

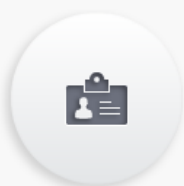
魔法酶切（Magic Enzyme Cutter）软件开发

作者：毛凤彪

学号：201318008515006

导师：孙中生

中科院.北京生命科学研究院



队长：毛凤彪
学号：201318008515006
研究所：中国科学院大学
邮箱：maofengbiao@gmail.com

团队成员：



姓名：蔡万世
学号：201118008515008
研究所：中国科学院大学
邮箱：524573104@qq.com



姓名：文艳玲
学号：201328016715035
研究所：中国科学院大学
邮箱：wyling19@126.com

2013 年 12 月 22 日

摘 要

为嵌合酶切技术在第二代技术的充分和合理运用,开发出一个灵活的、全面的、高效的、友好的酶切模拟和评估软件是迫在眉睫。因此,我们将开发一款名为“魔法酶切”(magic enzyme cutter)的软件,旨在完成和实现以下目的:

- (1) 采用 BM 算法大规模高效率地模拟单酶或多酶酶切,得到每一个酶切片段的黏端信息和碱基序列,解决以往酶切工具的基因组大小限制问题。
- (2) 统计和显示酶切片段大小、GC 含量和碱基平衡性,以评估特定长度大小片段的扩增效率。
- (3) 以电泳图的形式展示模拟酶切结果,形象生动的展示酶切效果。
- (4) 计算和统计特定长度范围上的特定序列的比例,为特定碱基或序列靶向测序提供依据和支持。
- (5) 使用最优解算法 PSO 优化酶切的区域范围的选择,使所选区域范围对目标碱基或特征序列测序效能最高。

最终将上面的结果以网页的形式汇总成评估报告,方便研究者和使用者一目了然地知道模拟酶切效果。



魔法酶切（Magic Enzyme Cutter）使用手册

1、快速教程

Magic-Enzyme-Cutter (MEC)

作者：毛凤彪、蔡万世、文艳玲.

联系邮箱: maofengbiao@gmail.com.

版本号: 1.0.

软件描述

Magic-Enzyme-Cutter 是操纵限制性内切酶 II 体外模拟酶切&&多方面评价消化降解产物特性&&计算最优选择的一个软件(MEC).

平台和环境

系统: Linux or UNIX

软件: 需要系统已经安装 perl >= v5.8.8, R >= R-2.15.1

perl 下载链接：<http://www.perl.org/>

R 下载链接：<http://www.r-project.org/>

如何使用 MEC

1. 安装方法

```
>tar xzvf MEC-xxx.tar.gz
```

```
>cd MEC-xxx
```

```
>sh install /path/to/R
```

