WebLab的简单介绍

汇报人：任珍静

组　　员：王　芳（组长）

魏中艳

腰苗苗

WebLab

● **WebLab 注册**

● 读码框分析

● 核苷酸序列分析

● 蛋白质序列分析

WebLab 注册

不要出现汉字

WebLab 注册

登陆 (Login)

WebLab 注册

注册后可以方便我们进行如下操作：

1.可以在**My Data**中随时保存输入的资料及分析结果以方便查看。
2.可以在**My literature** 中保存和管理文献。
3.可以在 **My toolbox** 中收集一些常用的分析工具。

读码框分析

◆ **PlotORF**

◆ **ShowORF**

◆**GetORF**

读码框分析

◆PlotORF

功能：用图形的形式来预测可能的开放阅读框。

注意事项：

1.在这个程序中的ORF定义为起始密码子与终止密码子之间的序列。但是，这种定义就会使真核生物基因组序列中不包含起始密码子的外显子序列被遗漏，所以PlotORF更适合用于原核生物或真核生物的mRNA序列的分析。

2.起始密码子默认为ATG,终止密码子默认为TAA、TAG、TGA。当然我们可以根据具体情况重新定义起始密码子和终止密码子。比如线粒体和叶绿体以ATG、ATT、ATA为起始密码子。

以PPF-1(Pisum sativum post-floral-specific gene)即豌豆内膜蛋白编码基因PPF-1 为例。
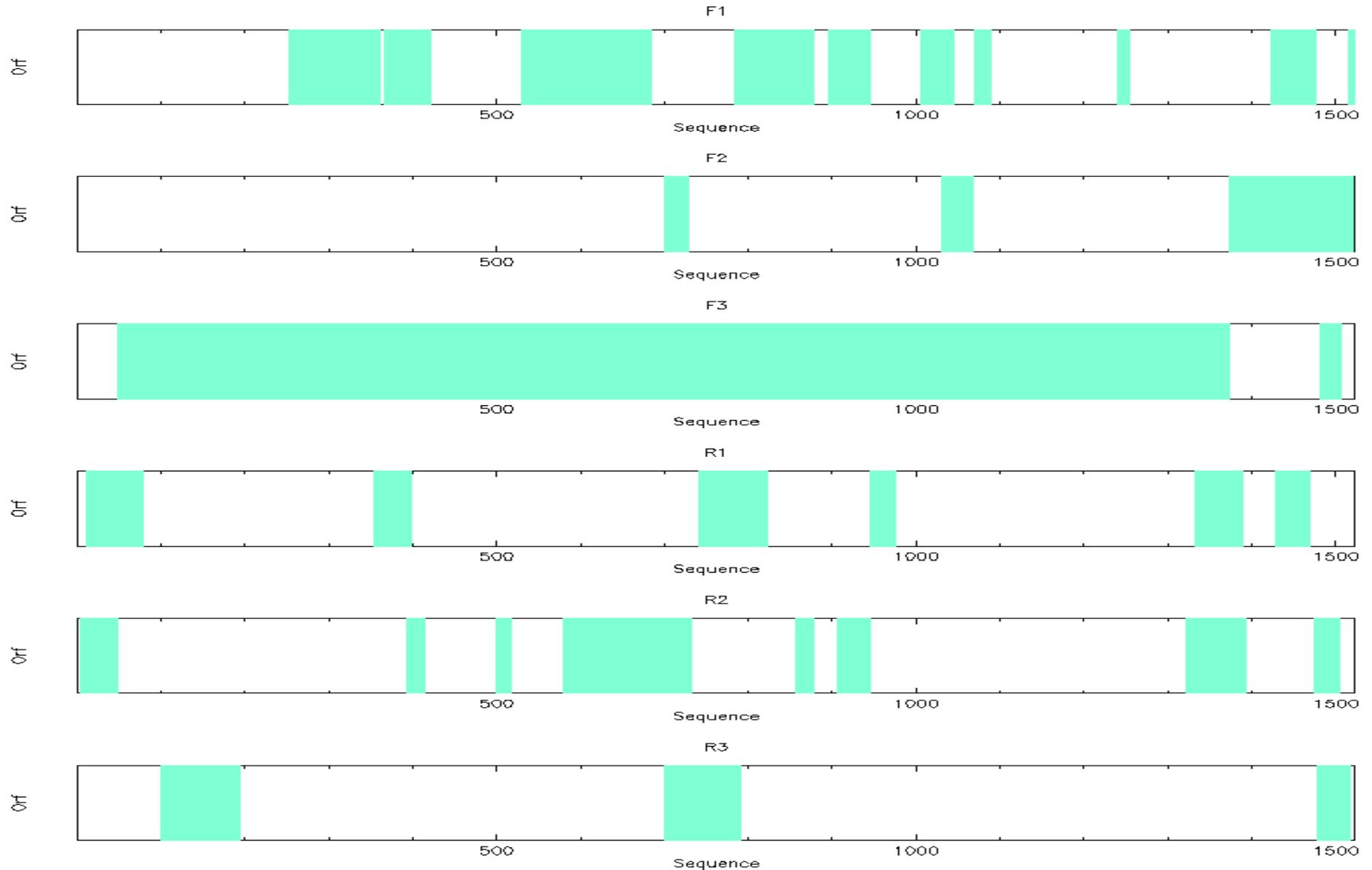
读码框分析

