGCTAGCTGACTCGATGC
```

```
# using setReplaceMethod
```

```
id(myseq2) <- "Sequence #1"
```

```
myseq2
```

```
## Object of class DnaSeq
## Id: Sequence #1
## Length: 79
## Alphabet: A C G T
## Sequence: AGTACGTAGTAGCTGCTGCTACGTGCGCTAGCTAGTACGTCAAGGTGAGATACGACGTAGATGCTAGCTGACTCGATGC
```

```
# using setReplaceMethod
```

```
`id<-`(myseq2, "this is sequence no. 1")
```

```
## Object of class DnaSeq
## Id: this is sequence no. 1
## Length: 79
## Alphabet: A C G T
## Sequence: AGTACGTAGTAGCTGCTGCTACGTGCGCTAGCTAGTACGTCAAGGTGAGATACGACGTAGATGCTAGCTGACTCGATGC
```