ypkpathway

Ypkpathway is a software tool for the automated assembly of pathways using the Yeast Pathway Kit.

Installation

Python 2.7 is required to run ypkpathway. Python is available from www.python.org. Better yet is to get the anaconda scientific python distribution from https://store.continuum.io/cshop/anaconda/.

Further requirements are the python packages networkx, biopython, pydna, and docutils.

Once python (or anaconda) is installed, the simplest way of installing is to use pip. Pip is installed by default in anaconda. Get pip here: https://pip.pypa.io/en/latest/installing.html.

When pip is installed, type the following into a terminal window:

```
pip install ypkpathway
```

followed by <return>. This will download and install all dependencies automatically.

Use

Ypkpathway is meant to be used in the terminal. The syntax is very simple as ypkpathway takes only one argument:

```
C:>ypkpathway datafile.txt
```

The datafile.txt can have any name as long as it is a text file containing the sequences to be assembled. See nest section for syntax.

Indata

The datafile.txt file (which can have a different name) could have the structure depicted in Fig 1.

```
>tp1
atcgac...

>gene1
agcatc...

>tp2
atttct...

>tp2
tttgag...

>gene2
ccaatg...

>tp3
caacta...
```

Fig 1

The file in Fig1 is simply a list of the Tps and genes that should be assembled. The sequences in Fig 1 are truncated for clarity and could also be given in Genbank format or a mix of FASTA and Genbank formats.

The sequences could be linear fragments (as in the example four_gene_xylose_pathway1.txt file accompanying this document).

The yeast pathway kit was designed for the reuse some of the genetic parts, especially terminator-promoter plasmids, pYPKa_Z_XXX and pYPKa_E_XX. Once cloned, Terminator-promoter plasmids can be reused for other pathways, in which case they do not need to be constructed again.

```
>pYPKa_Z_tp1
acacatgt...

>pYPKa_A_gene1
accaacatg...

>pYPKa_E_tp2
acacttggtgt...

>pYPKa_Z_tp2
ccttggt...

>pYPKa_A_gene2
acctgt....

>pYPKa_E_tp3
acctgt....
```

Fig 2

We can supply sequences in the form of pYPKa_Z, pYPKa_A or pYPKa_E clones, typically from a previous assembly experiment. These will be recognized by the ypkpathway algorithm and the assembly report will indicate that these were given.

They could be or pYPK0_tp_gene_tp sequences. These sequences can be given to indicate that they already exist and are not to be assembled. In the same way we can supply pYPK0_tp_gene_tp sequences (Fig3), in which the pYPKa vectors needed for their assembly will not needed.

```
>pYPK0_tp1_gene1_tp2
actcagctatcttctgtattcgac...

>pYPK0_tp2_gene2_tp3
taactactactgactggtacttgg...
```

Fig 3

Any valid combination of the three kinds of sequences is also permitted (Fig4).

```
>pYPKa_Z_tp1
acacatgt...

>pYPKa_A_gene1
accaacatg...

>tp2
tttgag...

>pYPK0_tp2_gene2_tp3
taactactactgactggtacttgg...
```

Fig 4

In Fig4 we supply two pYPKa sequences, one for the tp1 and one for the gene1. The tp2 was never cloned before, so it is supplied as a linear sequence.

The pYPK0_tp2_gene2_tp3 vector was made in a previous experiment and is also given.

Output

Use the test data file “four_gene_xylose_pathway1.txt” in this manner:

```
C:>ypkpathway four_gene_xylose_pathway1.txt
```

followed by return. The ypkpathway program creates a folder in the current working directory (the directory from which ypkpathway was called). The folder is called “ypk_assembly”.