◆ Showorf

功能：把输入的核酸序列翻译成蛋白质的氨基酸序列。

注意事项：

　　在Basic Options中有六个选项，分别为F1、F2、F3、R1、R2、R3。其中，F代表Forward 即正向，R代表Reverse即反向。即从正向（F）或反向（R）第几个核苷酸开始翻译。

☐ **Basic Options**

* Select one or more values:　　　　　None ▼

None
F1
F2
F3
R1
R2
R3

☐ **Additional section**

Genetic code to use:

```
SHOWORF of Y12618.1 from 1 to 1523
        ---------|---------|---------|---------|---------|
     1  CTCAAGCCTTCAAGGCTGAAGCGTCTCGTACACAAACCTTCTCATC[ATG] 50
P3   1     Q  A  F  K  P  E  A  S  R  T  Q  T  F  S  S [M] 16
        ---------|---------|---------|---------|---------|
    51  GCGAAGACACTGATTTCTTCTCCATCATTCCTCGGTACTCCACTTCCTTC 100
P3  17  A  K  T  L  I  S  S  P  S  F  L  G  T  P  L  F  S  33
        ---------|---------|---------|---------|---------|
   101  ACTTCACCGTACTTTCTCCCCTAATCGCACCAGGCTTTTCACCAAAGTTC 150
P3  34  L  H  R  T  F  S  P  N  R  T  R  L  F  T  K  V  Q  50
        ---------|---------|---------|---------|---------|
   151  AATTCAGTTTCCACCAACTTCCTCCGATTCAATCCGTAAGTCATTCTGTT 200
P3  51     F  S  F  H  Q  L  P  P  I  Q  S  V  S  H  S  V  66
        ---------|---------|---------|---------|---------|
   201  GACTTATCCGGAATCTTCGCTAGAGCCGAAGGTTTACTTTACACGCTCGC 250
P3  67  D  L  S  G  I  F  A  R  A  E  G  L  L  T  T  L  A  83
        ---------|---------|---------|---------|---------|
   251  CGATGCTACTGTTGCGGCGGATGCGGCTGCTTCCACTGATGTTGCTGCGC 300
P3  84     D  A  T  V  A  A  D  A  A  A  S  T  D  V  A  A  Q 100
        ---------|---------|---------|---------|---------|
   301  AGAAGAACGGAGGTTGGTTCGGTTTTTATTTCTGATGGAATGGAGTTTGTT 350
P3 101     K  N  G  G  W  F  G  F  I  S  D  G  M  E  F  V  116
        ---------|---------|---------|---------|---------|
   351  CTCAAGGTGTTAAAGGATGGTTTGTCTTCCGTGCACGTGCCTTACTCATA 400
P3 117  L  K  V  L  K  D  G  L  S  S  V  H  V  P  Y  S  Y  133
        ---------|---------|---------|---------|---------|
   401  TGGATTTGCTATCATATTACTAACTGTTATTGTTAAGGCTGCTACACTTC 450
P3 134  G  F  A  I  I  L  L  T  V  I  V  K  A  A  T  L  F 150
        ---------|---------|---------|---------|---------|
   451  CCTTGACAAAGCAACAGGTTGAATCAACACTAGCTATGCAAAACCTTCAA 500
P3 151     L  T  K  Q  Q  V  E  S  T  L  A  M  Q  N  L  Q  166
        ---------|---------|---------|---------|---------|
   501  CCTAAAATTAAGGCCATTCAAGAAAGATATGCTGGCAATCAGGAAAGAAT 550
P3 167  P  K  I  K  A  I  Q  E  R  Y  A  G  N  Q  E  R  I  183
        ---------|---------|---------|---------|---------|
```

