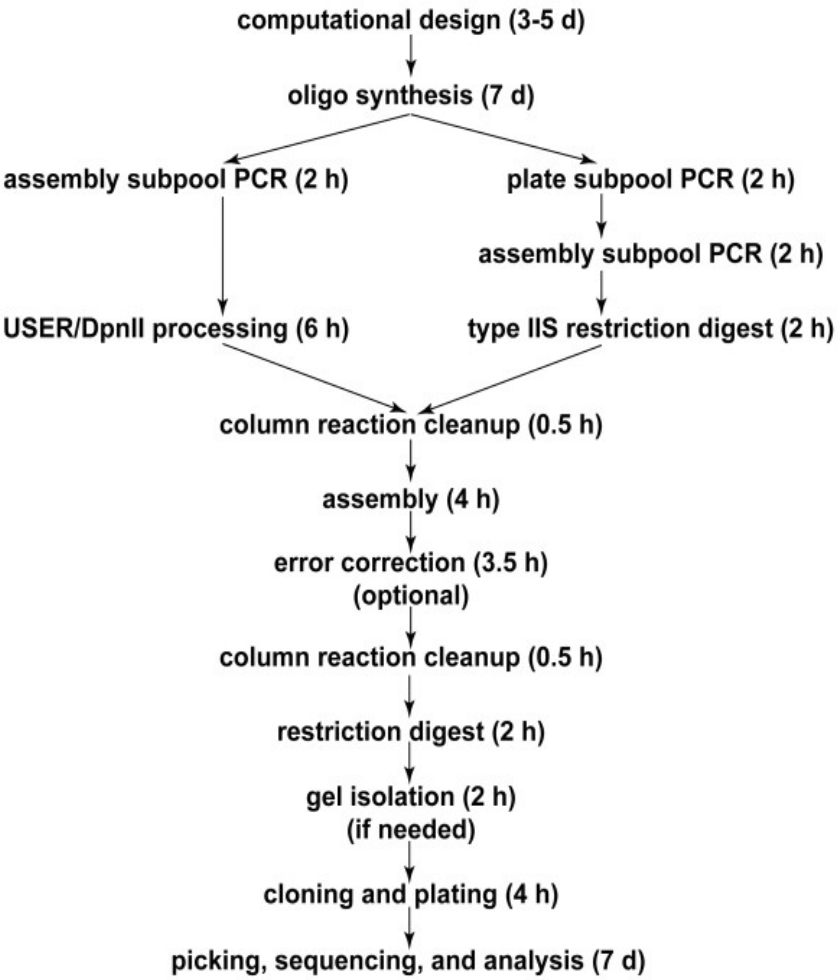
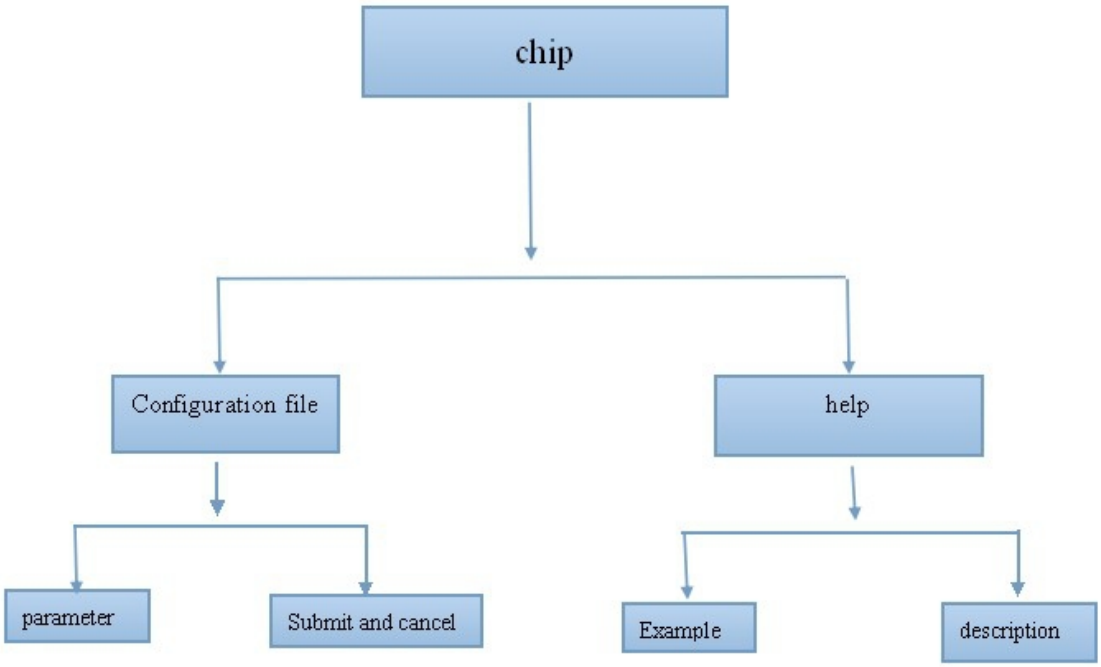
Workflow for gene synthesis from high-fidelity DNA microchips



UI design


chip
Configuration file
help
parameter
Submit and cancel
Example
description

后台需要安装的程序有：(都已经下载好了，只要装就行了)

Dependencies

UnaFold and BioPython are required to run the Python scripts in this package.

Please read followings for installation before run GSAP !!!

