

# 实用生物信息技术 期末报告

---

北京大学 G01+G08组

# ExpPASy: 这是什么?

The screenshot shows the homepage of the ExpPASy Bioinformatics Resource Portal. At the top left, there is a logo for SIB (Swiss Institute of Bioinformatics) and the text "ExpPASy Bioinformatics Resource Portal". Below the logo is a search bar with a dropdown menu set to "Query all databases". On the left side, there is a navigation menu with several red buttons: "Visual Guidance", "Categories", "Resources A..Z", and "Links/Documentation". Under the "Categories" button, a list of scientific fields is displayed: proteomics, genomics, structure analysis, systems biology, evolutionary biology, population genetics, transcriptomics, biophysics, imaging, IT infrastructure, medicinal chemistry, and glycomics. The main content area features a paragraph describing ExpPASy as a portal providing access to scientific databases and software in life sciences. Below this is a "Featuring today" section highlighting UniProtKB, described as a Protein Knowledgebase with functional information and rich manual and automated annotation, with a link to "[details]". At the bottom, a "How to use this portal?" section includes a question mark icon and a list of links: "Features and updates", "New to ExpPASy", and "Experienced ExpPASy users: what is different".

**Visual Guidance**

**Categories**

- proteomics
- genomics
- structure analysis
- systems biology
- evolutionary biology
- population genetics
- transcriptomics
- biophysics
- imaging
- IT infrastructure
- medicinal chemistry
- glycomics

**Resources A..Z**

**Links/Documentation**

ExpPASy is the **SIB Bioinformatics Resource Portal** which provides access to scientific databases and software in life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics. In this portal you find resources from many different SIB groups as well as external institutions.

**Featuring today**

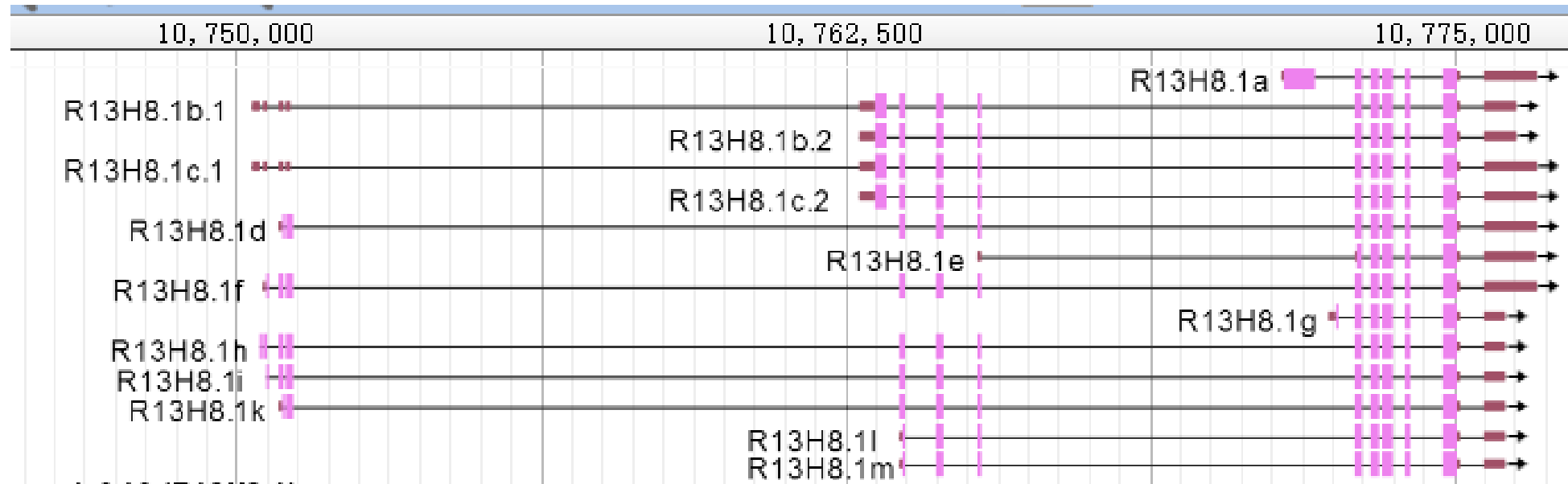
**UniProtKB**

Protein Knowledgebase with functional information and rich manual and automated annotation  
[\[details\]](#)

**How to use this portal?**

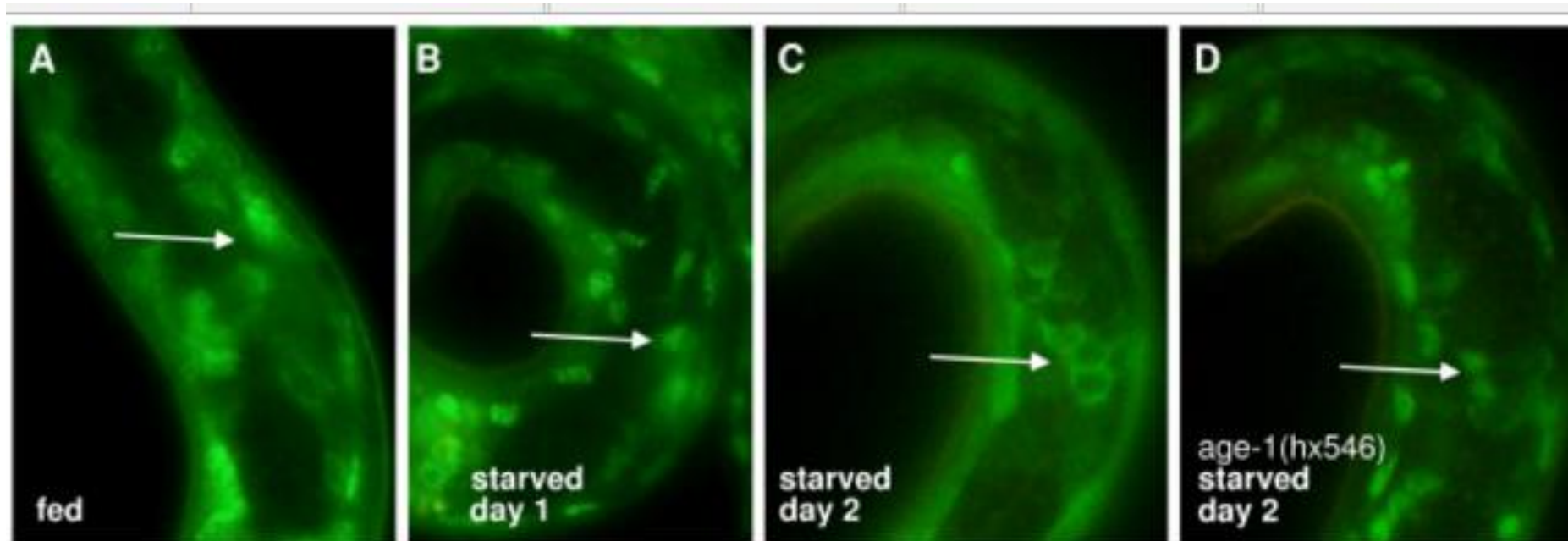
- Features and updates
- New to ExpPASy
- Experienced ExpPASy users: what is different

# *daf-16* isoform



# Subcellular localization from cytoplasm to nuclear

---



# 发挥功能的分子通路

---

GO:0000122	negative regulation of transcription from RNA polymerase II promoter <b>has regulation target:</b> <i>natc-1</i>
GO:0002119	nematode larval development
GO:0002376	immune system process
GO:0002821	positive regulation of adaptive immune response
GO:0006351	transcription, DNA-templated
GO:0006355	regulation of transcription, DNA-templated
GO:0006974	cellular response to DNA damage stimulus
GO:0006979	response to oxidative stress
GO:0007275	multicellular organism development
GO:0007568	aging

# ExPASy中的基因组学工具

## 序列比对工具：

- boxshade: 让序列比对结果打印出来更美观
- ClustalW: 多序列比对
- Decrease redundancy: 序列冗余度降低
- LALIGN: 配对序列比对
- MADAP: 基因注释数据聚类
- Newick Utilities: 高通量系统发生树构建
- Phylogibbs: 辨别调控位点
- SiBsim4: 剪接序列比对
- SwissTree: 金标准参考基因树
- T-Coffee: 多序列和结构比对











## Tools

-  [Alignment tools](#) • Four tools for multiple alignments • [\[more\]](#)
-  [boxshade](#) • MSA pretty printer • [\[more\]](#)
-  [ClustalW](#) • Multiple sequence alignment • [\[more\]](#)
-  [ClustalW - PBIL](#) • Multiple sequence alignment program • [\[more\]](#)
-  [ClustalW2](#) • Multiple sequence alignment program • [\[more\]](#)
-  [Decrease redundancy](#) • Sequence redundancy reduction • [\[more\]](#)
-  [DIALIGN](#) • Local multiple sequence alignment • [\[more\]](#)
-  [GENIO/logo](#) • RNA/DNA & Amino Acid Sequence Logos • [\[more\]](#)
-  [Kalign - EBI](#) • Fast and accurate multiple sequence alignment • [\[more\]](#)
-  [Kalign - SBC](#) • Fast and accurate multiple sequence alignment • [\[more\]](#)
-  [LALIGN](#) • Pairwise alignment • [\[more\]](#)
-  [MADAP](#) • clustering for genome annotation data • [\[more\]](#)
-  [MAFFT - CBRC](#) • Multiple sequence alignment • [\[more\]](#)
-  [MAFFT - EBI](#) • Multiple sequence alignment • [\[more\]](#)
-  [MaxAlign](#) • Gap removal from alignments • [\[more\]](#)
-  [Multalin](#) • Multiple sequence alignment • [\[more\]](#)
-  [MUSCLE](#) • Multiple alignment server • [\[more\]](#)
-  [Newick Utilities](#) • high-throughput phylogenetic tree processing • [\[more\]](#)
-  [Phylogibbs](#) • regulatory sites discovery • [\[more\]](#)
-  [SiBsim4](#) • spliced sequence alignment • [\[more\]](#)
-  [SwissTree](#) • Gold standard reference gene trees • [\[more\]](#)
-  [T-Coffee](#) • sequence and structure multiple alignments • [\[more\]](#)
-  [T-Coffee - EBI](#) • Multiple sequence alignment program • [\[more\]](#)
-  [T-Coffee - WUR](#) • Multiple sequence alignment program • [\[more\]](#)
-  [WebLogo](#) • Sequence logos • [\[more\]](#)