读码框分析

◆ GetORF

功能：可以直观地以多种不同的方式显示序列中的开放阅读框。

注意事项：

1.一个ORF可以被定义为两个终止密码子之间或者起始密码子与终止密码子之间区域。

2. ORF可以核酸序列或者翻译成的氨基酸序列输出。

核苷酸序列分析

在获得一个基因序列后，需要对其进行生物信息学分析，从中尽量发掘信息，从而指导进一步的实验研究。

根据所学知识，简单分为以下几个方面：

◆序列比对

◆密码子统计

◆内切酶分析

序列比对

包括全局比对和局部比对两类。

全局比对就是考虑序列间整体的相似性。

局部比对就是考虑分析序列间局部区域的相似性，如对某保守位点的比对，有较高的灵敏度。

Local Alignment

matcher(v6.0.1) - Waterman-Eggert local alignment of two sequences

seqmatchall(v6.0.1) - All-against-all word comparison of a sequence set

supermatcher(v6.0.1) - Calculate approximate local pair-wise alignments of larger sequences

water(v6.0.1) - Smith-Waterman local alignment of sequences

wordfinder(v6.0.1) - Match large sequences against one or more other sequences

wordmatch(v6.0.1) - Finds regions of identity (exact matches) of two sequences

Global Alignment

needle(v6.0.1) - Needleman-Wunsch global alignment of two sequences

stretcher(v6.0.1) - Needleman-Wunsch rapid global alignment of two sequences

密码子偏好统计

　　所用的软件是CUSP，以豌豆PPF-1基因密码子使用情况为例。

如图，选择目标序列进行分析，即选入豌豆PPF-1基因编码序列，选择输出位置，并可重新命名。

cusp(v6.0.1) - Create a codon usage table from nucleotide sequence(s)

☐ Input

⊞ * sequence:　　　　　|-135400672963.fasta ▼ ❓

☐ Output section

* save result in directory:　　　|- Y12618 ▼

* outfile: (bio:data:codon)　00.codon ❓

#CdsCount: 1

#Coding GC 43.04%
#1st letter GC 51.92%
#2nd letter GC 43.34%
#3rd letter GC 33.86%

#Codon AA Fraction Frequency Number
GCA A 0.205 18.059   8
GCC A 0.077   6.772   3
GCG A 0.154 13.544   6
GCT A 0.564 49.661 22
TGC C 0.000   0.000   0
TGT C 1.000   2.257   1
GAC D 0.182   4.515   2
GAT D 0.818 20.316   9
GAA E 0.750 40.632 18
GAG E 0.250 13.544   6

TTC F 0.526 22.573 10
TTT F 0.474 20.316   9
GGA G 0.355 24.831 11
GGC G 0.161 11.287   5
GGG G 0.032   2.257   1
GGT G 0.452 31.603 14
CAC H 0.600   6.772   3
CAT H 0.400   4.515   2
ATA I 0.130   6.772   3
ATC I 0.261 13.544   6
ATT I 0.609 31.603 14
AAA K 0.250 15.801   7
AAG K 0.750 47.404 21
CTA L 0.120 13.544   6
CTC L 0.120 13.544   6
CTG L 0.060   6.772   3
CTT L 0.340 38.375 17
TTA L 0.220 24.831 11

TCA S 0.262 24.831 11
TCC S 0.167 15.801   7
TCG S 0.024   2.257   1
TCT S 0.357 33.860 15
ACA T 0.407 24.831 11
ACC T 0.111   6.772   3
ACG T 0.037   2.257   1
ACT T 0.444 27.088 12
GTA V 0.097   6.772   3
GTC V 0.097   6.772   3
GTG V 0.194 13.544   6
GTT V 0.613 42.889 19
TGG W 1.000 15.801   7
TAC Y 0.455 11.287   5
TAT Y 0.545 13.544   6
TAA * 0.000   0.000   0
TAG * 0.000   0.000   0
TGA * 1.000   2.257   1