2. 使用软件

```
>cd MEC-xxx
```

```
>./Magic-Enzyme-Cutter -help
```






3. 测试软件

```
>cd MEC-xxx/test
```

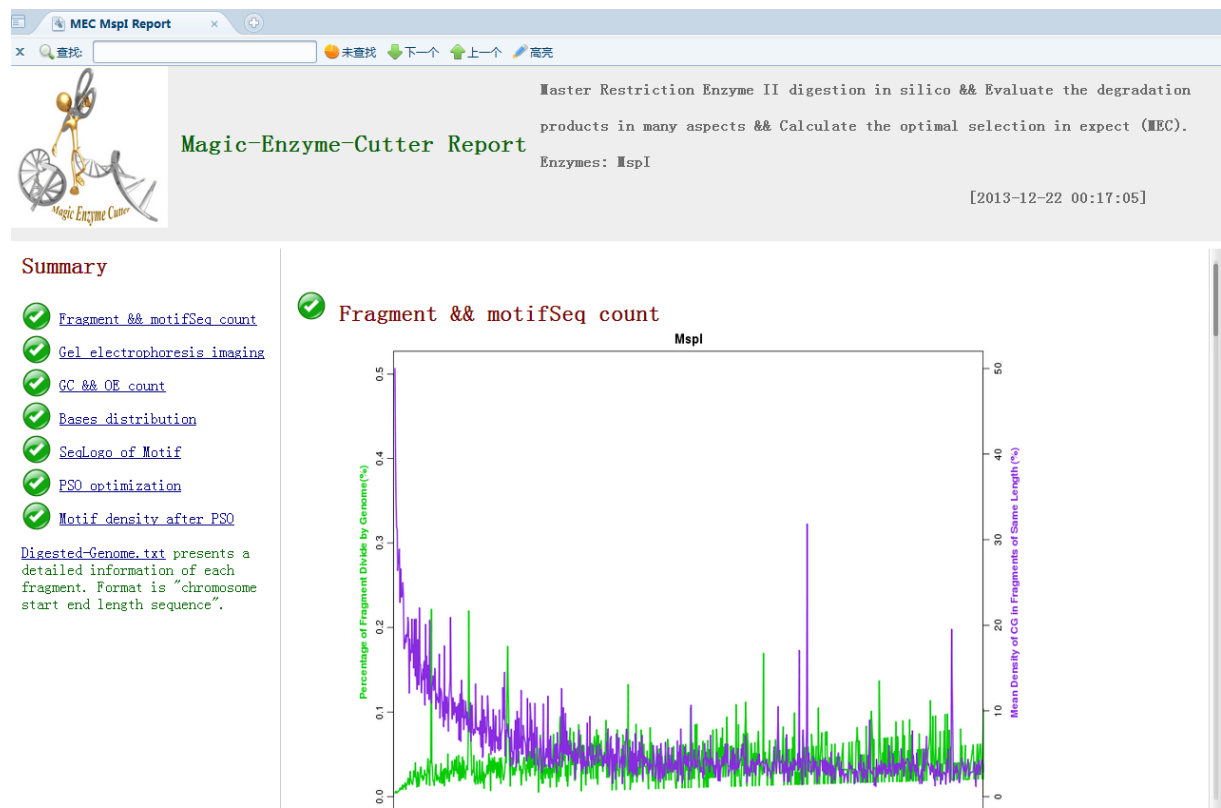
```
>sh test.sh /path/to/R
```

测试结果在文件夹：MEC-xxx/test/test-result

测试结果文件：

 Digest	2013/12/22 0:42	文件夹	
 Figure	2013/12/22 0:42	文件夹	
 Icons	2013/12/22 0:42	文件夹	
 Tables	2013/12/22 0:42	文件夹	
 Report-MEC	2013/12/22 0:41	HTML 文件	6 KB

双击网页 Report-MEC，显示为：



测序运行的参数为：

```
Magic-Enzyme-Cutter -i ref.fa.gz -o $pwd_dir/test-result -R $R
```

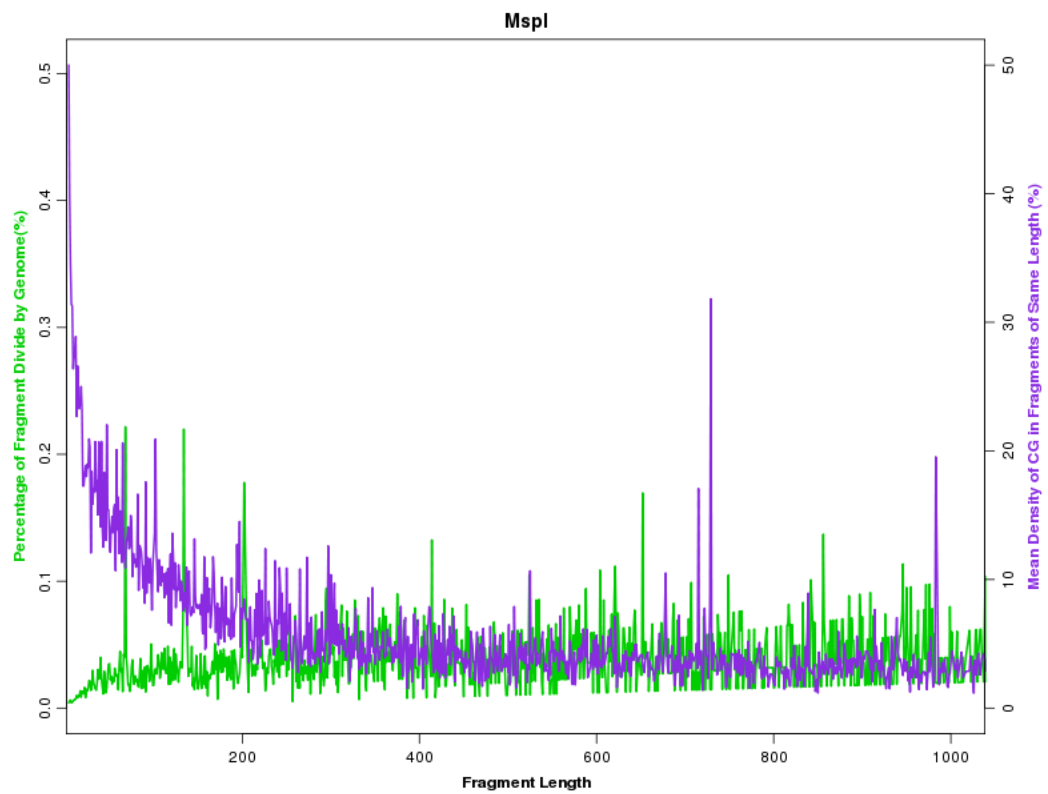
参数	参数值（默认值）	解释
-i	ref.fa.gz	Fasta 格式的参考基因组
-o	./test/test-result	输出结果文件夹
-m	CG	感兴趣的 motif
-R	Path/R	R 路径
-s	40	选取酶切片段最小长度
-e	1000	选取酶切片段最大长度
-n	200	PSO 最优化的 motif 密度的最小片段范围
-x	500	PSO 最优化的 motif 密度的最大片段范围
-f	4	Motif 上下游碱基数目
-g	100	碱基分布密度值
-c	F	是否改变图像视觉效果（F 表示否，T 表示是）
-t	50	凝胶电泳图透明度