# ExPASy中的基因组学工具

## 相似性搜索工具:

- BUSCO: 利用单拷贝直系同源基因,评估基因组,转录组的组装质量
- LALIGN: (同前)
- Phylogibbs: (同前)


















-  [BLAST - NCBI](#) • Biological sequence similarity search • [\[more\]](#)
-  [BLAST - PBIL](#) • BLAST search on protein sequence databases • [\[more\]](#)
-  [Blast2Fasta](#) • Blast to Fasta conversion • [\[more\]](#)
-  [BUSCO](#) • Benchmarking Universal Single-Copy Orthologs • [\[more\]](#)
-  [FASTA/SSEARCH/GGSEARCH/GLSEARCH](#) • Sequence similarity searching of protein db • [\[more\]](#)
-  [LALIGN](#) • Pairwise alignment • [\[more\]](#)
-  [Phylogibbs](#) • regulatory sites discovery • [\[more\]](#)
-  [SAMBA](#) • Systolic Accelerator for Molecular Biological Appl • [\[more\]](#)
-  [Sequerome](#) • BLAST similarity search and sequence profiling • [\[more\]](#)
-  [WU BLAST](#) • Sequence similarity search in protein databases • [\[more\]](#)

# ExpASy中的基因组学工具

## 描述与注释工具:

- EPD: 真核启动子辑录
- OMA: 全基因组同源性参考
- Association Viewer: 统一显示单核苷酸突变
- ChIP-Seq: 染色质免疫共沉淀分析
- COILS: 预测蛋白质的卷曲螺旋
- EIMMo: miRNA靶点预测
- ESTscan: 编码区探测
- GMM: 拷贝数探测
- IScan: 插入位点探测
- ISMARA: 基因组表达数据建模
- IsotopIdent: 理论同位素分布
- MAMOT: HMM模型
- SSM: 序列相似性作图
- ShoRAH: 分析NGS数据
- SSA: 核酸序列基序
- TriFLe: TRFLP物种识别和模拟
- tromeR: 转录组分析器

## Tools

-  [EPD](#) • collection of eukaryotic promoters • [\[more\]](#)
-  [OMA](#) • orthology inference among complete genomes. • [\[more\]](#)
  
-  [Association Viewer](#) • SNPs display in a genetic context • [\[more\]](#)
-  [ChIP-Seq](#) • ChIP-Seq data analysis tools • [\[more\]](#)
-  [COILS](#) • Prediction of Coiled Coil Regions in Proteins • [\[more\]](#)
-  [EIMMo](#) • miRNA target predictions • [\[more\]](#)
-  [ESTscan](#) • coding region detection • [\[more\]](#)
-  [GMM](#) • copy number variation detection • [\[more\]](#)
-  [IScan](#) • insertion sequence identification • [\[more\]](#)
-  [ISMARA](#) • genome-wide expression data modeling • [\[more\]](#)
-  [IsotopIdent](#) • theoretical isotopic distribution • [\[more\]](#)
-  [MAMOT](#) • HMM models • [\[more\]](#)
-  [Sequence Similarity Maps \(SSM\)](#) • data mining of viral isolates • [\[more\]](#)
-  [ShoRAH](#) • tools for the analysis of NGS data • [\[more\]](#)
-  [SSA](#) • nucleic acid sequence motifs • [\[more\]](#)
-  [TriFLe](#) • TRFLP species identification and simulation • [\[more\]](#)
-  [tromeR](#) • transcriptome analyser • [\[more\]](#)



# ExpASy中的基因组学工具

## Clustal-W多序列比对算法

ClustalW

Valid format for input is: FASTA(Pearson)  
max number of sequences = 30  
max total length of sequences = 10000

[Help page](#)

[More information on Clustal home page](#)

Scoring matrix :	Blosum ▾	Extending gap penalty :	0.05
Opening gap penalty :	10	Separation gap penalty :	0.05
End gap penalty :	10	Output order :	Input ▾
Output format :	Clustal ▾		

Input sequences:  
(see above for valid formats)

ExpASy  
Bioinformatics Resource Portal

ClustalW

Thanks. Your request has been filed with the following data:

Number: 5537  
e-mail:  
Format: aln

Processing, please wait...

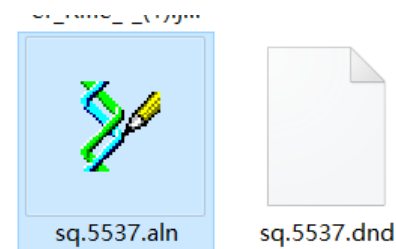
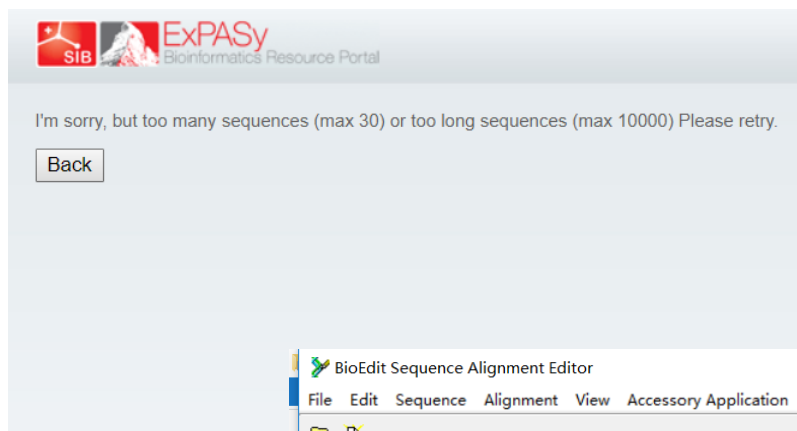
Here are your search results:

Multiple alignments	Dendrograms
<a href="#">clustalw (aln)</a>	<a href="#">clustalw (dnd)</a>
<a href="#">GCG/MSF</a>	
<a href="#">PIR</a>	
<a href="#">GDE</a>	
<a href="#">phylip</a>	

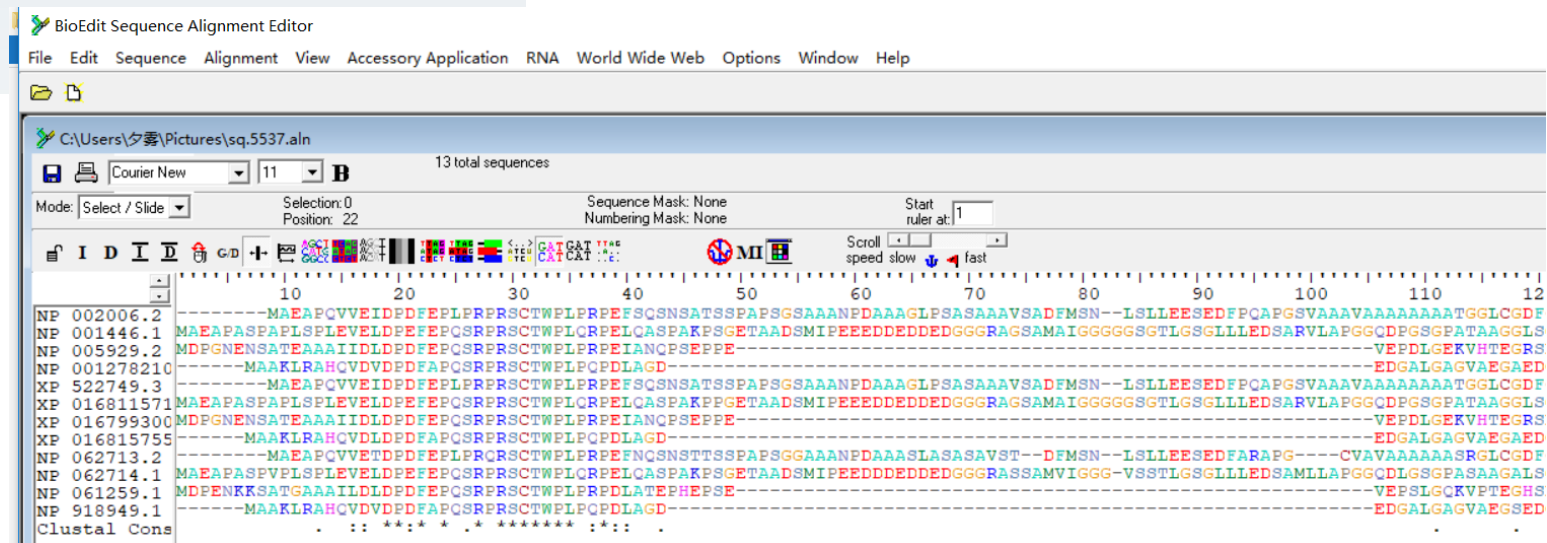
Thank you for using ClustalW

注意，ExPASy中的Cluster-W:

- 1、不能分析超过30个样本或10000个残基，
- 2、且得到的结果仍要用其他生信软件（如Bioedit）才能浏览、分析。



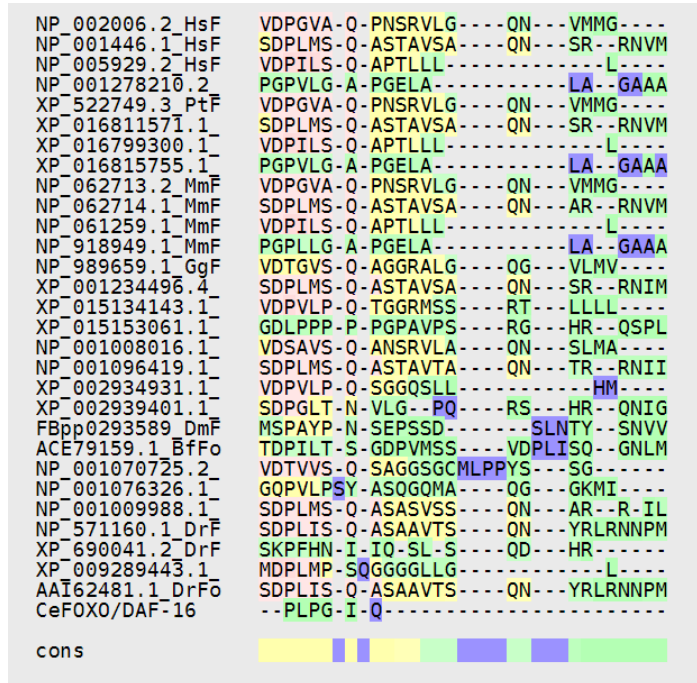
报错了.....