如图，这是部分分析结果，由图可看到密码子GCT，AAG，GAA，CTT，GTT等的使用频率是较高的。

内切酶分析

用Remap内切酶分析程序，来分析目标基因的酶 切位点，从而为引物设计提供依据。以豌豆PPF-1基因为例。

如图，选择目标基因，即选入豌豆PPF-1基因编码序列，选择输出位置，重新命名。

remap(v6.0.1) - Display restriction enzyme binding sites in a nucleotide s

Display sequence with restriction sites, translation etc.

⊟ **Input**

⊞ * sequence: |-1355400560506.fasta ▼ ❓

⊟ **Output section**

* save result in directory:	|- ppf ▼
* remap output: (unknown)	00.remap ❓
List the enzymes that cut:	Y ▼ ❓
Display RE sites in flat format:	N ▼ ❓
Limits reports to one isoschizomer:	Y ▼ ❓

下图为部分结果图

```
48-1376
Pisum sativum mRNA for PPF-1 protein

          Hin4I                      BccI
          |   BtsIMutI               BseDI
          |   |MboII                 BsaJI
          |   || BpiI                BssECI
          |   || BbsI            |     RsaNI
          |   || BpuAI           |     CviQI
          |   || BstV2I          |     Csp6I
          |   ||| TspRI          |     Hin4I
          |   ||| TscAI          |     |RsaI
          |   ||| MspJI          |     |AfaI
          |   ||| |MboII         |   || MnlI
          |   ||| || SgeI        |   ||  | SgeI
          |   ||| || MspJI       |   ||  | MspJI
          |   ||| ||MspJI        |   ||  |  HphI
          |   ||| |||| MspJI     |   ||  |  AsuHPI    HpyAV
          \   \\\ \\ \\ \        |   \\  \  \ \          \
ATGGCGAAGACACTGATTTCTTCTCCATCATTCCTCGGTACTCCACTTCCTTCACTTCAC
          10        20        30        40        50        60
----:----|----:----|----:----|----:----|----:----|----:----|
TACCGCTTCTGTGACTAAAGAAGAGGTAGTAAGGAGCCATGAGGTGAAGGAAGTGAAGTG
/       /   /    / ///      /   /    //// / // /     // /
Hin4I | MboII    |   |  |    || ||  MnlI | MspJI | HpyCH4III
    BtsIMutI |   |    |    || |Csp6I | SgeI | Bst4CI
    TscAI    |   |    |    || |CviQI AsuHPI | TaaI
    TspRI    |   |    |    || |RsaNI HphI  HpyAV
             |   |    |    ||  | AfaI
             |   |    |    ||  | RsaI
             |   |    ||    ||   BssECI
             |   |    ||    ||   BsaJI
             |   |    ||    ||   BseDI
             |   |    ||    |BccI
             |   |    ||    | Hin4I
             |   |    ||  MspJI
             |   |    ||MspJI
             |   |  MspJI
             |   | SgeI
             |   | MspJI
           BstV2I
           BpuAI
           MboII
           BbsI
           BpiI

 M  A  K  T  L  I  S  S  P  S  F  L  G  T  P  L  P  S  L  H
  W  R  R  H  *  F  L  L  H  H  S  S  V  L  H  F  L  H  F  T
   G  E  D  T  D  F  F  S  I  I  P  R  Y  S  T  S  F  T  S  P
----:----|----:----|----:----|----:----|----:----|----:----|
   X  A  F  V  S  I  E  E  G  D  N  R  P  V  G  S  G  E  S  *
  X  P  S  S  V  S  K  K  E  M  M  G  R  Y  E  V  E  K  V  E
 H  R  L  C  Q  N  R  R  W  *  E  E  T  S  W  K  R  *  K  V
```

```
          FatI
          MspJI
          HpyCH4V
          |CviAII
          ||FaiI
          ||| FaeI
          ||| NlaIII
          ||| Hin1II
          ||| Hsp92II
          \\\ \
GTTGCATGA

----:----

CAACGTACT
  // ///
  ||  ||MspJI
  ||  |||FatI
  ||  |CviAII
  || FaiI
  |HpyCH4V
  |Hsp92II
  |Hin1II
  |NlaIII
  |FaeI
  MspJI

 V  A  *
  L  H  X
   C  M  X
----:----
 T  A  H
Q  Q  M
  N  C  S
```