运行时间为：

```
real    0m13.394s
user    0m10.265s
sys     0m0.898s
```

图形展示：

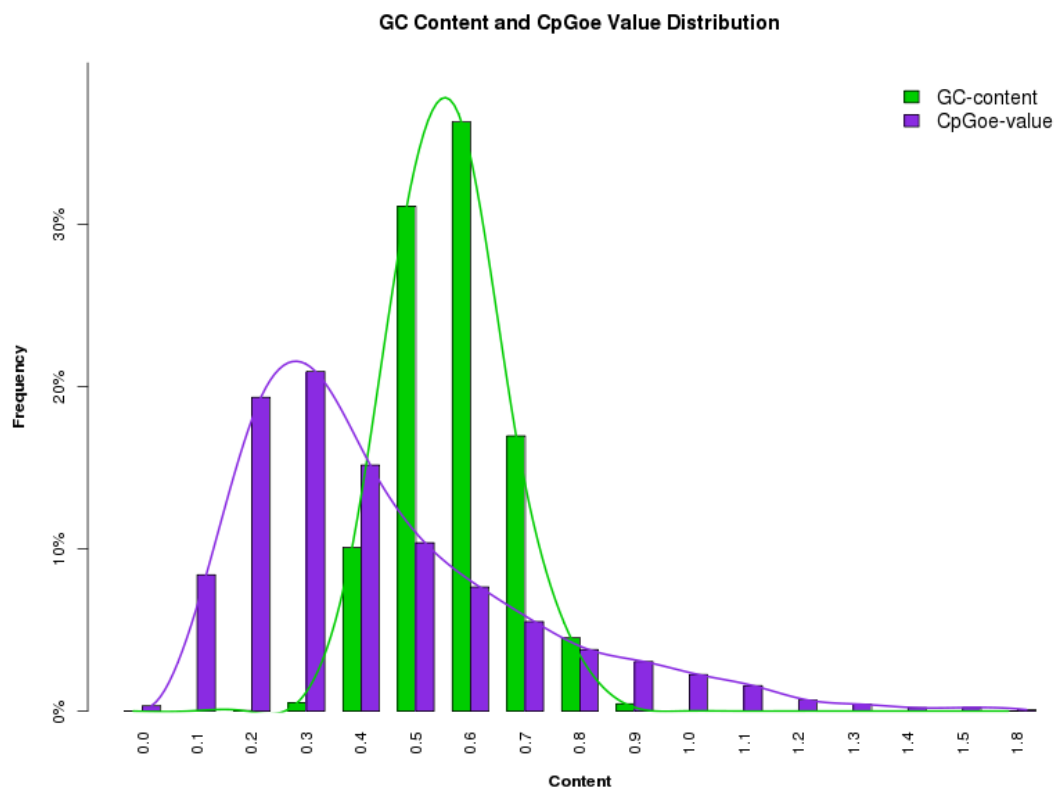