结果（在Bioedit中打开）

# T-Coffee: 寻找同源序列中的保守位点

FOX家族基因中的一些保守位点（红色）和家族中整体保守区域的状况（最下一排），可以看到一些特别保守的区段。



# ExPASy中的演化分析工具



A screenshot of the ExPASy website showing a list of bioinformatics tools. The tools are listed with their names, brief descriptions, and links to more information. The list includes:

- OMA • orthology inference among complete genomes. • [more]
- ALF • simulation of genome evolution • [more]
- BayeScan • identify natural selection • [more]
- BIONJ • Neighbor Joining (NJ) phylogenetic analysis • [more]
- BLASTO • BLAST on orthologous groups • [more]
- BUSCO • Benchmarking Universal Single-Copy Orthologs • [more]
- CEGA • Conserved Elements from Genomic Alignments • [more]
- Coev-Web • The Coev-Web is a web interface to Coev program. • [more]
- CT-CBN • estimate conjunctive Bayesian networks • [more]
- DendroUPGMA • Dendrogram creation using the UPGMA algorithm • [more]
- Evolutionary Trace Server (TraceSuite II) • Maps evolutionary traces to structures • [more]
- fastsimcoal • coalescent simulation of genomic data • [more]
- Genonets • Genotype network analysis • [more]
- MLtree • maximum likelihood optimization • [more]
- MLTreeMap • phylogenetics and functionalities of metagenomes • [more]
- Newick Utilities • high-throughput phylogenetic tree processing • [more]
- PHYLIP • Package of programs for phylogenetic analysis • [more]
- Phylo.io • Phylogenetic trees visualisation and comparison. • [more]
- Phylogeny.fr • Simple phylogenetic analysis • [more]
- Phylogeny programs • Links to phylogeny programs • [more]
- RAxML • ML inference of large phylogenetic trees • [more]
- REALPHY • Automatic inference of phylogenetic trees • [more]
- SuperTree • assemble phylogenetic trees • [more]
- SwissTree • Gold standard reference gene trees • [more]
- The Phylogenetic Web Repeater (POWER) • Phylogenetic analysis • [more]

- 计算选择压力

ALF、BayeScan

- 全基因组分析

OMA、BUSCO、  
fastsimcoal、MLTreeMap

- 查找保守序列

CEGA

- 共演化分析

Coev-Web

- 构建演化树相关

(演化树构建、可视化、优化、合并等)

Newick Utilities、

RAxML、REALPHY、

Phylo.io、MLtree、

Supertree

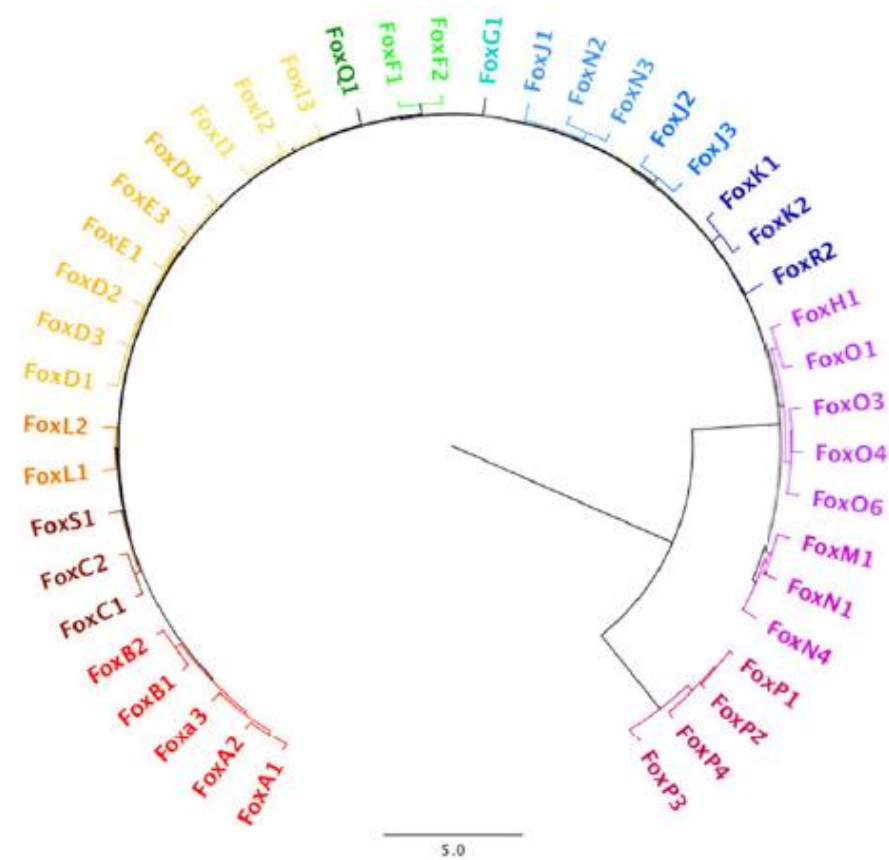
# Fox (forkhead box) 转录因子家族

广泛存在于线虫~哺乳动物

19个亚家族

命名为A~S

线虫**DAF-16 (FOXO)**  
属于Fox**O**亚家族



Phylogenetic tree of mouse Fox family members

*Maria L. Golson and Klaus H. Kaestner, Development (2016)*

# FoxO亚家族代表物种演化树的构建

## Genome-wide identification and characterization of Fox genes in the silkworm, *Bombyx mori*

JiangBo Song<sup>1,2</sup> · ZhiQuan Li<sup>1,2</sup> · XiaoLing Tong<sup>1,2</sup> · Cong Chen<sup>1,2</sup> ·  
Min Chen<sup>1,2</sup> · Gang Meng<sup>1,2</sup> · Peng Chen<sup>1,2</sup> · ChunLin Li<sup>1,2</sup> ·  
YaQun Xin<sup>1,2</sup> · TingTing Gai<sup>1,2</sup> · FangYin Dai<sup>1,2</sup> · Cheng Lu<sup>1,2</sup>

同源蛋白序列（全长氨基酸序列）的获得：

文献查找 & NCBI蛋白数据，BLASTP 共30条序列

线虫 *Caenorhabditis elegans* (1) : CeFOXO/DAF-16

文昌鱼 *Branchiostoma floridae* (1) : BfFoxO

斑马鱼 *Danio rerio* (7) : DrFoxO1a、DrFoxO1b、  
DrFoxO3a、DrFoxO3、DrFoxO4、DrFoxO6

果蝇 *Drosophila melanogaster* (1) : DmFoxO

热带爪蟾 *xenopus tropicalis* (4) : XtFoxO1、XtFoxO3、XtFoxO4、XtFoxO6

鸡 *Gallus gallus* (4) : GgoxO1、GgFoxO3、GgFoxO4、GgFoxO6

小鼠 *Mus musculus* (4) : MmFoxo1、MmFoxo3、MmFoxo4、MmFoxo6

黑猩猩 *Pan troglodytes* (4) : PtFoxO1、PtFoxO3、PtFoxO4、PtFoxO6

人 *Homo sapiens* (4) : HsFOXO1、HsFOXO3、HsFOXO4、HsFOXO6

## The Fox genes of *Branchiostoma floridae*

Jr-Kai Yu · Francoise Mazet · Yen-Ta Chen ·  
Song-Wei Huang · Kuo-Chen Jung ·  
Sebastian M. Shimeld

# FoxO亚家族代表物种演化树的构建

序列的比对:

ClustalW Mutiple (BioEdit)

保存为Phylip格式

系统发生树的构建:

RAxML (ExPSAy网站版)

最大似然法 (maximum likelihood, ML)

Home Manual Software Cluster status Mirror at CIPRES

## RAxML BlackBox

**Data**

Paste your sequence alignment

or upload a file  未选择任何文件

Gamma model of rate heterogeneity

Protein sequences

Outgroup

Constraint  未选择任何文件

Binary backbone  未选择任何文件

**Model**

Must be in Phylip format or convertible by ReadSeq

If not checked, CAT model will be used

If not checked, dna sequences will be assumed

The name of one or more (comma-separated) outgroup(s) can be specified

This option allows you to specify an incomplete or complete strictly bifurcating constraint tree in NEWICK format. More help here

This option allows you to pass a binary/bifurcating constraint/backbone tree in NEWICK format. More help here

激活 Windows  
转到“设置”以激活 Windows

步骤一：粘贴比对好的序列或上传文件

步骤二：勾选蛋白序列

如果有外类群或作为限制的NEWICK形式的树，在此设置

# FoxO亚家族代表物种演化树的构建

## Substitution Matrix

Substitution matrix

BLOSUM62

Use empirical base frequencies

步骤三:修改替换矩阵

## Model

Use a mixed/partioned model?