# Enzymes that cut	Frequency
AanI	1
Acc16I	2
Acc36I	1
AccIII	2
AciI	7
AclWI	5
AcoI	1
AcsI	1
AcvI	1
AfaI	2
AfiI	6
AgsI	9
AjnI	1
AjuI	1
AluBI	6
AluI	6
Alw21I	1
Alw26I	1
Alw44I	1
AlwI	5
Aor13HI	2
ApaLI	1
ApeKI	5
ApoI	1
AspLEI	2
AspS9I	1
AsuHPI	3

蛋白质序列分析

- Protein
  - Structure
    - Structure Comparison
      - domainrep(v0.1.0) - Reorder DCF file to identify representativestructures.
    - Structure Prediction
    - Structural Features
    - Structure Retrieval and Visualization
      - pepnet(v6.0.1) - Draw a helical net for a protein sequence
      - pepwheel(v6.0.1) - Draw a helical wheel diagram for a protein sequence
      - topo(v1.0.0) - Draws an image of a transmembrane protein
  - Similarity Detecting
  - Sequence Feature Identifying
    - Antigenic Sites
    - Biochemical Features
    - checktrans(v6.0.1) - Reports STOP codons and ORF statistics of a protein
    - Composition
    - Motifs
    - Domains and Motifs
    - Localization and Targeting
    - Transmembrane Region
      - tmap(v6.0.1) - Predict and plot transmembrane segments in protein sequences
  - Conceptual Back Translation
  - Enzyme
  - Phylogeny Reconstruction
  - Mass Spectrometry Analysis
  - Sequence Feature Display
  - Mutation

- Mutation
- Data Manipulation
- Literature
- Ontology
- DNA/RNA
- Pathway
- Sequence Comparison
  - Multiple Sequence Alignment
  - Database Searching
  - Analysis of Aligned Sequences
  - Pairwise Sequence Alignment
    - diffseq(v6.0.1) - Compare and report features of two similar sequences
    - dotmatcher(v6.0.1) - Draw a threshold dotplot of two sequences
    - dotpath(v6.0.1) - Draw a non-overlapping wordmatch dotplot of two sequences
    - dottup(v6.0.1) - Displays a wordmatch dotplot of two sequences
    - esim4(v1.0.0) - Align an mRNA to a genomic DNA sequence
    - Local Alignment
    - Global Alignment
    - polydot(v6.0.1) - Draw dotplots for all-against-all comparison of a sequence set
  - Alignment Editing and Visualization
-
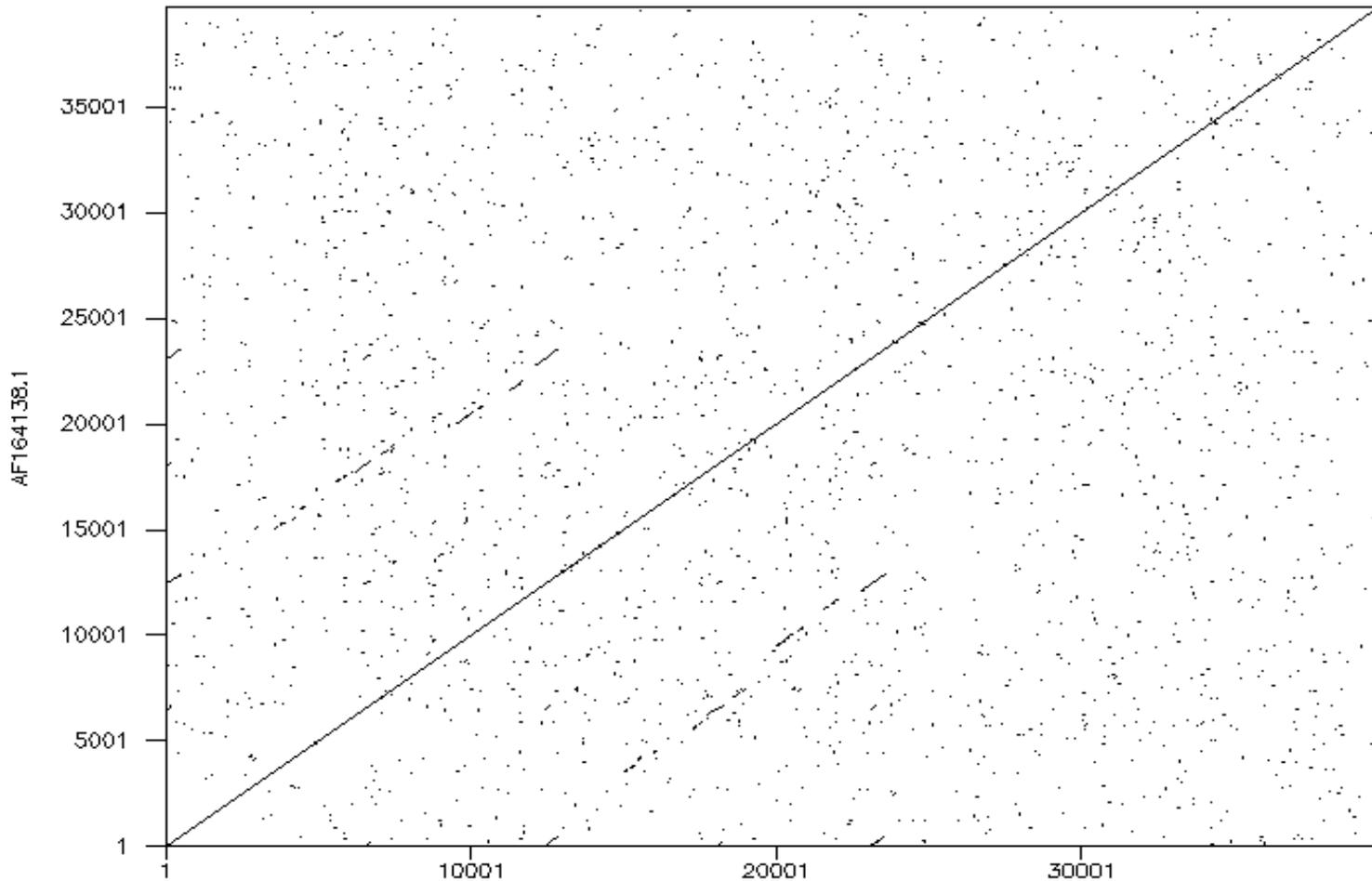- Expression
- Database Construction

