

README

SequenceTranslationForDF

PURPOSE:

Given a dataframe with a column of nucleotide sequences, translate the sequences and append a new column with translated strings.

REQUIREMENTS:

R packages:

- seqinr
- dplyr

```
library(seqinr)
library(dplyr)
```

METHOD:

1. Load a dataframe into R
2. The function, “translateFun” executes the following:
 - takes a nucleotide sequence argument
 - converts the sequence string to a vector of characters
 - translate the characters (seqinr::translate)
 - converts the translated characters to string format
3. Apply the function to the dataframe and use mutate() to add a new column as output

EXAMPLE:

Given the following dataframe (df) of two rows:

#Make a test dataframe with two sequences

```
ntd <- c("agctgctagt", "aagtcgc")
```

```
df <- as.data.frame(ntd)
df
```

```
##           ntd
## 1 agctgctagt
## 2   aagtcgc
```

Create the function

```
translateFun <- function(nuc) {
  translatedSeq <- c2s(translate(s2c(nuc)))
  return(translatedSeq)
}
```

Apply the function to the dataframe

```
newDF<- df %>%
  select(ntd) %>%
  mutate (newCol = apply(df, 1, translateFun))
newDF
```

```
##           ntd newCol
## 1 agctgctagt   SC*
## 2   aagtcgc    KS
```

- “newCol” = new column appended to df, containing the translated sequence(s)