“实用生物信息技术”课程小组讨论总结报告

组：G2　　次：1　　组长：边汉青　　执笔：刘奇

1. 时间

2026.4.8 上午

2. 方式

线上

3. 主题

关于Uniprot数据库的高级检索和序列条目注释信息以及外部链接的阐述和讨论

4. 内容

A Uniprot数据库的高级检索功能和序列条目注释信息

A1 Uniprot的高级检索

A2 Uniprot的序列条目注释信息

A3 Uniprot的多序列比对

B Uniprot数据库的外部链接

C Uniprot文献信息

D PlantTFDB序列注释信息

E 个人总结

A Uniprot 数据库的高级检索功能和序列条目注释信息

A1 Uniprot 的高级检索

以果蝇中的 *sxl* 基因为例进行检索，检索过程和结果如下：

Searching in: UniProtKB

Organism [OS]: 7227

AND

Gene Name [GN]: sxl

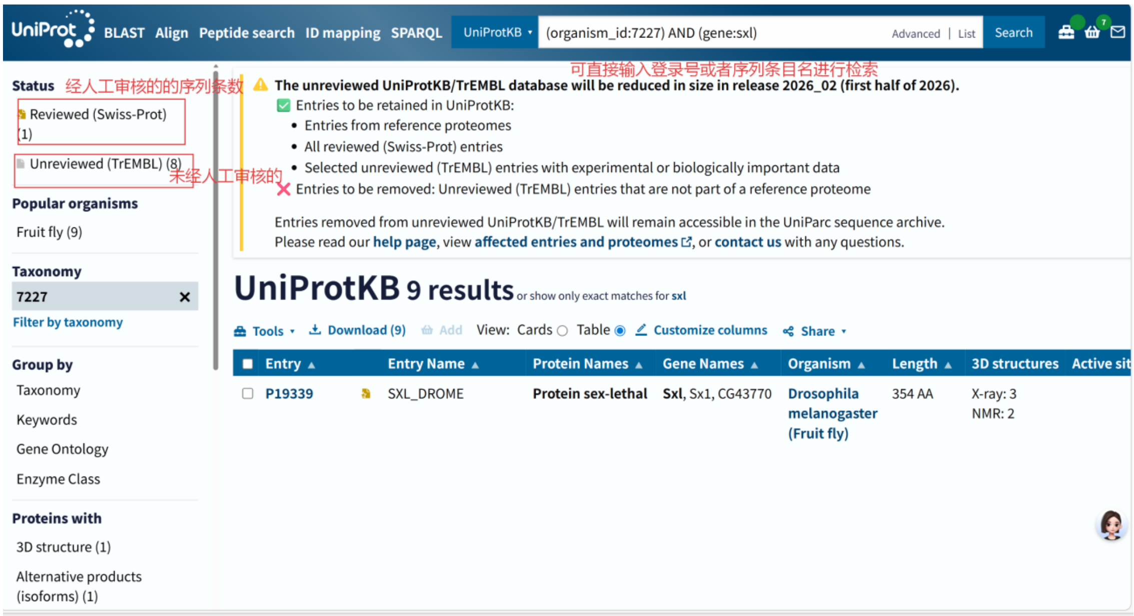
Type * in the search box to search for all values for the selected field.

可选择信息类别

选择多个检索信息之间是否并列

添加检索信息

检索



(organism_id:7227) AND (gene:sxl)

可直接输入登录号或者序列条目名进行检索

Status 经人工审核的的序列条数

Reviewed (Swiss-Prot) (1)

Unreviewed (TrEMBL) (8) 未经人工审核的

Popular organisms

Fruit fly (9)

Taxonomy

7227

Group by

Taxonomy

Keywords

Gene Ontology

Enzyme Class

Proteins with

3D structure (1)

Alternative products (isoforms) (1)

The unreviewed UniProtKB/TrEMBL database will be reduced in size in release 2026_02 (first half of 2026).

☑ Entries to be retained in UniProtKB:

- Entries from reference proteomes
- All reviewed (Swiss-Prot) entries
- Selected unreviewed (TrEMBL) entries with experimental or biologically important data

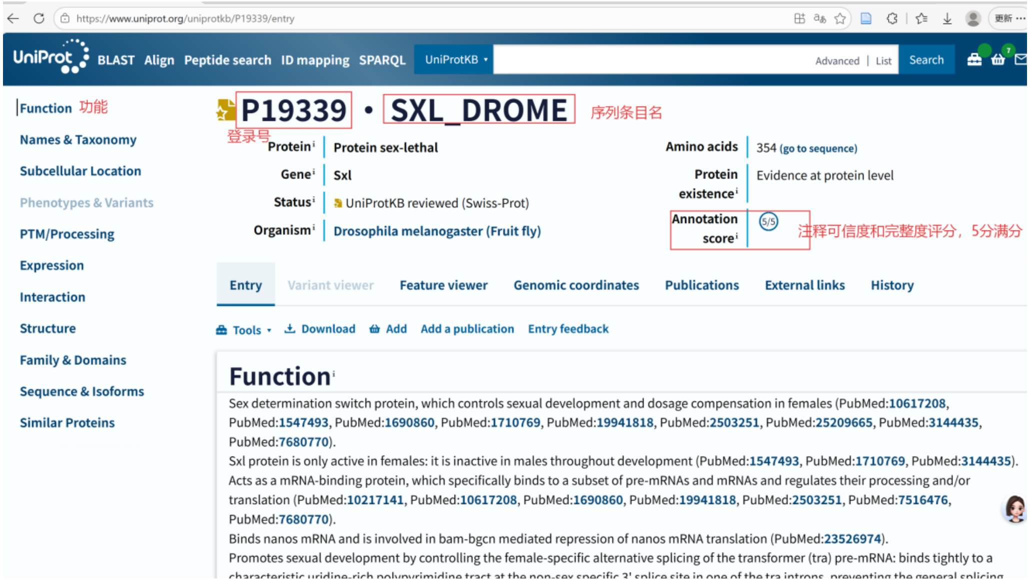
✗ Entries to be removed: Unreviewed (TrEMBL) entries that are not part of a reference proteome

Entries removed from unreviewed UniProtKB/TrEMBL will remain accessible in the UniParc sequence archive. Please read our **help page**, view **affected entries and proteomes**, or **contact us** with any questions.

