Workflow for producing a new chromosome

Experimental implementation

Computer-aided design

**Way1**

Neochr:

构建一条新染色体，其中包含了Grap Gene, Expression Unit， Neochromosome, Add, Delete

Nucleotide modification:

对新染色体进行碱基水平的修改优化，插件按优先级从大到小分别是CRISPRi site, Restriction site Remove, Codon Optimazation & Swap, Repeat Smash.

Segmentation:

从合成和组装策略的角度对即将合成的染色体进行切分. 并且找酶切位点（restriction site creation）含有针对合成仪合成和芯片合成两大类的插件。

**Way2**

Neochr:

不变

Primer Design:

致力于将整个基因单元从自然界中完整取出。并且添加符合3A 组装策略的接头

安装说明：

1. 安装blast+，安装cytoscape？
2. 安装python， biopython中的NLTK包
3. 安装我们添加了bin的jbrowse（我们将bin放进特定文件夹就好，jbrowse会自动配置？），数据储存在本地。

程序目录结构：

/\*/bin 放插件的地方

/\*/config/enzymes 可用酶的列表，如standard\_and\_IIB

/\*/config/features 放染色体特征的地方，有ARS，左端粒，右端粒，中心粒，loxpsym

/\*/config/markers 放染色体验证时所用marker的地方，有LEU2，URA3，HIS3，TRP1

/\*/config/species 放物种的gff和fa，还有相关辅助信息，如生长相关基因，必需基因的地方

/\*/pathway/kgml 放kgml，和解析kgml的程序的地方

/\*/pathway/diagram 放kgml画出的图的地方，除了在jbrowse上有pathywa信息外，在这里也放上jpg或者svg的图

/\*/rewire/logic\_gate 放逻辑门的说明文档和fa和gff

/\*/rewire/repressilator 放振荡电路

/\*/rewire/bistable 放双稳态

/\*/class/lib

/\*/docs 说明文档

/\*/README

/\*/LICENSE

插件的结果评估：

抑或和biostudio等的比对？

插件详解：

**GrapGene**

从KEGG调出模式物种的pathway信息，包含pathway的基因和基因间关系。

操作说明：点GrapGene，弹出可选模式物种，弹出可选pathway，弹出该pathway含有的基因，选择喜欢的基因，拖动其到已选框中（同时提供添加下一组pathway的选项（添加一个框即可），也提供手动添加基因的选项（添加一个框，输入基因名和fasta序列，对于手动添加的基因和库进行blast，获取该基因在该物种里的homolog，进行归类。）也提供基本的逻辑门电路的基因，调控序列以供使用（如与门，或门，转换开关，riboswitch，双稳态开关，甚至振荡器））。

后台操作：已经赋予了gene各种有用的信息（如基因的必须性和基因与生长的关系）以及序列。展示原有pathway的关系图，同时用cytoscape利用kgml的信息生成新基因列表的关系图指导基因的选择。并展示加入的逻辑门等基因线路。

**Expression Unit（可选）**

针对噬菌体进行转录单元的分离，以及针对一个基因出现另一个基因的5‘调控序列和部分CDS的情况，将ATG或者GTG同义变换。

**Neochromosome**

操作说明：获得了基因列表后，可根据Grapgene那生成的性关系图来针对基因的顺序和方向进行手动修改，或者点“推荐顺序，推荐方向”。然后生成染色体的基本骨架。

**Add&Delete**

Add可以加loxpsym，ARS，着丝粒，端粒，和基因间区。

Delete可以去除掉intron等。

**CRISPR site**

设计合成的染色体与野生的染色体的某23bp不同来进行区分染色体的属性（也就是marker，使用U6启动子的话是GN20GG，后20个N是unique），同时这个marker也是CRISPR识别位点，可以进行野生型的表达沉默。

此处输出每一个CRISPR位点的unique的情况，以及预估的成功率。

**Restriction site remove**

去除切分时所使用的酶切位点

**Codon Optimazation & Swap**

通过改变密码子使用率来调节表达量的大小，此处输出表达量的变化率。

**Repeat Smash**

降低GC含量或者碱基连续率来降低合成难度和PCR难度。此处输出合成难度变化率。

Restriction Info Online

从NEB上面

**Segmentation(30kto10k)**

**范例：**

参数：

gff: 进行酶切位点分析的染色体的注释文件 (string)

fa:进行酶切位点分析的染色体(string)

overlap30k: 30kchunk的大致长度（default：1k）

m1: marker for selection alternately (default:LEU2 1797)

m2: marker for selection alternately (default:URA3 1112)

m3: marker orinally residing in first 30k segmentaion (default:HIS3 657+x)

m4: marker orinally residing in first 30k segmentaion (default:TRP1 675+x)