Dottup

主要功能是用图形来表示两个序列中的重复片段。

Dottup: fasta::586402:AF164138.1 vs fasta::586403:AF1641...
Fri 18 Jan 2013 09:31:22

提取GenBank中刘勇等提交的柯氏质粒河豚鱼基因序列片段全长序列，得到的结果如上图所示：说明该基因包含重复序列。

关于蛋白质二级结构预测的工具有很多，例如：pepcoil、pepnet、tmap、pepwheel等工具，我们以tmap和pepwheel为例：

◆ tmap是预测蛋白质的跨膜螺旋区域；
◆ pepwheel是用α-螺旋轮显示氨基酸残基在α-螺旋上的分布情况。

tmap

tmap(v6.0.1) - Predict and plot transmembrane segments in protein sequences   +add to ToolBox

Displays membrane spanning regions

⊟ Input

⊟ * sequences: ------------- ▼   ❓

OR upload file from local disk: [_____] 浏览...

OR paste into window:

```
MAKTLISSPSFLGTPLPSLHRTFSPNRTRLFTKVQFSFHQLPPIQSVSHSVDLSGIFARA
EGLLYTLADATVAADAAASTDVAAQKNGGWFGFISDGMEFVLKVLKDGLSSVHVPYSYGF
AIILLTVIVKAATLPLTKQQVESTLAMQNLQPKIKAIQERYAGNQERIQLETSRLYTQAG
VNPLAGCLPTLATIPVWIGLYQALSNVANEGLLTEGFLWIPSLGGPTSIAARQSGSGISW
LFPFVDGHPLLGWYDTAAYLVLPVLLIVSQYVSMEIMKPPQTNDPNQKNTLLIFKFLPLM
IGYFSLSVPSGLTIYWFTNNVLSTAQQVWLRKLGGAKPAVNENAGGIITAGQAKRSASKP
EKGGERFRQLKEEEKKKKLIKALPVEEVQPLASASASNDGSDVENNKEQEVTEESNTSKV
SQEVQSFSRERRSKRSKRKPVA
```