选择文件 未选择任何文件

Maximum likelihood search

如果有用于建树的比对序列的区域的限制,在此设置  
步骤四:勾选ML法

## Notification

Email

Enter your email if you want to be notified upon job completion

可以填写邮箱,接收结果

Run

步骤五: run

When publishing results obtained via the web-server please cite:

A. Stamatakis, P. Hoover, J. Rougemont

[A Rapid Bootstrap Algorithm for the RAxML Web-Servers,](#)

Systematic Biology, 75(5): 758-771, 2008

激活 Windows



# FoxO亚家族代表物种演化树的构建

结果页面

## RAxML BlackBox

### 100 bootstrapped trees

[Download](#), [View](#).

### Consensus tree

[Download \(MRE\)](#), [View \(MRE\)](#).

[Download \(MR\)](#), [View \(MR\)](#).

[Download \(STR\)](#), [View \(Strict\)](#).

### Best-scoring ML tree

Tree with branch lengths only: [Download](#), [View](#).

Tree with branch lengths and support values: [Download](#), [View](#).

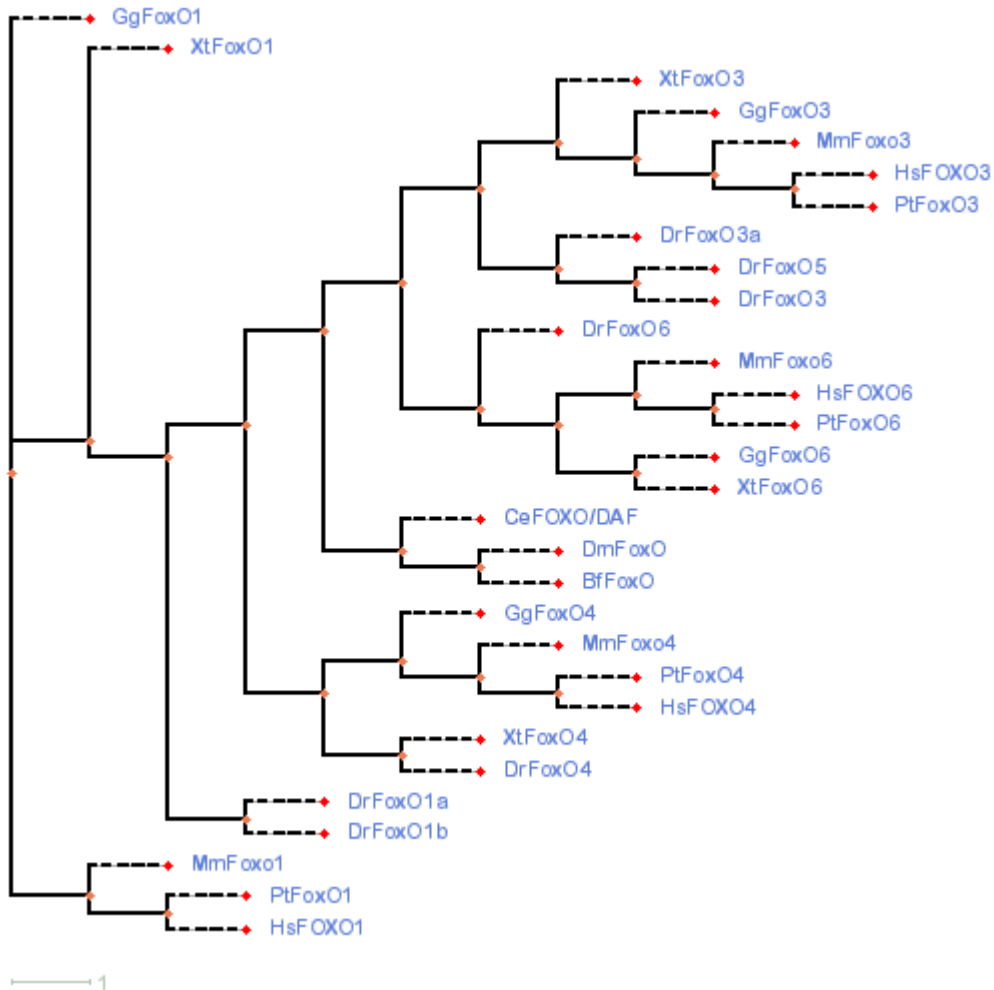
可以点击**view**查看结果  
或下载结果

### Job report

[Download](#), [View](#).

Results will be kept here for one week.

# FoxO亚家族代表物种演化树的构建



RAXML网页版优点是不用安装软件、操作简单；缺点是等待时间长，不能选择参数，不能对结果进行优化编辑。


另外，RAXML具有下载版软件，可以在Linux, MacOS, DOS下运行，优点是可以处理超大规模的数据，例如千万个物种，几百至上万个的碱基序列，但操作复杂，上手慢。

# 转录组（transcriptomics）相关数据库

## Databases

 [Bgee](#) • gene expression patterns comparison • [\[more\]](#)












 [SwissRegulon](#) • annotations of regulatory sites • [\[more\]](#)

 [EIMMo](#) • miRNA target predictions • [\[more\]](#)

**Bgee**: 提供了在正常生理条件（没有涉及基因突变、处理、疾病等）下各种动物体内的**基因表达模式**，这些数据来自RNA-seq、基因芯片、原位杂交、EST数据等；  
**SwissRegulon**: 提供了全基因组范围的调控位点的注释信息，主要是**转录因子结合位点**；

# 转录组（transcriptomics）相关工具

## Tools

-  EIMMo • miRNA target predictions • [\[more\]](#)
-  ESTscan • coding region detection • [\[more\]](#)
-  ExpressionView • explore biclusters in gene expression data • [\[more\]](#)
-  ISA • gene expression module discovery • [\[more\]](#)
-  ISMARA • genome-wide expression data modeling • [\[more\]](#)
-  Ping pong algorithm • coherent patterns across paired datasets • [\[more\]](#)
-  QuasR • Quantify and Annotate Short Reads in R • [\[more\]](#)
-  SIBsim4 • spliced sequence alignment • [\[more\]](#)
-  The Miner Suite • Tools for data analysis • [\[more\]](#)
-  Translate • nucleotide sequence translation • [\[more\]](#)
-  tromer • transcriptome analyser • [\[more\]](#)

**EIMMo**: 对microRNA作用的下游靶基因进行预测;

**ESTscan**和**Translate**: 首先进行CDS序列的预测, 然后使用6个读码框把核酸序列翻译成蛋白质序列;

**ExpressionView**: 在R语言环境下对**基因表达数据**进行双聚类的分析;

**Ping pong algorithm and ISA**: 对大规模的的数据, 比如说**微阵列 (microarray)**, 进行初步的模块处理来减少其复杂度;


**ISMARA**: 对**Chip-seq**数据和预测得到的**TF/microRNA 调控位点**进行建模和初步分析;

**QuasR**: 在R语言的环境下对测序得到的short reads进行定量分析和注释;

**SIBsim4**: 在Linux操作系统下, 对sim4进行改进, 提高了运算速度和信息的输出量, 对序列成熟的mRNA序列进行比对;

**Tromer**: 转录组分析, 输入基因名后可以找到转录信息;

# Translate – daf-16

 Translate  
Translation of a nucleotide (DNA/RNA) sequence to a protein sequence. [\[less\]](#)  
Keywords: codon, conversion tool, translation



## Translate tool

Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.

Please enter a DNA or RNA sequence in the box below (numbers and dashes are allowed).

```
TACTTCAGGGATAAGGGAGATTGGAACAGTTCAGCTGGATGGAAGAACTGATCCGTCACAATCTGTCTC
TTCATTCTCGTTTCATGCGAATTCAGAA TGAAGGAGCOGGAAGAGCTCGTGGTGGGTTATTAATCCAGA
TGCAAAAGCAGGAAGGAAATCCAGGCGTACACG TGAACGATCCAACTACTATTGAGACGACTACAAAGGCT
CAACTCGAAAAATCTCGCCGCGAGOC AAG AAG AGGATAAAGGAGAGAGCATGATGGGC TOCCTTCACT
CGACACTTAAATGGAAAATTCGATTGCOGGATCGATTCAAAACGATTTCTCAOGATTGTATGATGATGATTC
AATGCAAGGAGCATTGATAACGTTCATCATCTTTCCG TOCCOGAACTCAATCGAACCTCTCGATTOCT
GGATOG TOG TCTCGTGTTCTCCAGCTATTGGAAGTGTATCTATGATGATCTAGAAAT TCCCATCATGGG
TTGGGCAATCGT TCCAGCAAT TCCAAG TGATAT TGTTGATAGAAC TGA TCAAAATGCGTATOGATGCAAC
TACTCATATTGG TGGAGTTCAGATTAAAGCAGGAG TCGAAGCCGATTAAAGACGGAOCCAAT TGC TOCAOCCA
CCATCATACCACGAGTTG AACAGTG TCCGTGGATOG TGTGCTCAG AATCCACTTCTTCGAAATCCAATTG
TGCCAAGCACTAACTTCAAGCCAAATGCCACTACCGGG TGCTATGGAACACTATCAAAATGG TGAATAAC
TCCAATCAAT TGGCTATCAACA TCCAAC TCATC TOCACTGCTGGAAT TCAATCG TG TGG AAT TGTAGCT
GCACAGCATACTG TCGCTTCTTCATOGGCTCTTCCAATTGATT TGGAAAATCTGACACTTCCCGATCAGC
CACTGATGGATACTATGGATGTTGATGCATTGATCAGACATGAGCTGAG TCAAGCTGGAGGCCAGCATAT
TCATTTGATTTG TAA
```

Output format:   
Genetic code:

OR

第一种Verbose (“Met”, “Stop”, spaces between residues) ;

```
CNSSKSLHYIARSAEIFCKNSHRTC
INKSVRLLAE Stop PSILNFFAKNA
LQ Met EAILIRHRFLRISECPNRQTIP
WSR Met FHTSGIREIRTVQLDGRTRS
ILLRRLQRLNSKNLAAEPRRG Stop R
```