1、模拟酶切片段及 motif 分布图



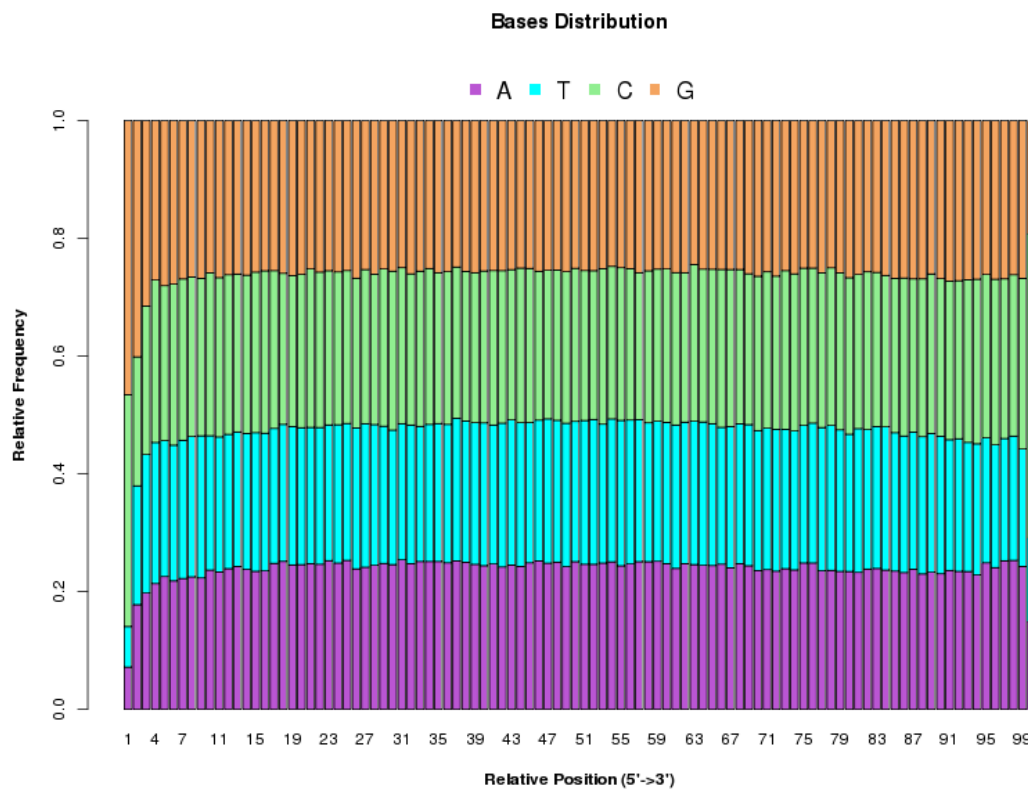
2、模拟酶切胶图。



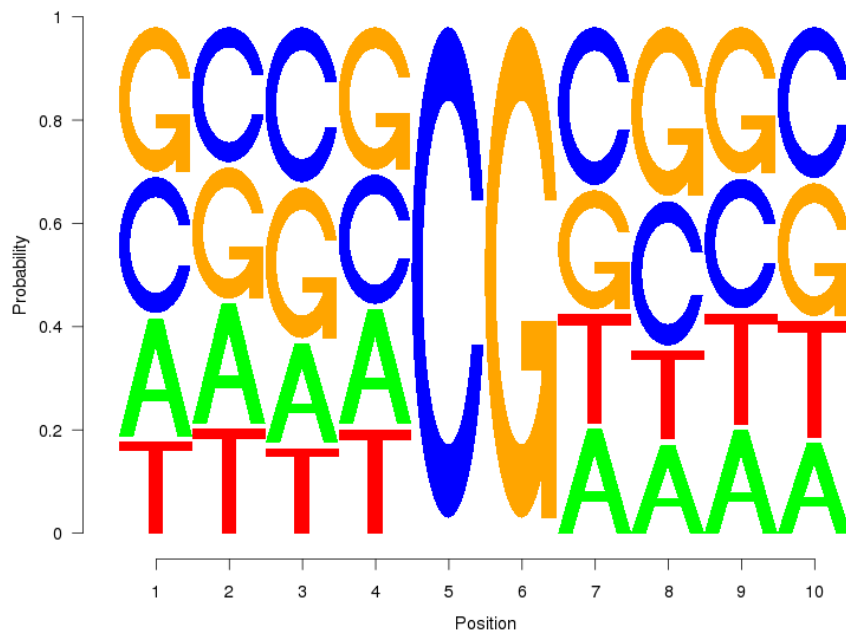
3、GC & OE 分布图:



4、碱基分布图:



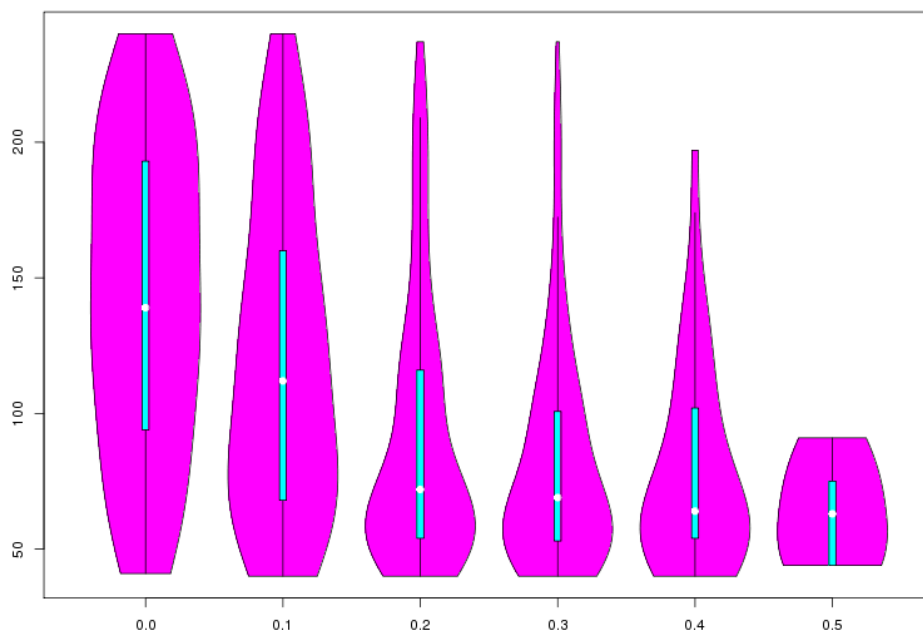
5、目标 motif 及两侧碱基的 seqLogo 图



6、PSO 最优化得到的起始长度

Start	End
40bp	240bp

7、PSO 最优化得到的 motif 密度分布图



8、酶切文件显示

Digested-Genome.txt presents a detailed information of each fragment. Format is "chromosome start end length sequence".

格式：染色体 起始位置 终止位置 片段长度 碱基序列

chrR 1 212 212

```
GATCTGATAAGTCCCAGGACTTCAGAAGAGCTGTGAGACCTTGGCCAAGTCACTTCCTCCTTCAGGAA
CATTGCAGTGGGCCTAAGTGCCTCCTCTCGGGACTGGTATGGGGACGGTCATGCAATCTGGACAACATTAC
CTTTAAAAGTTTATTGATCTTTTGTGACATGCACGTGGGTTCAGTAGCAAGAACTAAAGGGTCGCAGGC
```