⊟ Output section

* save result in directory: |- ppf   ▼

graph output file: (image)   tmap.jpg   ❓

graph output format: Portable Network Graphic (png)   ▼

* outfile: (bio:report:emboss:seqtable)   pepwheel.seqtable   ❓

⊟ Advanced Options

graph title: [_____]

graph subtitle: [_____]

graph x axis title: [_____]

graph y axis title: [_____]

Run
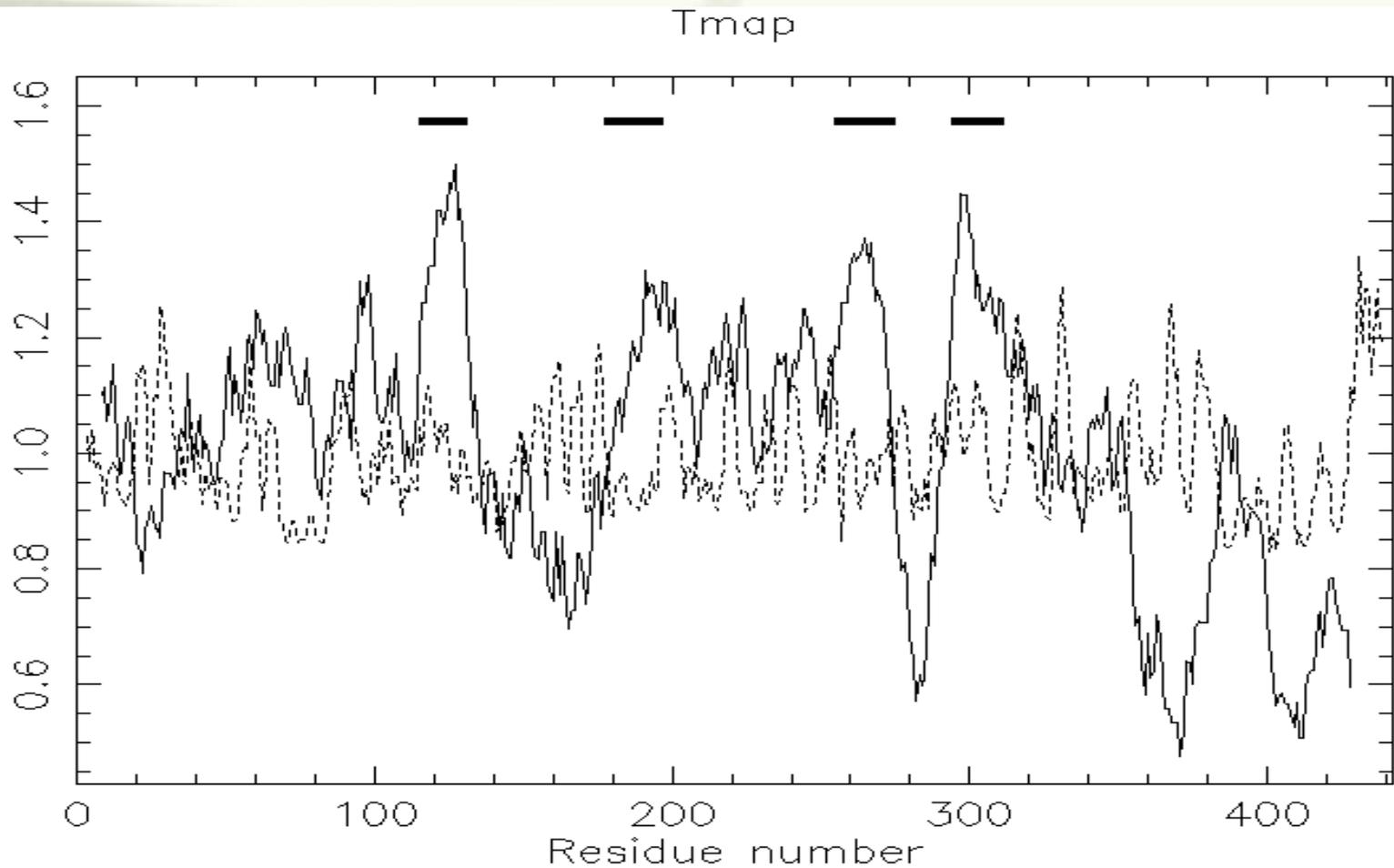
Tmap

提取**PPF-1**的蛋白质序列，利用**tmap**进行分析。从图中可以看出，有四个跨膜螺旋区域；

Result: `pepwheel.seqtable.2013-01-18 20:24 PM ▼` View: `Raw View ▼` `View`

Frier tmap..2013-01-18 20:24 PM
`pepwheel.seqtable.2013-01-18 20:24 PM`

```
###################################
# Program: tmap
# Rundate: Fri 18 Jan 2013 20:24:14
# Commandline: tmap
#     -auto
#     -graph png
#     -sequences 586585
#     -goutfile 586586
#     -outfile 586587
# Report_format: seqtable
# Report_file: 586587
###################################
#==================================
#
# Sequence: Consensus      from: 1    to: 442
# HitCount: 4
#==================================
  Start      End TransMem Sequence
   111      135         1 SVHVPYSYGFAIILLTVIVKAATLP
   173      201         2 SRLYTQAGVNPLAGCLPTLATIPVWIGLY
   251      279         3 LGWYDTAAYLVLPVLLIVSQYVSMEIMKP
   290      315         4 TLLIFKFLPLMIGYFSLSVPSGLTIY
#----------------------------------
#----------------------------------
#==================================
#
# Sequence:         from: 1    to: 442
# HitCount: 4
#==================================
  Start      End TransMem Sequence
   111      135         1 SVHVPYSYGFAIILLTVIVKAATLP
