

KK4D: 用于共线性分析, Ka,Ks,4DTv的计算

安装

```
bash Install.sh
```

```
source ~/.bashrc
```

测试是否安装成功

```
KK4D.sh -h
```

使用方法

1.修改config.ini里面的参数值,复制到工作目录

2.运行KK4D.sh all -c path/config.ini

现在也可以在命令行里指定具体参数来运行

```
KK4D.sh all -group 1 -cpu 24 -key ID -type mRNA -sample M.domestica -abbr Mdo -gff3 gene_models_20170612.gff3.gz -protein  
/share/home/Mdo.pep.fa -cds /share/home/Mdo.cds.fa -chrnum 17
```

```
KK4D.sh all -group 2 -cpu 32 -key ID ID -type mRNA mRNA -sample A.trichopoda M.domestica -abbr Ath Mdo -gff3 Ath.chr1.gff3 Mdo.chr1.gff3 -  
protein Ath.pep.fa.gz Mdo.genome.protein.fa -cds Ath.cds.fa.gz Mdo.cds.fa -chrnum 1 1
```

输入文件

Prepare Input file(can be normal or *.gz)

```
genome.gff3  
genome.pep.fa  
genome.cds.fa
```

详细使用方法

#for coline analysis

- KK4D.sh coline -c config.ini

#for kaks calculator

- KK4D.sh kaks -c config.ini

#for 4DTv calculator

- KK4D.sh 4DTV -c config.ini

#from gff3 get bed

- KK4D.sh bed -c config.ini

#from gff3 and protein.fa ,get the protein sequence

- KK4D.sh pep -c config.ini

#from gff3 and cds.fa, get the longest transcript sequence

- KK4D.sh cds -c config.ini

#from gff3 cds.fa protein.fa ,get 1 or 2 species all the above information.

- KK4D.sh all -c config.ini

Config.ini参数

#请一定要修改工作路径 默认运行是在软件安装的路径。建议新建一个目录，把文件都放在新的目录下，然后在此处指定为该目录。

工作路径

WorkPath="/share/home/database/test"

#1个基因组的设置

group=1 #样本组数，只能是1或2，应该和输入的参数的数目一致

key=(ID) #gff3文件的第9列的值

type=(mRNA) #gff3文件的第3列的字符

sample=(M.domestica) #物种的拉丁学名

abbr=(Mdo) #物种的三字符缩写

gff3=(gene_models_20170612.gff3.gz) #gff3文件，可以是gff3或gff3.gz

protein=(Mdo.pep.fasta) #蛋白文件，可以是fa或fa.gz

cds=(Mdo.cds.fasta) #cds文件，可以是cds或cds.gz

chrnum=(17) #染色体的数量（如果是scaffold，则设置共线性需要展示的scaffold的数量，一般scaffold设置为120）

#2个基因组的设置

group=2

key=(transcript_id ID) #gff3文件的第9列的值

type=(mRNA mRNA) #gff3文件的第3列的字符

sample=(A.trichopoda S.pommunis)

abbr=(Ath Spo)

gff3=(Ath.chr1.gff3 Spo.chr1.gff3)

protein=(Ath.pep.fa.gz Spo.genome.protein.fa)

cds=(Ath.cds.fa.gz Spo.cds.fa)

chrnum=(1,1)

输出结果

输出文件:

Ath_Spo.kaks4DTv.csv

Ath_Spo.kaks.csv

Ath_Spo.4DTv.csv

Seq	dtv	Ka	Ks	Ka/Ks
AT1G01010-Sp01G008830	0.333333333			
AT1G01020-Sp01G008670	0.25			
AT1G01030-Sp01G008660	0.529411765	0.521695	2.90182	0.179782
AT1G01040-Sp01G008650	0.363934426	0.143627	3.90695	0.036762
AT1G01050-Sp01G008640	0.260273973	0.081212	1.79963	0.045127
AT1G01060-Sp01G008590	0.422077922	0.424685	2.99626	0.141739
AT1G01070-Sp01G009000	0.4			
AT1G01090-Sp01G008310	0.415584416	0.119408	1.87359	0.063732
AT1G01100-Sp01G008270	0.333333333	0.192091	3.17101	0.060577
AT1G01110-Sp01G008190	0.404411765	0.266776	3.22401	0.082747
AT1G01120-Sp01G008140	0.6			
AT1G01140-Sp01G008130	0.34375	0.191443	3.67887	0.052038
AT1G01160-Sp01G008100	0.456521739	0.193255	3.49602	0.055279
AT1G01190-Sp01G007950	0.435028249	0.226201	2.89795	0.078056
AT1G01200-Sp01G007890	0.447368421	0.176725	3.6124	0.048922
AT1G01210-Sp01G007830	1			
AT1G01220-Sp01G007810	0.428571429			
AT1G01225-Sp01G007800	0.444444444			
AT1G01230-Sp01G007780	0.412698413	0.097985	2.39421	0.040926
AT1G01240-Sp01G007690	0.333333333			
AT1G01250-Sp01G007640	0.524590164	0.226618	3.06744	0.073878