获得cen和ars所在区域（cen和ars设计时距离不超过30k），截取30k左右起始片段，，左右加上不同的原始抗性，如HIS3和TRP1。左右再加上telo。然后分向左和向右两个方向，进行同源替换。

向左的与起始片段有875bp的overlap30k，同时左边加上抗性，再加上telo，向右的类似。

这里要设计30k的切分。30k的边界不需要在基因间区，只需要cen和ars的位置。

**Restriction Eyzyme Recognition Sites Parser**

分析现有和潜在的酶切位点（后者对于组装策略无用，优先级降低）

对基因间区分析现有位点，这些位点是不能改变的。

造后缀树，内容是所有酶切位点的6个翻译框（正3反3），使得每一个点都是可以反翻译成某个酶切位点的氨基酸序列。然后在外显子区域上寻找现有和潜在的位点。

对酶切位点评分，分数如何定义。

范例：

perl globalREmarkup\_v2.pl -gff sce\_chrI.gff -fa chr01.fa -re standard\_and\_IIB -ct Standard.ct -ot chr01.globalREmarkup\_stand\_IIB.out

参数：

#gff: 进行酶切位点分析的染色体的注释文件 (string)

fa:进行酶切位点分析的染色体(string)

re:酶切位点的库(string) (可选：standard\_and\_IIB, standard, inex\_and\_IIB, inex)

# inex==inexpensive

ct:染色体所在物种的密码子表(string) the genetic codes（可选：1 standard等）

|  |
| --- |
| 1 The Standard Code  2 The Vertebrate Mitochondrial Code  3 The Yeast Mitochondrial Code  4 The Mold, Protozoan, and Coelenterate Mitochondrial Code and the Mycoplasma/Spiroplasma Code  5The Invertebrate Mitochondrial Code  6 The Ciliate, Dasycladacean and Hexamita Nuclear Code  7 The Echinoderm and Flatworm Mitochondrial Code  8 The Euplotid Nuclear Code  9 The Bacterial, Archaeal and Plant Plastid Code  10 The Alternative Yeast Nuclear Code  11 The Ascidian Mitochondrial Code  12 The Alternative Flatworm Mitochondrial Code  13 Blepharisma Nuclear Code  14 Chlorophycean Mitochondrial Code  15 Trematode Mitochondrial Code  16 Scenedesmus Obliquus Mitochondrial Code  17 Thraustochytrium Mitochondrial Code  18 Pterobranchia Mitochondrial Code  19 Candidate Division SR1 and Gracilibacteria Code |

ot:output file of enzyme map (string)

**Segmentation(30kto10kto2k)**

例子：

perl ReClass2.pl -gff ../sce\_chrI.gff -mk global\_v2\_potent\_exit/globalREmarkup\_stand\_IIB.out -re standard\_and\_IIB -fa chr01.fa -ot chr01.seg.out

参数：

#gff: 进行酶切位点分析的染色体的注释文件 (string)

fa:进行酶切位点分析的染色体(string)

re:酶切位点的库(string) (可选：standard\_and\_IIB, standard, inex\_and\_IIB, inex)

# inex==inexpensive

ot: chr01.seg.out

以下为可选参数：

asem2-10（2k到10k的组装策略） 默认是gibson，可选gibson和goldengate

asem10-30（10k到30k的组装策略） 默认是goldengate，可选gibson和goldengate

（不可选--30k逐步替换YAC从而成为一条新的染色体）

miniChunk的大小（默认是2k）

minichunk上下浮动（默认是200）

chunk的大小（默认是10k）

chunk上下浮动（默认是1k）

保存chunk的质粒的两段同源区域（默认都是40）

megachunk的大小（默认是30k）

megachunk上下浮动（默认是3k）

marker的种类（默认是LEU和URA）

替换的YAC的种类

若选择gibson，则还需要考虑：

Overlap的长度（默认是40）没有上下浮动

Overlap的最小tm（默认是56）

Overlap的最大tm（默认是60

Overlap的最低自由能（默认是-3）

目标片段从质粒上切下来的酶的类型（默认IIP）

使用的外切酶的消化方向（默认是从5‘端开始，即使用T5外切酶）

这些酶的作用温度（默认是37度）

这些酶的最高单价（默认是0.5$/Unit）

（若add插件加入了loxpsym位点：

则还需要考虑overlap离loxpsym位点的最小距离（默认是40）

）

若选择goldengate，则需考虑

目标片段从质粒上切下来的酶的类型（默认IIA和IIB）

这些酶的作用温度（默认是37度）

**Segmentation-OLS**

。。。

**Primer Design**

自动设计好3A组装策略或者SOE-PCR的接口，并能支持点突变等。

实验指导：

首先将一个chunk的2k片段分别从载体上单酶切切下来，