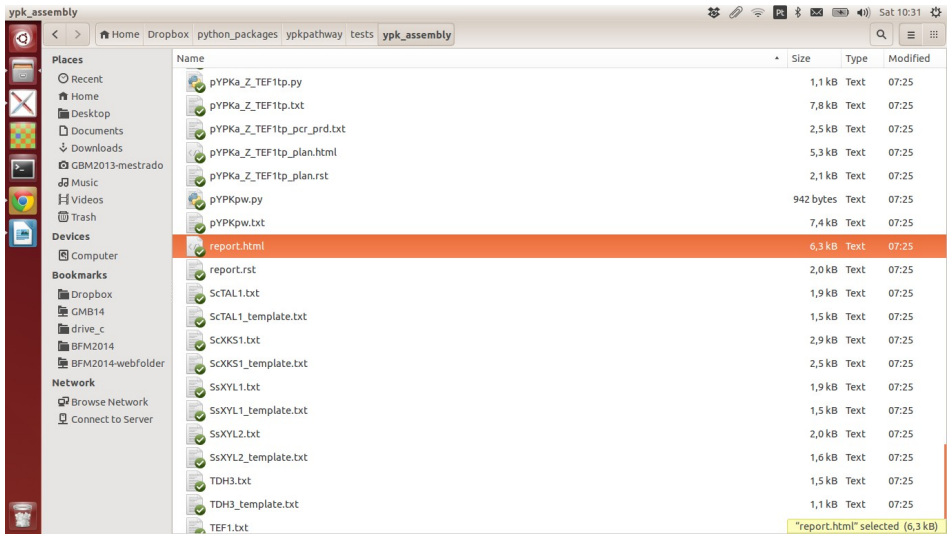


Fig 5

Open this folder and open the file “report.html” in your web browser (Fig5). You should now have a web page in your browser looking like Fig 6.

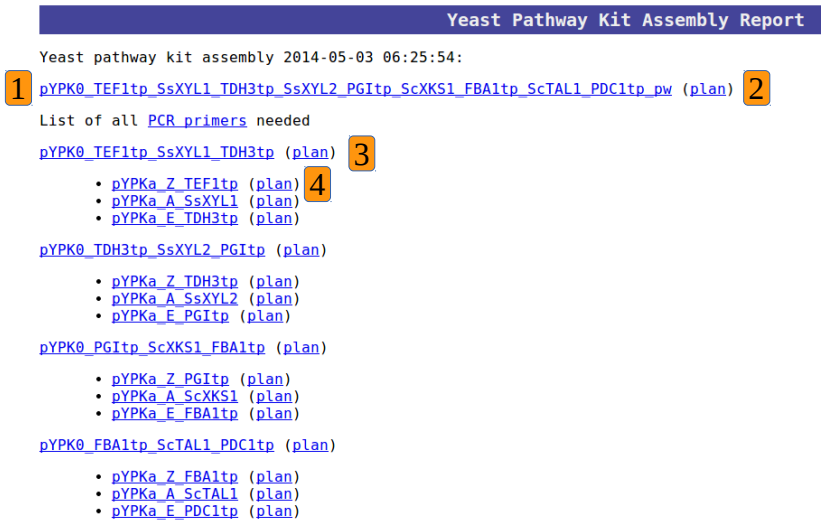


Fig 6:

Clicking on the first link “pYPK0_TEF1tp_SsXYL1_TDH3tp_SsXYL2_PGItp_ScXKS1_FBA1tp_ScTAL1_PDC1tp_pw” (Fig6-1) will display the final sequence of the pathway in the browser, a 14800 bp sequence in this case (Fig7).

```
LOCUS       pYPK0_pathway          14800 bp    DNA     circular UNK 14-MAY-2014
DEFINITION  pYPK0_pathway
ACCESSION   pYPK0_pathway
VERSION     pYPK0_pathway
KEYWORDS    .
SOURCE      .
  ORGANISM  .
            .
FEATURES             Location/Qualifiers
     overlap         363..486
                     /note="olp_GqXRnFdj0bUssSD9FyJPEMLnbh0"
                     /ApEinfo_fwdcolor="#D6B6CB"
                     /chksum="GqXRnFdj0bUssSD9FyJPEMLnbh0"
                     /ApEinfo_revcolor="#FEE9B6"
     primer_bind     537..555
                     /note="pfw579"
                     /ApEinfo_revcolor="red"
                     /ApEinfo_fwdcolor="green"
     primer_bind     complement(1090..1115)
                     /note="prv579"
                     /ApEinfo_revcolor="red"
                     /ApEinfo_fwdcolor="green"
     primer_bind     complement(1150..1172)
                     /note="567"
                     /ApEinfo_revcolor="red"
                     /ApEinfo_fwdcolor="green"
     primer_bind     1175..1194
                     /note="pfw957"
                     /ApEinfo_revcolor="red"
                     /ApEinfo_fwdcolor="green"
     primer_bind     complement(2111..2131)
                     /note="prv957"
                     /ApEinfo_revcolor="red"
                     /ApEinfo_fwdcolor="green"
     primer_bind     complement(2138..2162)
                     /note="467"
                     /ApEinfo_revcolor="red"
                     /ApEinfo_fwdcolor="green"
```

