Create fasta DB

Witold Wolski

30 September 2015

```
library(pepfdr)
```

Create db

Contaminants and mscqc1 proteins

```
rm(list=ls())
file = file.path(path.package("pepfdr"),"extdata/fgcz_contaminants_20150123.fasta")
contaminants <- readPeptideFasta(file)
msqc1 <- readPeptideFasta("../data/fastaFiles/msqc1-sequences.fasta")
```

remove mscq1 proteins from contaminants, so there are no duplicates.

```
toremove <- NULL
for(i in 1:length(msqc1)){
  idx <- grep(msqc1[[i]],contaminants)
  toremove<-c(toremove,idx)
}
print(toremove)
```

```
## [1]  33  24 197 200 193
```

```
length(contaminants)
```

```
## [1] 263
```

```
contNoMSQC1 <- contaminants[-toremove]
length(contNoMSQC1)
```

```
## [1] 258
```

Create Reverse Sequences

```
contaminantsrev <- reverseSeq(contNoMSQC1)
msqc1rev <- reverseSeq(msqc1)
```

Prepare e-coli and human databases

```
ecoli <- readPeptideFasta("../data/fastaFiles/uniprot-taxonomy83333.fasta")
human <- readPeptideFasta("../data/fastaFiles/uniprot-taxonomyHomoSapiensHuman9606.fasta")

length(ecoli)
```

```
## [1] 6098
```

```
length(human)
```

```
## [1] 20197
```

```
ecoliRev <- reverseSeq(ecoli)
humanRev <- reverseSeq(human)
```

Create new database, with reverse and forward sequences

```
all_d <-c(msqc1, ecoli, human, contNoMSQC1,  msqc1rev, ecoliRev, humanRev , contaminantsrev)
length(all_d)/2
```

```
## [1] 26559
```

```
stopifnot(length(all_d)/2 == length(ecoli) + length(human) + length(msqc1) + length(contNoMSQC1))
writeFasta(all_d, file="../data/fastaFiles/output/p1755_db1_d_20151016_msqc1ecolihuman.fasta")

all <- c(msqc1, ecoli, human, contNoMSQC1)
stopifnot(length(all) == length(ecoli) + length(human) + length(msqc1) + length(contNoMSQC1))

writeFasta(all, file="../data/fastaFiles/output/p1755_db1_20151016_msqc1ecolihuman.fasta")
```