chrR 213 1357 1145

```
CGGTTTCTGCTAATTTCTTAATTCCAAGACAGTCTCAAATATTTTCTTATTAACCTCCTGGAGGGAGGC
TTATCATTCTCTCTTTTGGATGATTCTAAGTACCAGCTAAAATACAGCTATCATTCTTTTCTTGATTGGGAG
CCTAATTTCTTTAATTTAGTATGCAAGAAAACCAATTTGGAAATATCAACTGTTTTGGAAACCTTAGACCTAGG
TCATCCTTAGTAAGATCTCCCATTTATATAAATACTTGCAAGTAGTAGGCCATAATTACCAAACATAAAGCCA
ACTGAGATGCCCAAAGGGGGCCACTCTCCTTGCTTTTCTCCTTTTTAGAGGATTTATTTCCCATTTTTCTTAA
AAAGGAAGAACAACTGTGCCCTAGGGTTTACTGTGTCAGAACAGAGTGTGCCGATTGTGGTCAGGACTCC
ATAGCATTTCACCATGAGTTATTTCCGCCCTTACGTGTCTCTCTCAGCGGTCTATTATCTCCAAGAGGGC
ATAAAACTGAGTAAACAGCTCTTTATATGTGTTTCTGGATGAGCCTTCTTTAATTAATTTGTTAAGGG
ATTTCTCTAGGGCCACTGCACGTCTGAGGGAGTCACCCCGACACTCCCAATTGGCCCCTTGTCACCCAG
GGGCACATTTAGCTATTTGTAAACCTGAAATCACTAGAAAAGGAATGTCTAGTGACTTGTGGGGGCCAAG
GCCCTTGTATGGGGATGAAGGCTCTTAGGTGGTAGCCCTCCAAGAGAATAGATGGTGAATGTCTCTTTTCA
GACATTAAAGGTGTCAGACTCTCAGTTAATCTCTCCTAGATCCAGGAAAGGCCTAGAAAAGGAAGGCCTGA
CTGCATTAATGGAGATTCTCTCCATGTGCAAAATTTCTCCACAAAAGAAATCCTTGCAGGGCCATTTAATG
TGTTGGCCCTGTGACAGCCATTTCAAATATGTCAAAAAATATATTTTGGAGTAAAATACTTTCAATTTCTTTC
AGAGTCTGCTGTCTATGATGCCATACCAGAGTCAGGTTGGAAAGTAAGCCACATTATACAGCGTTAACCTA
AAAAAACAAAAAACTGTCTAACAAGATTTTATGGTTTATAGAGCATGATTCCC
```

输入文件为基因组文件，格式为 **FASTA**，其格式如下：

>chrR

```
GATCTGATAAGTCCCAGGACTTCAGAAGagctgtgagaccttggccaagt
cacttcctcctcagGAACATTGCAGTGGGCCTAAGTGCCTCCTCTCGGG
ACTGGTATGGGGACGGTCATGCAATCTGGACAACATTACCTTTAAAAGT
TTATTGATCTTTTGTGACATGCACGTGGGTTCAGTAGCAAGAACTAA
AGGGTCGCAGGCCGGTTTCTGCTAATTTCTTTAATTCCAAGACAGTCTCA
AATATTTTCTTATTAACCTCCTGGAGGGAGGCTTATCATTCTCTCTTTG
GATGATTCTAAGTACCAGCTAAAATACAGCTATCATTCTTTTCTTGAT
TTGGGAGCCTAATTTCTTTAATTTAGTATGCAAGAAAACCAATTTGAAA
TATCAACTGTTTTGGAAACCTTAGACCTAGGTCATCCTTAGTAAGATctt
cccttatataaatacttgaagtagtagtgcataattaccaaacata
aagccaactgagatgcccgaaggggccactctccttcttctcctt
```


2、详细教程（英文）

Magic-Enzyme-Cutter

Master Restriction Enzyme II digestion in silico && Evaluate the degradation products in many aspects && C

Author : Mao FengBiao,Cai WanShi,Wen YanLing.

Contact: maofengbiao@gmail.com.

Usage:

Magic-Enzyme-Cutter -i <ref.fa> -o <outdir> [options]

Require:

-i <in_file: reference in FASTA format>

-o <out_directory>

Options:

-R <R path> [R]

-z <enzyme name or cut-Type> [MspI]

-m <motif which is focus on> [CG]

-s <start length of fragment> [40]

-e <end length of fragment> [1000]

-n <min regions of PSO search> [200]

-x <max regions of PSO search> [500]

-f <bases of flanking included> [4]

-g <graduated scale number> [100]

-c <change the vision,T or F> [F]

-t <transparency of Gel Figure> [50]

-h <display this help information>

Note:

The delimiter of enzymes is ",", "_" or "+" while the delimiter of cut-Type is "-".