Fig 7

The “(plan)” link (Fig6-2) displays a small representation of how the final sequence was assembled (Fig8). This image shows how four PCR fragments were assembled from PCR products derived from pYPK0 tp_gene_tp clones and linearized pYPKpw sequence to form the final circular construct.

pYPK0_TEF1tp_SsXYL1_TDH3tp_SsXYL2_PGItp_ScXKS1_FBA1tp_ScTAL1_PDC1tp_pw

Step 1 Prepare vector

Linearize pYPKpw with EcoRV resulting in the linearized vector.

Step 2 tp-gene-tp PCR reactions

Perform the following 4 PCR reactions:

primers 577, 778 and pYPK0_TEF1tp_SsXYL1_TDH3tp => 2524bp_PCR_prod

primers 775, 778 and pYPK0_TDH3tp_SsXYL2_PGItp => 2924bp_PCR_prod

primers 775, 778 and pYPK0_PGItp_ScXKS1_FBA1tp => 3567bp_PCR_prod

primers 775, 578 and pYPK0_FBA1tp_ScTAL1_PDC1tp => 3009bp_PCR_prod

Step 3 Transformation and Assembly

Mix the DNA fragments and transform a S. cerevisiae ura3 mutant. The DNA fragments will be assembled by in-vivo homologous recombination:

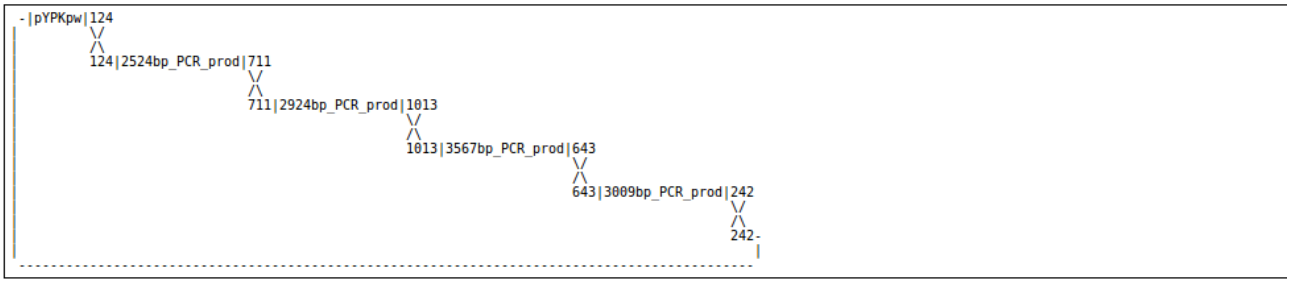
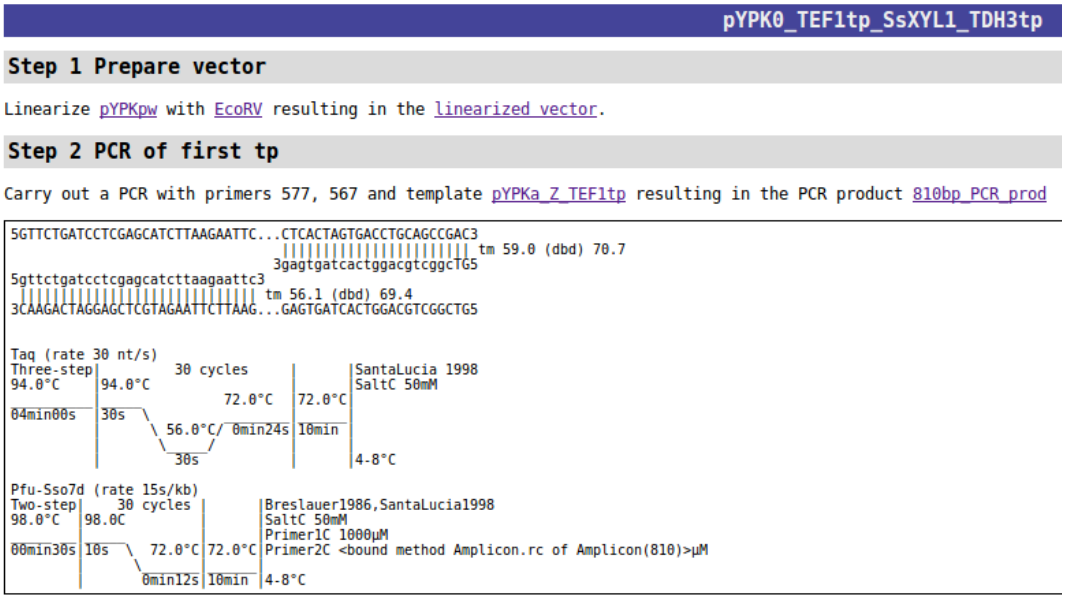