第二种Compact (“M”, “-”, no spaces);

```
CNSSKSLHYIARSAEIFCKNSHRTCKRGTWSSIRRYHL
LTMISSISTINKSVRLLAE-PSILNFFAKNARSLRI
VDSSAEHLQ TGRQCFILQMEAILIRHRFLRISECPNRQ
LNL SLQPLWL VQRNG -LLHKFTNGWSR MFHTSGIREIRT
FRMKEPERARGLLIQMQSQEGIHGVHNDPILLRRLQR
```

第三种Includes nucleotide sequence;

```
tcacactgcacaatctcgaaccggcaaggcctgattccggaatgag
S H C T I S N R Q G L I P E -
ttgacgatgatttcttcaatctcgacctcatcaacaagagcgttc
L T M I S S I S T S I N K S V
gagtaaccagatattctcaacaattcttcgcaaaaatgctcgtt
E - P S I L N N F F A K N A R
acacatctttagagactgttgacagcggaagaactagcctatacgg
T H L - R L L T A E E L A Y T
```

16种密码子表

# Translate - daf-16

## 5'3' Frame 1

Met QLEQKSSLHC SKCRN FLQKFSQD Met QAWNCRELDSPLPSDITLHNLEPARPDSG Met SFSTDFDDFFNLDLHQQERSASFSGVVTQYSQQFLREKCSFSPYFHTSLETVDSGRTSL  
YGSNEQCQQLGGASSNGSTAMet LHTPDGSGNSHQTSFSPDFR Met SESPDDTVSGKKTTRRNAWGN Met SYAELITTAI Met ASPEKRLTLAQVYEW Met VQNVVPYFRDKGDSNSSAGWK  
NSIRHNL SLHSRF Met RIQNEGAGKSSWWVINPDAKPGRNPRRTRERSNTIETTTKAQLEKSRRGAKKRIKERAL Met GSLHSTLNGNSIAGSIQTISHDLYDDDS Met QGAFDNVPSSFR  
PRTQSNLSIPGSSSRVSPAIGSDIYDDLEFP SWVGESVPAIPSDIVDRTDQ Met RIDATTHIGGVQIKQESKPIKTEPIAPPPSYHELNSVRGSCAQNP LLRNPIVPSTNFKP Met PLPGAY  
GNYQNGGITPINWLSTSNSSPLPGIQSCGIVAAQHTVASSSALPIDLENLTLPDQPL Met DT Met DVDALIRHEL SQAGGQHIHFDL Stop

## 5'3' Frame 2

CNSSKSLHYIARSAEIFCKNSHRTCKRGT VVSSIRRYHLTSHCTISNRQGLIPE Stop VFPLILT Met ISSISTSINKSVRLLLAE Stop PSILNFFAKNARSLRISTHL Stop RLLTAEELAYTGA  
Met SNVDSSAEHLQTGRQOCFILQ Met EAILIRHRFLRISECPNRQTIPYREKRQRPDGTLGEICH Met LNL SLQLPLWLVRNG Stop LLHKFTNGWSR Met FHTSGIREIRTVQLDGRTRSVT  
CLFILVSCEFR Met KEPERARGGLIQ Met QSQEGIHGVHVNDPILLRRLQRLNSKNLAAEPRRG Stop RREH Stop WAPFTRHL Met EIRLPDRFKRFLTC Met Met Met IQCKEHLITFHHLVFP  
ELNRTSRFLDRRLVFLQLLEVIS Met Met Stop NSHHGLANRFQQFQVILLIELIKCVS Met QLLILVEFRLSRSRRLRRNQLLHHHTTS Stop TVSVDRVLRHFFEIQLCQALTSSQCHYRV  
P Met ETIK Met VE Stop LQSIGYQHPTHLHCLEFNVEL Stop LHSILSLLHRLFQLIWKI Stop HFPISH Stop WILW Met L Met H Stop SD Met S Stop VKLEGSIFILIC

## 5'3' Frame 3

ATRAKVFITLLEVPKFSAKILTGHASVELS Stop ARFAATI Stop HHTAQSR TGKA Stop FRNEFFH Stop F Stop R Stop FLQSRPPSTRAF GFFWRSNPVVFSTISSRK Met LVLSVFP HIFRDC Stop  
QRKN Stop PIREQ Stop A Met WTARRSIFKRVD SNASY SRWKQFSSDIVSFGFQNVRIARRYRIGKKDNDQTERLGKYVIC Stop TYHSHYG Stop SRETVNSCTSLR Met DGPECSILQG Stop G  
RFEQFSW Met EELDPSQSVSSFSFHANSE Stop RSRKELVVG Y Stop SRCKARKESTAYT Stop TIQYY Stop DDYK GSTRKISPRSQEEDKGESIDGLPSLDT Stop WKFD CRIDSNDFSRFV  
Stop Stop Stop FNARSI Stop Stop RSIIFPSPNSIEPLDSWIVVSCFSSYWK Stop YL Stop Stop SRIP Met GWRIGSSNSK Stop YC Stop Stop N Stop SNAYRCNYSYWWSSD Stop AGVEAD Stop DGTN  
CSTTIIPRVEQCPWIVCSESTSSKSNCAKH Stop LQANATTGCLWKLSKWWNNSNQLAINIQLISTAWNSIVWNC SCTAYCRFFIGSSN Stop FGKSDTSRSATDGYG C Stop CIDQT Stop A  
ESSWRAAYSF Stop FV

## 3'5' Frame 1

LQIK Met N Met LPSSLTQL Met SDQCINIHSIHQW LIGKQCIFQINWKS R Stop RSDS Met LCSYNSTR LNSRQWR Stop VGC Stop Stop PIDWSYSTILIVSIGTR Stop WHWLEVS AWHNWISKKWI  
LSTRSTD TVQLVV Stop WWWSNWFRLNRL RLLLNLNSTN Met SSCIDTHLISSINNITWNCWNR FANP Stop WEF Stop IIIDITSNSWRNTRRRSRNREVRLSSGTER Stop WNVIKCSLH Stop I  
IIIQIVRNRLNRSNRSIKCRVKG AHQCSLLYPLLGSAAFFELSLCSRLNSIGSFTCTPWIPSWLCIWINNPPRALSGS FILNSHETR Met KRQIVTDRVLPSS Stop TVRISLIPEVWNILD  
HPFVNLCKS Stop PFLW TSHNGCSDKFSI Stop HISPSVPSGRCLFSRYGIVWRFGHSEIRRK RCL Met RIASIWS Met KHCCRPV Stop RCSAELSTLLIAPV Stop ASSAVNSL Stop RCVEIRR  
ERAFFAKLLRILGYS AKRSRTLL Met EVEIEEIIVKISGKTHSGIRPCRFEIVQCDVRW Stop RRIELTTV PRLHVL Stop EFLQKISALRA Met Stop Stop RLLLELH

## 3'5' Frame 2

YKSK Stop ICCPPA Stop LSSCLINASTSIVSISG Stop SGSVRF SKSIGRADEEATVCCAATIPHD Stop IPGSGDELVDVSQLIGVIPPFF Stop Stop FP Stop APGSGIGLKLVLGTIGFRRSGF Stop  
AHDPRTLFNSWYDGGGAIGSVLIGFDSCLI Stop TPPI Stop VVASIRI Stop SVLSTISLGIAGTDSPTHGNSRSS Stop ISLPIAGETRDDD PGIERFD Stop VRGRKDDGTLSNAPCIESSYK  
S Stop EIV Stop IDPAIEFPLSVE Stop REPINALS FILFAPRRDFSS Stop AFVVSIVLDRSRVRRGFLPGFASGLITHHELFPAPSF Stop IR Met KRE Stop RDRL Stop RIEFFHPAELFESPLSLK  
YGTFTIHS Stop TCARVNRFSGLAI Met AVVISSAYDIFPQAFRLVVVFPDTVSSGSDILKSE GNDV Stop Stop ELLPSGV Stop SIAVDPFEDAPPSCPHCSLLPYRLVPLSTVSKDVWK  
YGENEHFSRRNC Stop EYWVTPPKEAERSC Stop WRSRLKSSSSKSVEKLIPESGLAGSRLCSV Met SDGSGESSRQFHAC Met SCENFCRKRHFHQCNEDFCSSC

## 3'5' Frame 3

TNQNEYAALQLDSA HV Stop S Met HQHP Stop YPSVADREVSDFPNQLEEP Met KKRQYAVQLQFH TIEFQAVE Met SW Met LIAN Stop LELFHFD SFHRHPVVALA Stop S Stop CLAQLDFEEV  
DSEHTIHGHCSTRG Met Met VVEQLVPS Stop SASTPA Stop SELHQYE Stop LHRYAFDQFYQQYHLELLEPIRQP Met Met GILDHRYHFQ Stop LEKHETTIQESRGSIEFGDGK Met Met ERYQ  
Met LLALNHHHTNREKSFESIRQSNFH Stop VSSEGPS Met LSPLSSSWLRGEIFRVEPL Stop SSQ Stop YWIVHVYAVDSFLALHLD Stop Stop PTTSSFRLHSEFA Stop NENEETDCD GSS  
SIQLNCSNLPYP Stop S Met EHSGPSIRKLVQELTVSLD Stop P Stop WL Stop Stop Stop VQH Met TYFPKRSVWLSFFPIRYRLAIRTF Stop NPKET Met SDENCFHLEYEALLSTR LK Met LRAVH  
AHCSRIG Stop FFRCQQLK Met CGNTERTSIFREEIVENTGLLRQKKNALVDGGRD Stop RNHRQNQWKN SFRNQALPVRDCAV Stop CQ Met VAANRAHDSSTLACPVRIFAENFGTSS  
NV Met KTFARVA