The brief-codes of cut-Type are supported, They are:

R=G/A;Y=C/T;M=A/C;K=G/T;S=G/C;W=A/T;B=C/G/T;D=A/G/T;H

Gzipped FASTA format input is supported.

Example:

Magic-Enzyme-Cutter -i <ref.fa> -o <outdir> -z MspI -s 100 -e 2000

Magic-Enzyme-Cutter -i <ref.fa> -o <outdir> -R /Path/R -z C-CGG,C-GGC,G-CGC

Magic-Enzyme-Cutter -i <ref.fa> -o <outdir> -R /Path/R -z MspI_AcI+TaqI

Magic-Enzyme-Cutter -i <ref.fa> -o <outdir> -R /Path/R -z MspI+TTT-AAA+YAC-GTR

Magic-Enzyme-Cutter -i <ref.fa> -o <outdir> -n 100 -x 300 -g 50 -f 5 -c T

Detailed scripts:

cd ./bin/

1. perl Magic-enzyme-cutter.pl

perl Magic-enzyme-cutter.pl <ref.fa> <type1>,<type2>... <brief code> [outfile]

example:

```
perl Magic-enzyme-cutter.pl hg19.fa C-CGG,C-GGC,G-CGC brief-code.txt >
outfile.txt
```

```
perl Magic-enzyme-cutter.pl hg19.fa MspI_AciI+TaqI brief-code.txt > outfile.txt
```

```
perl Magic-enzyme-cutter.pl hg19.fa MspI+TTT-AAA+YAC-GTR brief-code.txt >
outfile.txt
```

2. perl Fragment-motifSeq-count.pl

Usage :

perl Fragment-motifSeq-count.pl

```
-i <enzymed_file>
-m <motif> [CG]
-n <name of enzymes> [MspI]
-s <start length> [40]
-e <end length> [1000]
-o <out_directory>
-R <R path> [R]
-h <display this help info>
```

3. perl Gel-Ele.pl

Usage :

perl Gel-Ele.pl

```
-i <in_file,outfile of Fragment-motifSeq-count.pl>
-o <out_directory>
-s <start length of fragment> [40]
-e <end length of fragment> [1000]
-n <names of enzyme> [MspI]
-t <transparency,0~100> [50]
-R <R path> [R]
-h <display this help info>
```

4. perl GC-OE-count.pl

Usage :

perl GC-OE-count.pl

-i <enzymed_file>

-s <start length> [40]

-e <end length> [1000]

-o <out_directory>

-R <R path> [R]

-h <display this help info>

5. perl Base-distribution.pl

Usage :

perl Base-distribution.pl

-i <enzymed_file>

-f <Relative fragments> [100]

-s <start length> [40]

-e <end length> [1000]

-R <R path> [R]

-o <out_directory>

-h <display this help info>

6. perl Motif-seqLogo.pl

Usage :

perl Motif-seqLogo.pl

-i <enzymed_file>

-f <flanking length> [4]

-m <motif> [CG]

-c <ic: "T" or "F"> [F]

-s <start length> [40]

-e <end length> [1000]

-o <out_directory>

-R <R path> [R]

-h <display this help info>

7. perl PSO.pl

Usage :

perl PSO.pl

-i <in_file,outfile of Fragment-motifSeq-count.pl>

-o <out_directory>

-s <start length of fragment> [40]

-e <end length of fragment> [1000]

-min <min distance between start and end> [200]

-max <min distance between start and end> [500]
-R <R path> [R]
-h <display this help info>

8. perl PSO-Motif-dense.pl

Usage :

perl PSO-Motif-dense.pl

-i <enzymed_file>
-m <motif> [CG]
-s <start length> [40]
-e <end length> [1000]
-c <horizon or not,T:true,F:False> [F]
-o <out_directory>
-R <R path> [R]
-h <display this help info>

9. perl Report.pl

Usage :

perl Report.pl

-i <Figure directory by Magic Enzyme Cutter>
-t <The table of PSO results>
-l <Logo of Icons>
-m <Logo of MEC>
-n <names of enzymes> [MspI]
-s <start length> [40]
-e <end length> [1000]
-o <out_directory>
-h <display this help info>

