SEQUENCE OUTPUT TEST

GSI Bioinformatic Support

27 September, 2019

```
library(knitr)




formatDNA<-function(fasta.file){
  dna<-scan(fasta.file,what="character")
  dna<-dna[-1]
  dna<-paste(dna,collapse="")

  dna.lines<-strsplit(dna, '(?<=.{60})', perl = TRUE)[[1]]
  ### set the positions
  n<-60
  names(dna.lines)<-seq(from=1,by=n,length.out=length(dna.lines))
  dna.df<-as.data.frame(dna.lines)

  dna.chunked<-apply(dna.df,1,function(x){paste(strsplit(x, "(?<=.{10})", perl = TRUE)[[1]],collapse=" "
  dna.chunked.df<-as.data.frame(dna.chunked)
  colnames(dna.chunked.df)<-NULL
  dna.chunked.df
}

fasta.file<-"dna.fasta"
dna<-formatDNA(fasta.file)
print(dna,right=FALSE)
```

```
1    GCTCCCGGCT TAGAGGACAG CGGGGAAGGC GGGCGGTGGG GCAGGGGGCC TGAAGCGGCG
61   GTACCGGTGC TGGCGGCGGC AGCTGAGGCC TTGGCCGAAG CCGCGCGGTG AGTCTAGGGC
121  CTGGCACGAC CCCTCTAGGG CGGCGTAATG TCCAGACCCA CGGCCAGCCT GCCCGTGGGG
181  GGGCCAGGGG AAGCCGCCCG TCTGGGACGT GGGGTCCCCA CATTGCCGGC CCCACGACCT
241  GGGCACCGTC TGAGGCCTTG ACTCCCACCC CTCGAGGCAA CGCCCACCTC CCCGGGACCC
301  CCAAGAGACC CCTGGACCCT TCGGCCCGCG GTACGTCCGC CCGAAGGCCG GCCCATCAGC
361  TCTGGTATTG CCCTTTGGGT CTTTTGCCTG TCACAGCCAC CCTTTCTTTT GAGCCACGTA
421  TAAAAATTGT GTATCTCTGC CCCAGGGAAC ACTAGCGTCC GTGTCAGCCC CTTGTCCTCC
481  ACTGTGGACA CCTCTCAGAG ATCCAGTCCC CGAAACTGAG CTTGGTTACA TCGTTGGGGT
541  ACATCTTCTC TCCCTCTCCC CCAGCCCCAT CCCTGTTTCC CCCATCCAAT CTTAACTTCC
601  TTAAGCCCTA TCAACACCAT TAGGATATTT GACTTCAGAT ATCCTAAGTT TAATTGAATT
661  CAGTCTGGAG CAGATGGCCT GTGGGCCTCA GATCATATAA AGATACACTG GTTCTTTCCT
721  GAGAATAGAA ATCCCTTGCC AGCCACCTTC CCCCCTTTCG GACGCACACA CATACACACA
781  AACTGGAGTA GTTTTCTTAG CAGGGATTCT TAAGTTTCTT CTCCCCTAAA GGATGACATT
841  TCTTTCTGCA GTCTTCTTCT TGGCAGTGGA GTATTTGAAA GCTTTACAAA ACCAATTATT
901  CCCAGGTTTT TCCTCTGTGC CTTACAAATT CTTTCAAAAA TAAGAATTTT GAGAAATTGG
961  TTTTGTAGCA TAAGCTGAAC ATACTTGGGA TAGGTGTATG TTACACCCAT GGCAGTGTGG
1021 GCATAATTGA GGAATGAGGG AATGAGCTCA GGAACTAATT GGTGTTTTTT GTTTTGTTTT
1081 GAGACGGAGT TTCGCTCTGT CGCCCAGGCT GGAGTGCAGT GGCTCGATCT CTGCTCACTG
1141 CAAGCTCCGC CTCCCGGGTT CCCGCCATTC TCCTGCCTCA GCCTCCCGAG TAGCTGGGAC
```

```
1201 TACAGGCGCC CGCCACCATA CGGCTAATTT TTTTTTTTTT TTGTATTTTT AGTAGAGGCG
1261 GGATTTCACC GTGTTGGCCA GGATGGTCTC GATTTCCTGA CCTCGTGATC CGCCCGCCTC
1321 GGCCTCCCAA ACCCGGCCAG GAACTAATTA CTTTCTTTTA GCTTACATTT GAAGAGCTAG
1381 TCCCCTGTAG AACTGCGTTG TTCAGTACAA GAACCACCAG CCACATGTGG CTGTGGAGCA
1441 CTTGAAATGT GACTAGTCCA GGCCGGGCAT GGTGGCTCAC ACCTGTAATC CCAGCACTTT
1501 GGGAGGCCAA GGCGGGCAGA TCACCTGAGG TCAGGAATTT GAGACCAGCC TGGCCAACAT
1561 AATAAAACCC CATCTCTACT AAAAACACAC AAAAATTAGC CAGGCGTGGT GCTGCACATC
1621 TGTAGTCCCA GCTACTCGGG AAGCTGAGGC AGGAGAATCA TTTGAACCCG GGAGGTGGAT
1681 GTTGCAGTGA GCCGAGATCA TGCCACTGCA CTCCAGCCTG GGTGACAGAG CGAGACTCTG
1741 TCTCAAAAAA AAAAAAAAAA ATGACTAGTC CAAATTGACA TTGTTGTAAG TGTAAAATGC
1801 ACATTAGATT TTGAAGACTC GCTAAAAAAA AAGAATGGAC ACTATATCAA TTTTTTAAAT
1861 ATTGATAACA TGTTGAAATG ATATTTTGGA TATATTGGTT TAGGTAATTA ATTTCACCTG
1921 TTTCTTTTTA CCTTTTAAAA TATGGCTACA AGCAGCCTTA AAATTACATT TGTGGATCAT
1981 TGTATTTCCT CTTTTTTTTT TTTTTAATTG AGATGGAGTT TTGCTCTTGT TGCCCAGGCT
2041 GTAGTGCAAT GGCCCAATCT CAGCTCACTG CAACCTCTGC CT
```