# Translate - daf-16

```
M QLEQKSSLHCSKCRNFLQKFSQDM QAWNCRELDSPLPSDITLHNLEPARPDSG M SFSTDFDDDDFFNLDL
HQQERSASFGGVTQYSQQFLREKCSFSPYFHTSLETVDSGRTSGLYGSNEQCGQLGGASSNGSTA MLHTP
DGSNSHQTSFSPDFR M SE SPDDTVSGKKTTRRNAWGN M SYAELITTAI MASPEKRLTLAQVYEW M VQNV
PYFRDKGDSNSSAGWKNSIRHNLSLHSRF M RIQNEGAGKSSWWVINPDAKPGRNPRRTRERSNTIETTTK
AQLEKSRRGAKKRIKERAL M GSLHSTLNGNSIAGSIQTISHDLYDDDS M QGAFDNVPSSFRPRTQSNLSIP
GSSSRVSPAIGSDIYDDLEFP SWVGESVPAIPSDIVDRDQ M RIDATTHIGGVQIKQESKPIKTEPIAPPSY
HELNSVRGSCAQNPLLRNP IVPSTNFKP M PLPGAYGNYQNGGITPINWLSTSNSSPLPGIQSCGIVAQAHT
VASSALPIDLENLTL PDQPL MDT M DVDALIRHEL SQAGGQHIHFDL Stop
```

```
ID VIRT60794 Unreviewed; 541 AA.
AC VIRT60794;
DE Translation of nucleotide sequence generated on ExPASy
DE on 10-Jan-2018 by 162.105.250.50.
CC -!- This virtual protein sequence will automatically be deleted
CC from the server after a few days.
DR SWISS-2DPAGE; VIRT60794; VIRTUAL.
SQ SEQUENCE 541 AA; 8A36A86311A32CBB CRC64.
MQLEQKSSLH CSKCRNFLQK FSQDMQAWNC RELDSPLPSD ITLHNLEPAR PDSGMSFSTD
FDDDDFFNLDL HQQERSASFG GVTQYSQQFL REKCSFSPYF HTSLETVDSG RTSGLYGSNEQ
CGQLGGASSN GSTAMLHTPD GNSNSHQTSFP SDFRMSESPD DTVSGKKTTR RNAWGNMSY
AELITTAIMA SPEKRLTLAQ VYEWVQNV YFRDKGDSNS SAGWKNSIRH NLSLHSRFMR
IQNEGAGKSS WWWINPDAKP GRNPRRTRER SNTIETTTKA QLEKSRRGAK KRIKERALMG
SLHSTLNGNS IAGSIQTISH DLYDDDSMQG AFDNVPSSFR PRTQSNLSIP GSSSRVSPA
GSDIYDDLEF P SWVGESVPA IPSDIVDRD QMRIDATTHI GGVQIKQESK PIKTEPIAPP
PSYHELNSVR GSCAQNPLLR NP IVPSTNFKP M PLPGAYGN YQNGGITPIN WLSTSNSSPL
PGIQSCGIVA AQHTVASSA LP IDLENLTL PDQPLMDTMD DVALIRHEL SQAGGQHIHFD
L
```

//

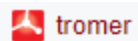
Sequence in FASTA format

## Sequence in FASTA format

>VIRT60794

```
MQLEQKSSLHCSKCRNFLQKFSQDMQAWNCRELDSPLPSDITLHNLEPARPDSGMSFSTD
FDDDDFFNLDLHQQERSASFGGVTQYSQQFLREKCSFSPYFHTSLETVDSGRTSGLYGSNEQ
CGQLGGASSNGSTAMLHTPDGNSNSHQTSFP SDFRMSESPDDTVSGKKTTRRNAWGNMSY
AELITTAIMASPEKRLTLAQVYEWVQNV YFRDKGDSNSSAGWKNSIRHNLSLHSRFMR
IQNEGAGKSSWWVINPDAKPGRNPRRTRERSNTIETTTKAQLEKSRRGAKKRIKERALMG
SLHSTLNGNSIAGSIQTISHDLYDDDSMQGAFDNVPSSFRPRTQSNLSIPGSSSRVSPA
GSDIYDDLEFP SWVGESVPAIPSDIVDRDQMRIDATTHIGGVQIKQESKPIKTEPIAPP
PSYHELNSVRGSCAQNPLLRNP IVPSTNFKP M PLPGAYGNYQNGGITPINWLSTSNSSPL
PGIQSCGIVAQAHTVASSALPIDLENLTL PDQPLMDTMDVDALIRHEL SQAGGQHIHFD
L
```

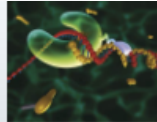
# TromER 数据库



tromer

The transcriptome analyser project aims to provide tools to determine and document all the transcribed parts of a genome. The transcribed parts are defined by analysing experimental evidence, like expressed sequence tags (EST) and other mRNA sequences. [\[less\]](#)

Keywords: [expressed sequence tag](#)

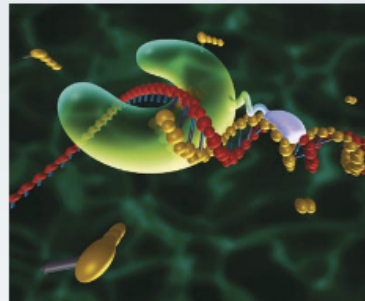


## TromER



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

Computational Cancer Genomics | ExPASy | EPFL



### TromER Database

#### Transcriptome Analysis Database

**Description:** The transcriptome analyser (**TromER**) project aims at providing powerful tools to determine and document all the transcribed elements of a genome. The transcribed parts are defined by analysing experimental evidence, like expressed sequence tags (EST) and other mRNA sequences. The transcriptome data include transcribed regions, splice variants, SAGE mappings, MPSS mappings, Affymetrix chips mappings, etc., for several organisms. The [Tromer Web Service](#) provides a user-friendly web interface to gene predictions, or transcriptome units of several genomes (such as human, mouse, rat, drosophila and worm).

#### Access to the TromER Database

Genome

H. sapiens

Assembly

Feb. 2009 (GRCh37/hg19)

Quick Search

in **TROMER DB** by TROMER ID, SwissProt ID  
NCBI-GENE ID or GENE name

- [Retrieve TROMER Graphs](#)
- [View/Download tag mappings: \[MPSS\]](#)
- [local FTP site for software download](#)

Documentation



Daf-16 Gene ID: 172981

# TromER – daf-16

Genome:  Assembly:

in **TROMER DB** by TROMER ID, SwissProt ID  
NCBI-GENE ID or GENE name

**TROMER Entry : WTr022738**  
[\[General TROMER Report\]](#) [\[List of Transcript Sequences\]](#)

**TROMER Transcriptional Unit Report**

TROMER ID	WTr022738
ASSEMBLY	ce4
Description	Caenorhabditis elegans abnormal DAuer Formation family member (daf-16) (daf-16) partial mRNA.
NCBI AceView WormGenes	<a href="#">WBGene00000912</a>
NCBI Entrez Gene ID	<a href="#">172981</a>
Visualisation	<a href="#">Graph View - NC_003279_2105</a> <a href="#">TROMER-Predicted Transcript Variants</a>

List of Transcript Sequences in GenBank  
A reduced set has been chosen for display (top five for each transcript class)

Total nb. of sequences	63
SEQUENCE	ACC= <a href="#">AU207803</a> ; DB=E; CLONE=yk1006c10; END=5'; LID=24032; LEN=533
SEQUENCE	ACC= <a href="#">AU207952</a> ; DB=E; CLONE=yk1008b03; END=5'; LID=24032; LEN=534
SEQUENCE	ACC= <a href="#">AU221147</a> ; DB=E; CLONE=yk1006c10; END=3'; LID=24032; LEN=645
SEQUENCE	ACC= <a href="#">AU221301</a> ; DB=E; CLONE=yk1008b03; END=3'; LID=24032; LEN=460
SEQUENCE	ACC= <a href="#">AV190520</a> ; DB=E; CLONE=yk572a12; END=5'; LID=6040; LEN=359
SEQUENCE	ACC= <a href="#">AF020342</a> ; DB=M; LEN=2579
SEQUENCE	ACC= <a href="#">AF020343</a> ; DB=M; LEN=2565
SEQUENCE	ACC= <a href="#">AF020344</a> ; DB=M; LEN=2345
SEQUENCE	ACC= <a href="#">AF032112</a> ; DB=M; LEN=3035

# TromER – daf-16

Graph accession number : [NC\\_003279\\_2105](#)

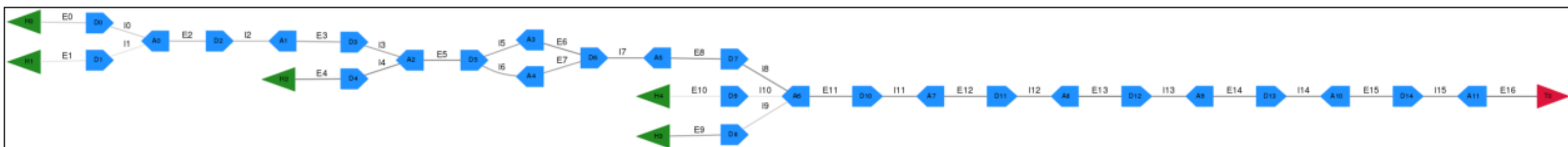
Locus : [NC\\_003279.5](#) Chromosome : [I](#)

Start : **10750557** Stop : **10776690**

Nbr. of exons : **17** Nbr. of introns : **16**

Size : **26133 bp's** Species : **C. elegans (6239)**

Viewers : [UCSC](#) [ENSEMBL](#)



**Legend and color code for Exon/Intron experimental evidence**

**Warning:** This is a graphical display of a gene frameset corresponding to a tromer-predicted splicing graph. Graph elements are clickable and they are not drawn to scale!

