Shared Peptides Explorer for Mass Spectrometry

Aneta Foks
Karolina Krzyszkiewicz

Main functionalities:

- listing tryptic peptides for a given set of proteins,
- finding shared peptides,
- interactive visualization of shared peptides,

```r
### data loading function:
load_file <- function(path, file, sep){
  lst_file <- list.files(path, pattern = file)
  lst_table <- lapply(lst_file, read.table, header = TRUE, sep = sep, stringsAsFactors = FALSE)
  do.call(rbind, lst_table)
}

### filter function:
filter_table <- function(table, col_1, col_2, col_3, q_value){
  table <- filter(table, table[[col_3]] < q_value)
  table <- select(table, col_1, col_2)
}

### separation of data in the column:
separate_table <- function(table, col, sep){
  table_separate <- cSplit(table, col, sep)
}

### one table from two columns; reject "<NA>":
connect_table <- function(table_separate){
  table_protein <- table_separate[,1:2]
  for (i in seq(from = 3, to = dim(table_separate)[2])){
    table_protein <- rbind(table_protein,
                           filter(select(table_separate, 1, i),
                                  table_separate[[colnames(table_separate)[i]]] != "<NA>"),
                           use.names = FALSE)
  }
  table_protein
}
```

```r
### delete duplicate rows
unique_table <- function(table){
  table <- unique(table)
}

### columns names change
change_name_cols <- function(table, name_cols){
  colnames(table) <- name_cols
  table
}

### sort table
sort_table <- function(table, name_col){
  table_sort <- table[order(table[[name_col]])]
}

### include table
create_table_0_1 <- function(table){
  table <- as.data.frame.matrix(table(table))
}

### count values peptides/proteins:
count_table <- function(table, id){
  if (id == 2){
    table_qty <- data.frame(colnames(table), as.vector(apply(table, 2, sum)))

  } else {
    table_qty <- data.frame(rownames(table), as.vector(apply(table, 1, sum)))
  }
  table_qty
}
```

```r
### import function:
import_comet_percolator <- function(path, file,
                                    sep, sep_2,
                                    col_1, col_2, col_3,
                                    q_value, name_cols){
  table <- load_file(path, file, sep)
  table <- filter_table(table, col_1, col_2, col_3, q_value)
  table <- separate_table(table, col_2, sep_2)
  table <- connect_table(table)
  table <- unique_table(table)
  table <- change_name_cols(table, name_cols)
  table
}
```

```
table_proteins_peptides <- import_comet_percolator(path, file,
                                                   sep, sep_2,
                                                   col_1, col_2, col_3,
                                                   q_value, c("peptides", "proteins"))
```

```
> head(table_proteins_peptides)
                            peptides    proteins
1: RLPFPEPYILVYANDAAISEPESVVSSLQGHR HPRR2310052
2:               NAYHNVTAEQLFLKDIIEK HPRR4160484
3:       TVEFQHIIPISAVTGEGIEELKNCIR HPRR2440154
4:            KLSFIEALIQQYQEQLDKSTK HPRR2551185
5:  TPDPAPEKPSESSAGPSTEEDFAVDFEKIYK HPRR3720532
6:        GSHPWQVALLSGNQLHCGGVLVNER HPRR2390055
```

```r
### manipulate function:
manipulation_table <- function(table, name_col, id, name_cols){
  table <- sort_table(table, name_col)
  table <- create_table_0_1(table)
  table <- count_table(table, id)
  table <- change_name_cols(table, name_cols)
  table
}
```

```
table_qty_peptides <- manipulation_table(table_proteins_peptides,
                                         "proteins", 2,
                                         c("proteins", "qty_peptides") )
```

```
> head(table_qty_peptides)
     proteins qty_peptides
1 HPRR1010001            8
2 HPRR1060049            2
3 HPRR1310038            1
4 HPRR1370013            7
5 HPRR1370116            1
6  HPRR140662           10
```

```
table_qty_proteins <- manipulation_table(table_proteins_peptides,
                                         "proteins", 1,
                                         c("peptides", "qty_proteins") )
```

```
> head(table_qty_proteins)
             peptides qty_proteins
1        AADFIDQALAQK             1
2     AADFIDQALAQKNGR             1
3      AADIHTFSSVLEKV             1
4             AALSIER             2
5         AAQPSPSTVPK             2
6 AAQPSPSTVPKTEDQRPQ             2
```

Visualization