   173      201         2 SRLYTQAGVNPLAGCLPTLATIPVWIGLY
   251      279         3 LGWYDTAAYLVLPVLLIVSQYVSMEIMKP
   290      315         4 TLLIFKFLPLMIGYFSLSVPSGLTIY
#----------------------------------
#----------------------------------
#----------------------------------
# Total_sequences: 2
# Total_hitcount: 8
#----------------------------------
```

Pepwheel

☐ Input

☐ * sequence: `-------------` ▼   ❓

OR upload file from local disk: [   ] [浏览...]

OR paste into window:

```
MAKTLISSPSFLGTPLPSLHRTFSPNRTRLFTKVQFSFHQLPPIQSVSHSVDLSGIFARA
EGLLYTLADATVAADAAASTDVAAQKNGGWFGFISDGMEFVLKVLKDGLSSVHVPYSYGF
AIILLTVIVKAATLPLTKQQVESTLAMQNLQPKIKAIQERYAGNQERIQLETSRLYTQAG
VNPLAGCLPTLATIPVWIGLYQALSNVANEGLLTEGFLWIPSLGGPTSIAARQSGSGISW
LFPFVDGHPLLGWYDTAAYLVLPVLLIVSQYVSMEIMKPPQTNDPNQKNTLLIFKFLPLM
IGYFSLSVPSGLTIYWFTNNVLSTAQQVWLRKLGGAKPAVNENAGGIITAGQAKRSASKP
EKGGERFRQLKEEEKKKKLIKALPVEEVQPLASASASNDGSDVENNKEQEVTEESNTSKV
SQEVQSFSRERRSKRSKRKPVA
```

☐ Output section

* save result in directory:	|- ppf ▼	
Plot the wheel:	Y ▼	
Number of steps:	20	❓
Number of turns:	7	❓
graph output file: (image)	pepwheel.jpg	❓
* graph output format:	Portable Network Graphic (png) ▼	

⊞ Advanced Options

☐ Markup section

Prompt for amphipathic residue marking:		❓
Mark as squares:	ILVM	❓
Mark as diamonds:	DENQST	❓
Mark as octagons:	HKR	❓

[Run]

Helical wheel of raw::586409
Fri 18 Jan 2013 09:54:41

　　提取PPF-1基因的对应的蛋白质序列，利用pepwheel分析。其中黑色带圆圈的是极性带电荷的氨基酸，亲水性比较强；红色带方框的是极性不带电荷的氨基酸，也是亲水性的；紫色字母不带框的是非极性氨基酸，具有疏水性；蓝色带方框的是非极性氨基酸，疏水性比紫色的强。

Thanks for your attention !