Fig 8

The second “(plan)” link (Fig6-3) show a plan for the construction of the first pYPK0_tp_gene_tp clone. These clones are assembled from three pYPKa derived pcr products for each element and and linearized pYPKpw (Fig9).



pYPK0_TEF1tp_SsXYL1_TDH3tp

Step 1 Prepare vector

Linearize pYPKpw with EcoRV resulting in the linearized vector.

Step 2 PCR of first tp

Carry out a PCR with primers 577, 567 and template pYPKa_Z_TEF1tp resulting in the PCR product 810bp_PCR_prod

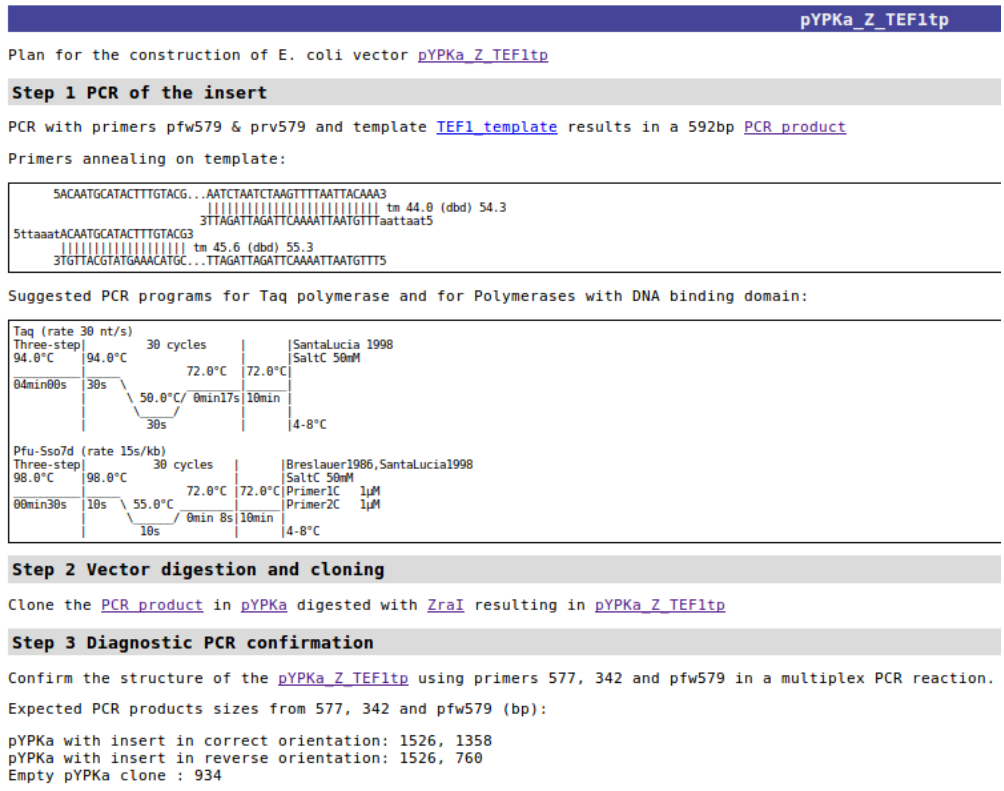
```
5GTTCTGATCCTCGAGCATCTTAAGAATTC...CTCACTAGTGACCTGCAGCCGAC3
                                 |||||||||||||||||||||||| tm 59.0 (dbd) 70.7
                               3gagtgatcactggacgtcggcTG5
5gttctgatcctcgagcatcttaagaattc3
 |||||||||||||||||||||||||||||| tm 56.1 (dbd) 69.4
3CAAGACTAGGAGCTCGTAGAATTCTTAAG...GAGTGATCACTGGACGTCGGCTG5


Taq (rate 30 nt/s)
Three-step|         30 cycles      |       |SantaLucia 1998
94.0°C    |94.0°C                  |       |SaltC 50mM
          |                 72.0°C |72.0°C |
04min00s  |30s  \                  |       |
          |      \ 56.0°C/ 0min24s |10min  |
          |       \____/           |       |
          |        30s             |       |4-8°C

Pfu-Sso7d (rate 15s/kb)
Two-step  |    30 cycles      |       |Breslauer1986,SantaLucia1998
98.0°C    |98.0C              |       |SaltC 50mM
          |                   |       |Primer1C 1000µM
00min30s  |10s  \  72.0°C     |72.0°C |Primer2C <bound method Amplicon.rc of Amplicon(810)>µM
          |      \____________|_______|
          |       0min12s     |10min  |4-8°C
```

