

CountTableGenerator

Joshua Burkhardt

9/1/2017

```
transcriptome_fasta %>% length()

## [1] 80322

seqs_csv %>% nrow()

## [1] 70540

ma1_num_seq <- mas[[1]] %>% nrow()
for(ma in 2:length(mas)){
  assertthat::assert_that(nrow(mas[[ma]]) == ma1_num_seq)
}
print(ma1_num_seq)

## [1] 67626

seqs_csv$Status %>% table()

## .
##      0      1      2      3
## 10660 38270 16747  4863

seqs_csv$Paralog.. %>% table()

## .
##      0      1      2      3      4      5      6      7      8      9     10
## 65677 3441  944  287  101   47   21   11    6    3    2

table(seqs_csv$Status, seqs_csv$Paralog..)

##
##           0      1      2      3      4      5      6      7      8      9     10
## 0 10660      0      0      0      0      0      0      0      0      0      0
## 1 38270      0      0      0      0      0      0      0      0      0      0
## 2 16747      0      0      0      0      0      0      0      0      0      0
## 3      0 3441  944  287  101   47   21   11    6    3    2

seqs_csv <- seqs_csv %>%
  dplyr::rowwise() %>%
  dplyr::mutate(SeqType = ifelse(grepl("CONTIG",ID),
                                "CONTIG",
                                ifelse(grepl("F5",ID),
                                        "SINGLETON",
                                        "UNKNOWN")))

seqs_csv$SeqType %>% table()

## .
##      CONTIG SINGLETON
##      24438      46102
```

```
table(seqs_csv$SeqType, seqs_csv$Status)
```

```
##  
##           0      1      2      3  
##  CONTIG    4601  7210 11125  1502  
##  SINGLETON 6059 31060  5622  3361
```

```
table(seqs_csv$SeqType, seqs_csv$Paralog..)
```

```
##  
##           0      1      2      3      4      5      6      7      8      9  
##  CONTIG    22936  1274   187    30     7     3     1     0     0     0  
##  SINGLETON 42741  2167   757   257    94    44    20    11     6     3  
##  
##           10  
##  CONTIG         0  
##  SINGLETON      2
```