Exon / Intron : 1 EST 2 EST's 3 - 5 EST's 5 - 10 EST's 10 - 20 EST's 20 - 50 EST's 50 - 100 EST's 100 - 500 EST's more than 500 EST's

Nodes : H = Start A = Acceptor D = Donor T = Stop

# TromER – daf-16

---

## GRAPH INFO

Graph accession number : **NC\_003279\_2105**      Contig : **NC\_003279.5**      Chromosome Start : **:1**      Stop : **10750557**      **10776690**

---

## EXON / INTRON : E0

Position (start - stop) : **10750557 -- 10750642** Size : **86 bp's** Nbr. EST : **2** Rest : ---

---

## REFERENCES :

<b>1</b>	EST - EMBL	<a href="#">EC034273</a>	5..90	(10750557..10750642) 100%
<b>2</b>	RefSeq	<a href="#">NM_001026425</a>	1..86	(10750557..10750642) 100%

# TromER – daf-16













ID EC034273; SV 1; linear; mRNA; EST; INV; 505 BP.  
XX  
AC EC034273;  
XX  
DT 28-MAY-2006 (Rel. 87, Created)  
DT 28-MAY-2006 (Rel. 87, Last updated, Version 1)  
XX  
DE 3857413 KZ41 *Caenorhabditis elegans* cDNA clone 1781086, mRNA sequence.  
XX  
KW EST.  
XX  
OS *Caenorhabditis elegans*  
OC Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida;  
OC Rhabditoidea; Rhabditidae; Peloderinae; *Caenorhabditis*.  
XX  
RN [1]  
RP 1-505  
RA Zeng K., Garrick B., Chen F., Gnirke A., Swan K., Bjerke L., Baille M.,  
RA Cheung L., Chong A., Goldschmidt S., Hussain S., Laufer A., Oliva J.,  
RA Murray L., Park C., Reyes J., Wong M., Milash B., Amundsen C., Orton A.,  
RA Shao A., Platt D., Swimmer C., Nicoll M.;  
RT "Exelixis *Caenorhabditis elegans* EST project";  
RL Unpublished  
XX  
DR MD5: 98c9df66f5482512cb87a959a368be5e.  
DR UNILIB; 42560; 19760.  
XX  
CC Contact: Karin Schmitt  
CC Open Biosystems, Inc.  
CC 6705 Odyssey Drive, Huntsville, AL 35806, USA  
CC Email: KSCHMITT@OPENBIOSYSTEMS.COM.  
XX  
FH Key Location/Qualifiers  
FH

```
rn
FT source 1..505
FT /organism="Caenorhabditis elegans"
FT /mol_type="mRNA"
FT /clone_lib="KZ41"
FT /clone="1781086"
FT /db_xref="taxon:6239"
FT /db_xref="UNILIB:42560"
XX
SQ Sequence 505 BP; 140 A; 129 C; 109 G; 127 T; 0 other;
   ggggagcaaa agtcttcatt acattgctcg aagtgcgaa atttctgca aaaattctca 60
   caggacatgc aagcgtggaa ctgtcgtgag ctcgattcgc cgctaccatc tgacatcaca 120
   ctgcacaatc tcgaaccggc aaggcctgat tccggaatga gttttccac tgattttgac 180
   gatgatttct tcaatctcga cctccatcaa caagagcgtt cggcttcttt tggcggagta 240
   acccagtatt ctcaacaatt tcttcgcgaa aaatgctcgt tctctccgta tttccacaca 300
   tctttagaga ctgttgacag oggaagaact agcctatacg ggagcaatga gcaatgtgga 360
   cagctoggcg gagcatcttc aaacgggtcg acagcaatgc ttcatactcc agatggaagc 420
   aattctcatc agacatcgtt tcttcggat ttcagaatgt ccgaatcgcc agacgatacc 480
   gtatcgggaa aaaagacaac gacca 505
```

//

# ExPASy Structure Analysis

**Tools**

-  [SWISS-MODEL Workspace](#) • structure homology-modeling • [\[more\]](#)
-  [SwissDock](#) • protein ligand docking server • [\[more\]](#)
  
-  [CAMEO](#) • Continuous Automated Model EvaluatiOn • [\[more\]](#)
-  [Click2Drug](#) • Directory of computational drug design tools • [\[more\]](#)
-  [COILS](#) • Prediction of Coiled Coil Regions in Proteins • [\[more\]](#)
-  [MARCOIL](#) • coiled-coils prediction • [\[more\]](#)
-  [OpenStructure](#) • molecular modelling and visualization • [\[more\]](#)
-  [Protein Model Portal](#) • structural information for a protein • [\[more\]](#)
-  [QMEAN](#) • estimate quality of protein models • [\[more\]](#)
-  [Swiss-PdbViewer](#) • analyse protein 3D structures • [\[more\]](#)
-  [SwissParam](#) • topology, parameters for small organic molecules • [\[more\]](#)
-  [SwissTargetPrediction](#) • Target prediction for bioactive small molecules • [\[more\]](#)

- 同源建模工具

Swiss-model、open structure

- 结果评估工具

CAMEO、QMEAN

- 核心结构域预测

COILS、MARCOIL

- 结构可视化工具

SWISS-PDB VIEWER

- 药物设计工具

Click2Drug、Swiss TargetPrediction

# Swiss-Model

## Search for templates

Paste protein sequence  
from uniprot

Start a New Modelling Project ⓘ

Target Sequence:  
*(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)*

Target `MOLEQKSSLHCCKRNFLQKFSQDMQAWNCRELDSPLPSDITLHNLEPARPDSGMSFSTDFDDFFNLDLHQQERSASFGGVTQYSQQFLREKCSFSPYFHTSLETVDSGRTSLYGSNEQCGQLG` 125

Target `GASSNGSTAMLHTPDGNSHQTSFSPDFRMSESPDDTVSGKKTTRRNNAWGNMSYAEELITTAIMASPEKRLTLAQVYEWVQNVVPYFRDKGDSNSSAGWKN SIRHNL SLHSR FMRIQNEGAGKSS` 250

Target `WVWV INPDAK PGRNPRRTRERSNTIETTTKAQLEKSRRAKRIKERALMGSLHSTLNGNSIAGSIQTI SHDLYDDDSMQGAFDNPSSFRPRTQSNLSIPGSSSRVSPAIGSDIYDDLEFPSSWVG` 375

Target `ESVPAIPSDIVDRTDQMRIDATTHIGGVQIKQESKPIKTEPIAPPPSYHELNSVRGSCAQNPLLRNPIVPSTNFKEMPLPGAYGNYQNGGITPINWLSTSNSSPLPGIQSCGIVAAQHTVASSSA` 500

Project Title:

Email:

By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding [articles](#).

# Swiss-Model

Select an optimal template to build model

Templates   Quaternary Structure   Sequence Similarity   Alignment of Selected Templates   More ▾

Name	Title	Coverage	Identity	Method	Oligo State	Ligands
<input checked="" type="checkbox"/> 1e17.1.A	AFX	<input type="text" value=""/>	57.69	NMR	monomer ✓	None

**Method** SOLUTION NMR

**Found By** HHblits

**GMQE** 0.09

**Seq Similarity** 0.48

**QSQE** -

**Biounit Oligo State** monomer

**Target Prediction** It is only possible to build a monomer.

**Build Model**

Target **M**OLEFOKSSLHCCKCRNPLQKFSQDMQAWNCRELDSPLFSDITLHNLEPARPDSGMSFSTDFDDDFNLDLHQERSASFGGVTOYSQQFLREKCSFSPYFHTSLE 105  
1e17.1.A -----

Target **T**VDSGRTSLYGSNEQCQQLGGASSNGSTAMLHTPDGNSHQTSFPSPDFRMSESPDDTVSGKTTTRRNAWGNMSYAEIITAIMASPEKRLTLAQVYEWVQNVF 210  
1e17.1.A ----- **V**PRGSHMIEDPCAVTGPRKGGRRNAWGNQSYAELISQAT**S**PEKRLTLAQVYEWV**R**Q**V**E 75

Target **Y**FRDKGDSNSSAGWKN SIRHNL SLHSRPMRIQNEGAGKSSWWVINPDARPGNPRTRERSNTIETTTKAQLEKSRGAKKRIKERALMGSLHSTLNGNSIAGSI 315  
1e17.1.A **Y**FRDKGDSNSSAGWKN SIRHNL SLHSR**F**IK**V**NEATGKSS**W**LN**P**EGGKSGK**A**PR -- **R**RAASMDSS**S**K**I** ----- 143

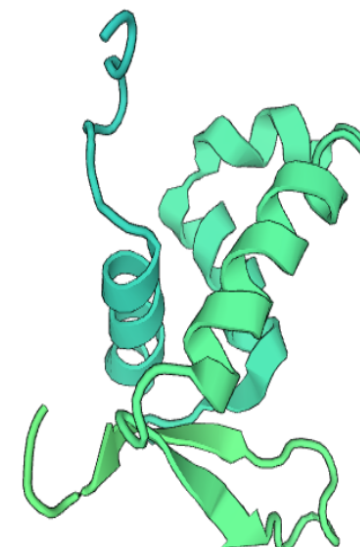
Target **Q**TISHDLYDDDSMQGAFDNPVPSFRPRTQSNLSIPGSSSRVSPAIGSDIYDDLEFPSWVGESVPAIPSDIVDRDQMRIDATTHIGGVQIKQESKPIKTEPIAPP 420  
1e17.1.A -----

Target **P**SYHELNSVRGSCAQNPLLRNPIVPSTNFKPMPLPGAYGNYQNGGITPINWLSTSNSSPLPGIQSCGIVAAQHTVASSSALPIDLENLTLDPDQLMDTMDVDALI 525  
1e17.1.A -----

