

一个R包的公开处刑

作者: rabbit (hui1303@j.smu.edu.cn)

时间: 2019-11-27

1.安装和加载

```
#install
devtools::install_github("PPEachi/rpackage/rabbitpackage")
#library
library(rabbitpackage)
```

报错了请及时联系作者（作者多半也不会解决

2.批量下载genbank数据

因为最后只需要对两条序列进行全局比对，这里只下载两个gb文件

首先拿两个你感兴趣的“accession number”

```
accn<-c("AJ534528", "AJ534529")
```

然后调用函数 `download_GBorFASTA`

```
download_GBorFASTA(accn, "nucleotide", "gb", "text")
```

如果不知道怎么用可以用 `?download_GBorFASTA` 查询一下，不过因为这个描述文件也是作者自己写的，看不懂就直接联系我八

运行完你就可以在你的目录下看到多出了两个gb文件！

3.genbank转fasta

得到了两个gb文件后，我们需要再调用一个函数 `genbank2fasta` 来将gb格式转成fasta格式，然后保存为“.fas”结尾的fasta文件

用到了上一步创建好的变量“accn”

```
for (i in accn){
  genbank2fasta(i)
}
```

再废话一句，如果不知道怎么用可以用 `?genbank2fasta` 查询一下

运行完你就可以在你的目录下看到又多出了两个fasta文件！

4.读入fasta文件并且计算碱基含量

首先读入fasta文件，用到函数 `read_fas`

```
x<-read_fas(accn[1])
y<-read_fas(accn[2])
```

这样就可以得到两个数据框格式的序列数据，第一列是accn，第二列是序列

然后计算碱基含量

```
bx<-base_freq(x)
by<-base_freq(y)
```

结果示意

```
## 1 sequences in total

## #Labels:
## AJ534528
##
## #Base composition:
##      length      A      T      G      C
## AJ534528  1143 0.2729659 0.2432196 0.1268591 0.3534558
```

5.全局比对

第一步就是得到两条序列，所以我们用到上一步获得的序列数据

序列保存在数据框的第二列，我们将它们取出来

```
x1<-x$seq
y2<-y$seq
```

调用函数 `global_align`，这里为了结果好看，只各自取了两条序列的前十个字符，另外比对得分可以改成你们喜欢的

```
global_align(paste(unlist(strsplit(x1,split = ""))[1:10], collapse =
""),paste(unlist(strsplit(y2,split = ""))[1:10], collapse = ""),1,-1,-3)
```

结果示意

```
## Sequence x: atggcncnca
## Sequence y: atggcaccca
## Scoring system: 1 for match, -1 for mismatch, -3 for gap
##
## Dynamic programming matrix 1:
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
## [1,]    0  -3  -6  -9 -12 -15 -18 -21 -24  -27  -30
```

```

## [2,] -3 1 -2 -5 -8 -11 -14 -17 -20 -23 -26
## [3,] -6 -2 2 -1 -4 -7 -10 -13 -16 -19 -22
## [4,] -9 -5 -1 3 0 -3 -6 -9 -12 -15 -18
## [5,] -12 -8 -4 0 4 1 -2 -5 -8 -11 -14
## [6,] -15 -11 -7 -3 1 5 2 -1 -4 -7 -10
## [7,] -18 -14 -10 -6 -2 2 4 1 -2 -5 -8
## [8,] -21 -17 -13 -9 -5 -1 1 5 2 -1 -4
## [9,] -24 -20 -16 -12 -8 -4 -2 2 4 1 -2
## [10,] -27 -23 -19 -15 -11 -7 -5 -1 3 5 2
## [11,] -30 -26 -22 -18 -14 -10 -6 -4 0 2 6
## Dynamic programming matrix 2:
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
## [1,] "0"  "←"  "←"  "←"  "←"  "←"  "←"  "←"  "←"  "←"  "←"
## [2,] "↑"  "↖"  "←"  "←"  "←"  "←"  "↖←"  "←"  "←"  "←"  "↖←"
## [3,] "↑"  "↑"  "↖"  "←"  "←"  "←"  "←"  "←"  "←"  "←"  "←"
## [4,] "↑"  "↑"  "↑"  "↖"  "↖←"  "←"  "←"  "←"  "←"  "←"  "←"
## [5,] "↑"  "↑"  "↑"  "↖↑"  "↖"  "←"  "←"  "←"  "←"  "←"  "←"
## [6,] "↑"  "↑"  "↑"  "↑"  "↑"  "↖"  "←"  "↖←"  "↖←"  "↖←"  "←"
## [7,] "↑"  "↑"  "↑"  "↑"  "↑"  "↑"  "↖"  "↖←"  "↖←"  "↖←"  "↖←"
## [8,] "↑"  "↑"  "↑"  "↑"  "↑"  "↖↑"  "↖↑"  "↖"  "↖←"  "↖←"  "←"
## [9,] "↑"  "↑"  "↑"  "↑"  "↑"  "↑"  "↖↑"  "↑"  "↖"  "↖←"  "↖←"
## [10,] "↑"  "↑"  "↑"  "↑"  "↑"  "↖↑"  "↖↑"  "↖↑"  "↖"  "↖"  "←"
## [11,] "↑"  "↖↑"  "↑"  "↑"  "↑"  "↑"  "↖"  "↑"  "↑"  "↖↑"  "↖"
## Dynamic programming matrix 3:
##      [,1] [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
##      ## [,9]
## [1,] "0"  "left"      "left" "left"      "left"      "left"      "left"
## "left"      ## "left"
## [2,] "up" "diag"      "left" "left"      "left"      "left"      "diag_left"
## "left"      ## "left"
## [3,] "up" "up"        "diag" "left"      "left"      "left"      "left"
## "left"      ## "left"
## [4,] "up" "up"        "up"  "diag"      "diag_left" "left"      "left"
## "left"      ## "left"
## [5,] "up" "up"        "up"  "diag_up" "diag"      "left"      "left"
## "left"      ## "left"
## [6,] "up" "up"        "up"  "up"      "up"      "diag"      "left"
## "diag_left" ## "diag_left"
## [7,] "up" "up"        "up"  "up"      "up"      "up"      "diag"
## "diag_left" ## "diag_left"
## [8,] "up" "up"        "up"  "up"      "up"      "diag_up"  "diag_up"
## "diag"      ## "diag_left"
## [9,] "up" "up"        "up"  "up"      "up"      "up"      "diag_up"  "up"
##      ## "diag"
## [10,] "up" "up"        "up"  "up"      "up"      "diag_up"  "diag_up"
## "diag_up"   ## "diag"
## [11,] "up" "diag_up"  "up"  "up"      "up"      "up"      "diag"      "up"
##      ## "up"
##      [,10]      [,11]
## [1,] "left"      "left"
## [2,] "left"      "diag_left"
## [3,] "left"      "left"
## [4,] "left"      "left"
## [5,] "left"      "left"
## [6,] "diag_left" "left"
## [7,] "diag_left" "diag_left"
## [8,] "diag_left" "left"
## [9,] "diag_left" "diag_left"

```

```
## [10,] "diag"      "left"
## [11,] "diag_up"   "diag"
##
## Alignment:
## x:  a t g g c n c n c a
##      | | | | |   |   | |
## y:  a t g g c a c c c a
##
## #1 score: 6
## #2 hamming-distance: 2
```

♥谢谢你们看我表演♥