Prerequisites: *UNAFold*and *Biopython.*

1、UnaFold

http://dinamelt.rit.albany.edu/download.php

2、Biopython：

http://biopython.org/DIST/docs/install/Installation.html

Parameter：

Configuration file :

-a : fixedIndex
-b fixedPrimerSetThroughoutFile //true or flase
-c positionleeway
-d fixedPlateNumber
-e avgoverlapsize
-f lengthleeway
-g insertionSizeToKillRESite
-z forwardPrimersLibraryFile //输入文件必须要有的
-i initialPlateNum
-j initialPlateNum
-l buildSequencesFile // 必须要有的 这个文件需要的是 fasta 格式的
-m buildSequencesFile // 必须要有的 这个文件需要的是 fasta 格式的
-n selfDimersThreshold
-u selfDimersThreshold
-o the name of user .//这个是用来标示不同的 configuration file 和 gasp-para file（当用户多的时候可能会出现重名的现象）

Email_para:

-p SMTPserver: "your.smtp.server.locations"
-q emailAddressFrom: "blah@blah.com"
-t SMTPlogin :"yourLoginNameForYourSMTPServer"
-r "SMTPpw":"yourPasswordForYourSMTPServer"

-l 和-m 是需要用户提供的输入文件，需要交这个两个输入文件放入程序目录下的 input 文件夹中

The script of creating configuration file:
这个脚本的功能根据用户的需求填写相应的参数：用户填写后，提交保存在程序的目录下
Usage : perl Get_configFile.pl <option> <output file name>

用户提交了 configuration file 后，再按下确定就可以了。
后台的操作是，调用运行程序：python python gasp.py <configuration file name> <Email file name>
为了方便，最终的结果是需要在前端展示的，而是直接以邮件的形式发送到用户指定的邮箱。

Help:

这个页面主要是对软件的一些 description 和 参数说明

The description of software:

GASP:Gene Assembly by Subpool PCR This set of scripts designs oligonucleotides that can be used to synthesize genes from high-complexity DNA pools

Parameter description

The parameters, which are described in detail below, may have to be further adjusted if the DNA will be processed using methods that deviate from the workflow described here.

InitialPlaneNum: 96-well plates of assemblies will be numbered sequentially initiating at this value. This should never be set to 1, as plate #1 is reserved for construction primers.

avgoverlapsize: Each construct will be broken up into assembly oligos that will be fused using a polymerase. The fusion reaction requires priming through overlaps between neighboring oligos. This setting specifies the mean length of the overlap region.

deltaGThresholdForOverlaps: Rejects any overlaps with a secondary structure that has a hybridization free energy less than the value specified (in units of kcal/mol).

selfDimerThreshold: Rejects assembly oligos that have any self-dimerization configura-tions with a hybridization free energy less than the value specified (arbitrary units).

lengthleeway: Sets allowable variation in the length of the overlap regions.

positionleeway: Sets allowable variation in the assembly oligo junction position. Increas-ing this value results in a less constrained search space, but increases the computation time and increases variation in synthesized oligonucleotides' lengths.

oligoSizeMax: The maximum oligo size that will be designed. This includes the full-length oligos that include the coding region, the restriction enzyme processing site, and the assembly-specific and plate-specific priming sites. This value should typically be constrained by the commercial synthesis platform used. Note that many of the oligos will be shorter than this maximum value.

seqsToAvoidInOverlapRegion: Specifies positions to be avoided in the overlap between neighboring assembly oligos. This should usually be left blank, but can be used in specialized applications, such as constructing proteins with known repeated regions.

Example :

Configuration :

```
{"configDictList":[ // each entry in this list corresponds to a single input file with many seqs
            {
    "initialPlateNum": 3, // never set this = to 1
    "buildSequencesFile": "input-seqs/All_MC_Scaffolding_Seq_PostProc.txt", // input file name
     "RESpacing": [ // offset of cut site from recognition site for type II enzyme
                              2,
                              5,
                              4
                         ],
       "REVector": [
                              "BtsI",
                              "BsmBI",
                              "BspQI"
                         ],
         "SearchForRE": "True",
      "REToUse": "BtsI", // fixed RE to use, if applicabl
        "forwardPrimersLibraryFile": "primer-library/forward_finalprimers.fasta", // from Sri
        "reversePrimersLibraryFile": "primer-library/reverse_finalprimers.fasta", // from Sri
          "fixedPrimerSetThroughoutFile": "False", // true if all seqs in file get same primer set
        "fixedIndex": 0, // plate position index of primer set when using fixed primer set
          "fixedPlateNumber": 1, // plate # of primer set when using fixed primer set
          "avgoverlapsize": 20, // keep this fixed
          "deltaGThresholdForOverlaps": -3, // keep this fixed
          "selfDimersThreshold": 5, // default = 3, deltaG threshold for “primer dimer” effect
            "insertionSizeToKillRESite": 2, // keep this fixed
             "lengthleeway": 15, // default = 10, length of oligo can be + or - this #
             "overlaptemps": [ // keep these fixed, range of overlap melting temps for PCR
                              55,
                              65
                         ],
           "positionleeway": 10, // oligo center position can be + or - this
            "seqsToAvoidInOverlapRegions":["CATCCGTCGAAGACGGATG"] // set to [] if can
             “skip” : ["ATbias_Seq1"] // names of sequences not to process in the file
                }
]
}
```

Email_para:

{

"SMTPserver":"smtp.gmail.com",

"emailAddressFrom":"adam.h.marblestone@gmail.com",

"SMTPlogin":"gaspserver@gmail.com",

"SMTPpw":"syngaspbiogaspsisgaspgeorgechurch"

}

Result:

The first one will contain a report that contains: (1) The sequences to be synthesized on the DNA chip in FASTA format; (2) The plate-specific, position-specific, and construction primers needed to build the set of assemblies; (3) The plate-specific, position-specific, and construction primers that correlate with each individual assembly. The second e-mail will contain a FASTA file that contains the sequences that should be synthesized on the DNA chip.