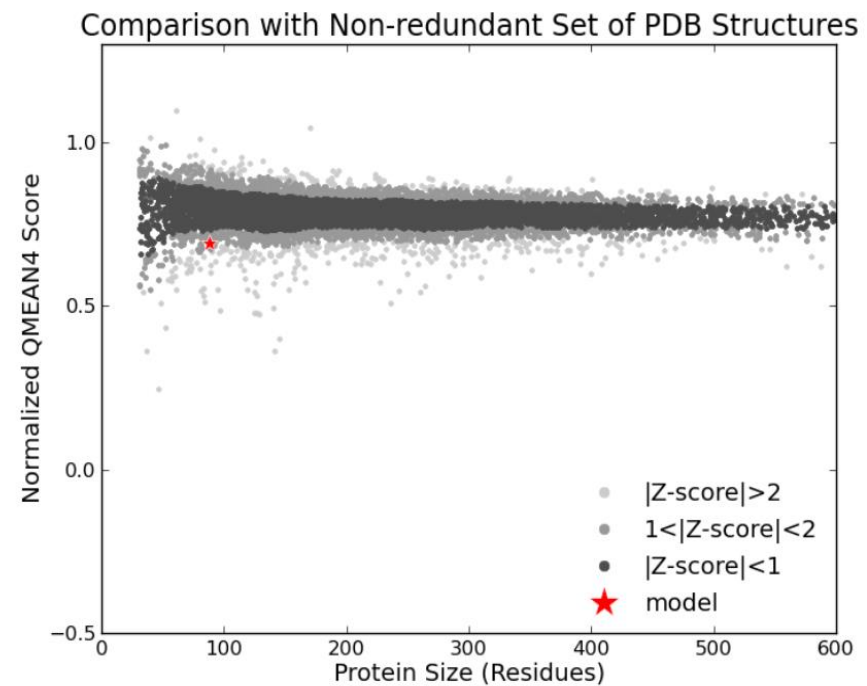
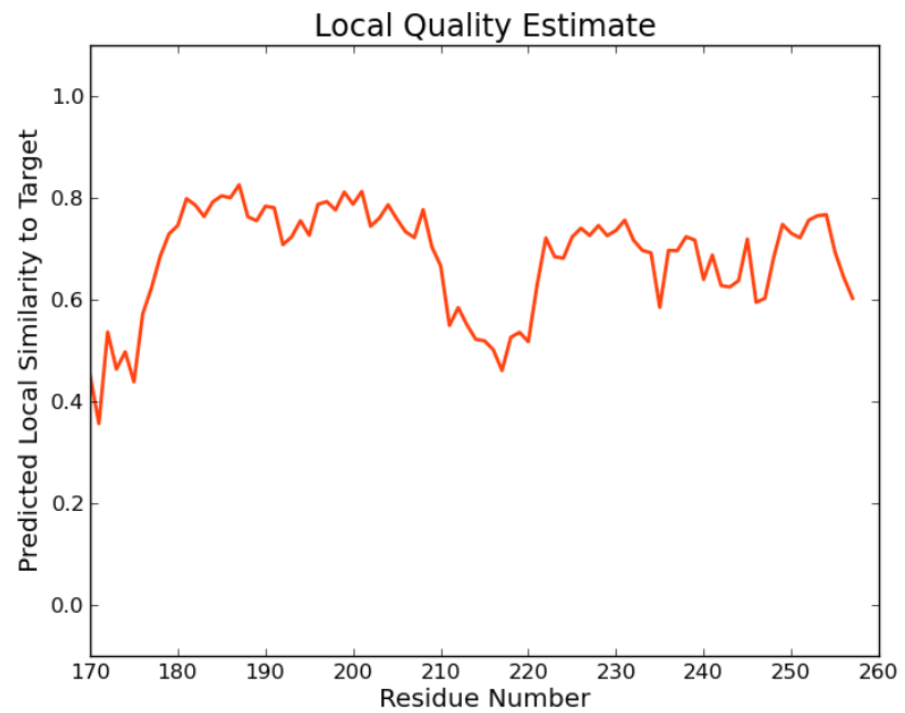
Target **R**HELSQLGGQHIHFDL 541  
1e17.1.A -----

**Build Models 1**

Clear Selection

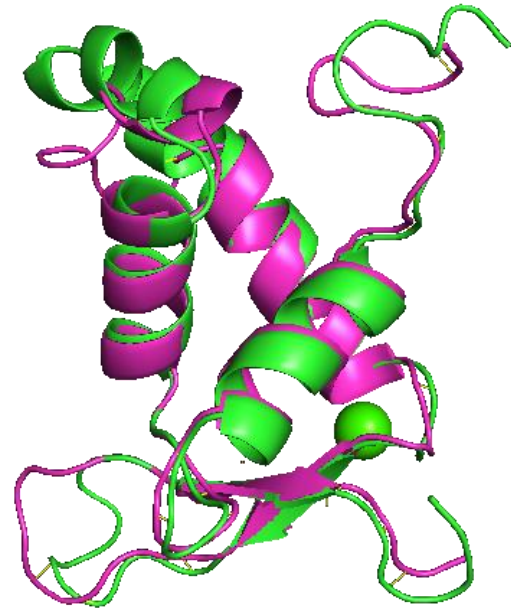


# Swiss-Model Results





# Pymol-align Evaluate two results differences



## Templates

█ FOXO4

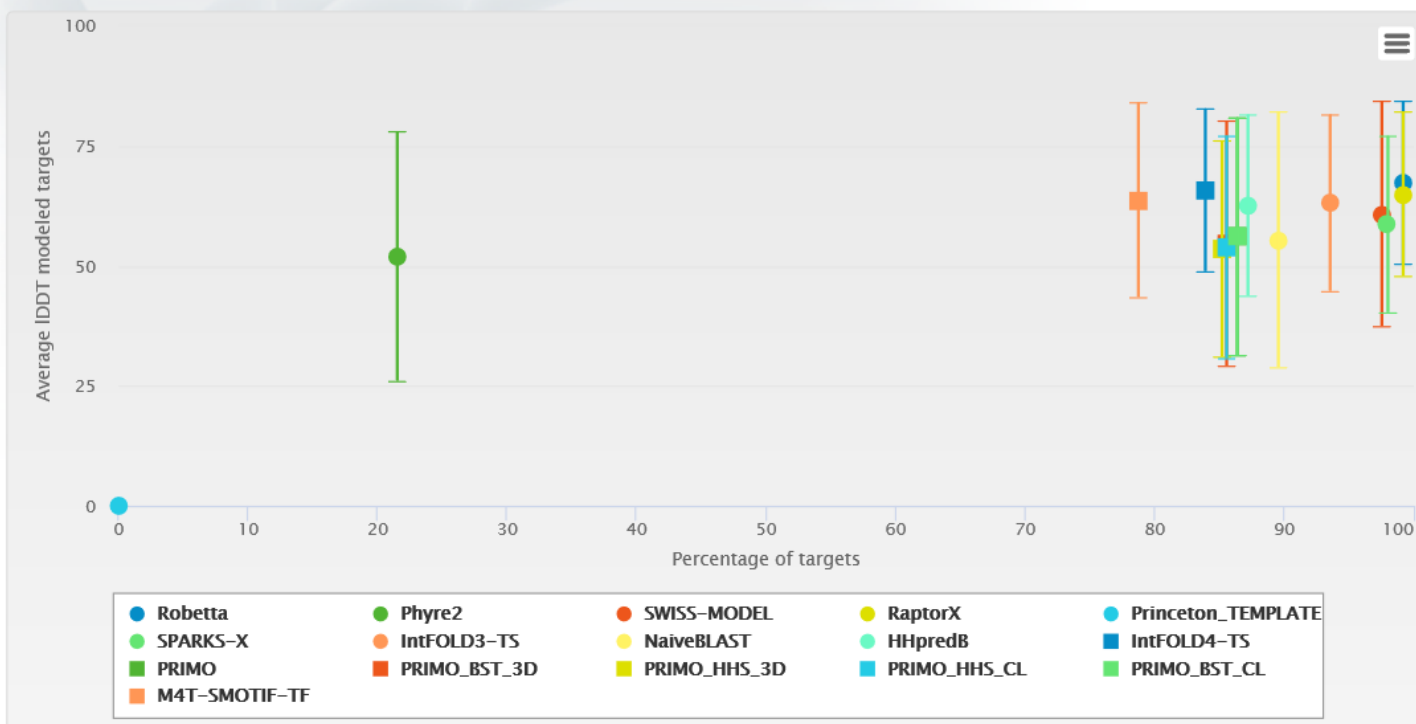
█ FOXO1-ETS1 DNA complex

```
ExecutiveAlign: 87 atoms aligned.  
ExecutiveRMS: 6 atoms rejected during cycle 1 (RMS=2.45).  
ExecutiveRMS: 5 atoms rejected during cycle 2 (RMS=1.65).  
ExecutiveRMS: 6 atoms rejected during cycle 3 (RMS=1.25).  
ExecutiveRMS: 3 atoms rejected during cycle 4 (RMS=0.98).  
ExecutiveRMS: 1 atoms rejected during cycle 5 (RMS=0.90).  
Executive: RMS = 0.874 (66 to 66 atoms)  
Executive: object "aln_model_01_to_model_06" created.
```

# CAMEO-3D: Protein Structure 3-months Performance (2017-10-13 - 2018-01-06)

1 Year 6 Months **3 Months** 1 Month 1 Week

All targets Easy Medium Hard ⓘ



## Dataset composition

	All Targets	Easy	Medium	Hard
Monomer	111	18	54	39
Oligomer (as part of hetero)	139 (54)	24 (7)	69 (25)	46 (22)
containing ligand(s)	94	18	46	30
<b>Total</b>	<b>250</b>	<b>42</b>	<b>123</b>	<b>85</b>

Total Number of Models: 3195

# 谢谢！

---

小组成员：

- 毕莹 1701110438
- 戴睿成 1701110442
- 韩旭 1701111467
- 蒋家浩 1701110480
- 李伯源 1701110448
- 王卓青 1701110558
- 吴美莹 1701110486
- 杨婷玉 1701111494