Fig 9

The last “(plan)” link (Fig6-4) show a plan for the construction of the first pYPKa clone (Fig10). These clones are made from pYPKa vectors linearized with ZraI, AjiI or EcoRV and a linear PCR product.



pYPKa_Z_TEF1tp

Plan for the construction of E. coli vector pYPKa_Z_TEF1tp

Step 1 PCR of the insert

PCR with primers pfw579 & prv579 and template TEF1_template results in a 592bp PCR product

Primers annealing on template:

```
     5ACAATGCATACTTTGTACG...AATCTAATCTAAGTTTTAATTACAAA3
                           |||||||||||||||||||||| tm 44.0 (dbd) 54.3
                         3TTAGATTAGATTCAAAATTAATGTTTaattaat5
5ttaaatACAATGCATACTTTGTACG3
     |||||||||||||||| tm 45.6 (dbd) 55.3
     3TGTTACGTATGAAACATGC...TTAGATTAGATTCAAAATTAATGTTT5
```

Suggested PCR programs for Taq polymerase and for Polymerases with DNA binding domain:

```
Taq (rate 30 nt/s)
Three-step|         30 cycles      |       |SantaLucia 1998
94.0°C    |94.0°C                  |       |SaltC 50mM
          |                 72.0°C |72.0°C |
04min00s  |30s  \                  |       |
          |      \ 50.0°C/ 0min17s |10min  |
          |       \____/           |       |
          |        30s             |       |4-8°C

Pfu-Sso7d (rate 15s/kb)
Three-step|         30 cycles      |       |Breslauer1986,SantaLucia1998
98.0°C    |98.0°C                  |       |SaltC 50mM
          |                 72.0°C |72.0°C |Primer1C   1µM
00min30s  |10s  \ 55.0°C _________ |_______|Primer2C   1µM
          |      \_____/ 0min 8s   |10min  |
          |       10s              |       |4-8°C
```

Step 2 Vector digestion and cloning

Clone the PCR product in pYPKa digested with ZraI resulting in pYPKa_Z_TEF1tp

Step 3 Diagnostic PCR confirmation

Confirm the structure of the pYPKa_Z_TEF1tp using primers 577, 342 and pfw579 in a multiplex PCR reaction.

Expected PCR products sizes from 577, 342 and pfw579 (bp):

pYPKa with insert in correct orientation: 1526, 1358
pYPKa with insert in reverse orientation: 1526, 760
Empty pYPKa clone : 934

Fig 10