UniProtKB 9 results or show only exact matches for **sxl**

☐	Entry		Entry Name	Protein Names	Gene Names	Organism	Length	3D structures	Active sit
☐	P19339		SXL_DROME	Protein sex-lethal	**Sxl**, Sx1, CG43770	Drosophila melanogaster (Fruit fly)	354 AA	X-ray: 3 NMR: 2	

A2 Uniprot 的序列条目注释信息（以 Sxl 蛋白为例）



功能

P19339 • **SXL_DROME** 序列条目名

登录号

Protein	Protein sex-lethal	Amino acids	354 (go to sequence)
Gene	Sxl	Protein existence	Evidence at protein level
Status	UniProtKB reviewed (Swiss-Prot)	Annotation score	5/5
Organism	Drosophila melanogaster (Fruit fly)		

注释可信度和完整度评分，5分满分

Function

Sex determination switch protein, which controls sexual development and dosage compensation in females (PubMed:10617208, PubMed:1547493, PubMed:1690860, PubMed:1710769, PubMed:19941818, PubMed:2503251, PubMed:25209665, PubMed:3144435, PubMed:7680770).

Sxl protein is only active in females: it is inactive in males throughout development (PubMed:1547493, PubMed:1710769, PubMed:3144435).

Acts as a mRNA-binding protein, which specifically binds to a subset of pre-mRNAs and mRNAs and regulates their processing and/or translation (PubMed:10217141, PubMed:10617208, PubMed:1690860, PubMed:19941818, PubMed:2503251, PubMed:7516476, PubMed:7680770).

Binds nanos mRNA and is involved in bam-bgcn mediated repression of nanos mRNA translation (PubMed:23526974).

Promotes sexual development by controlling the female-specific alternative splicing of the transformer (tra) pre-mRNA: binds tightly to a characteristic uridine-rich polypyrimidine tract at the non-sex specific 3' splice site in one of the tra introns, preventing the general splicing

亚细胞定位

表型与变异位点

Names & Taxonomy

名称和分类

Protein names

Recommended name	Protein sex-lethal 1 Publication

Gene names

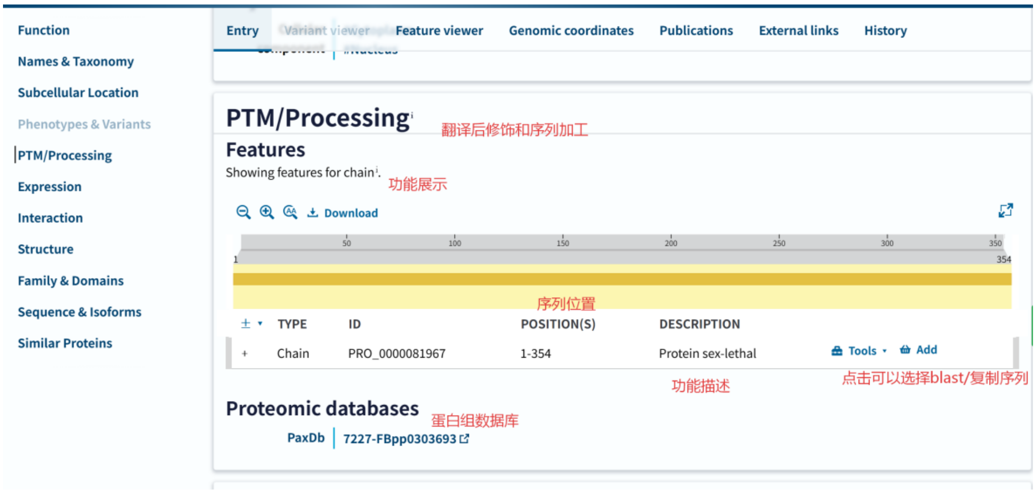
Name	Sxl 1 Publication Imported
Synonyms	Sx1
ORF names	CG43770 Imported

Organism names

Taxonomic identifier	7227 (NCBI)
Organism	Drosophila melanogaster (Fruit fly) Imported
Strains	Oregon-R Berkeley
Taxonomic lineage	cellular organisms > Eukaryota (eukaryotes) > Opisthokonta > Metazoa (animals) > Eumetazoa > Bilateria > Protostomia > Ecdysozoa > Panarthropoda > Arthropoda (arthropods) > Mandibulata > Pancrustacea > Hexapoda (hexapods) > Insecta (insects) > Dicondylia > Pterygota (winged insects) > Neoptera > Endopterygota > Diptera (flies) > Brachycera > Muscomorpha > Eremoneura > Cyclorrhapha > Schizophora > Acalyptratae > Ephydroidea >

可跳转到NCBI相关信息

分类谱系



PTM/Processing

翻译后修饰和序列加工

Features

Showing features for chain.

功能展示

序列位置

TYPE	ID	POSITION(S)	DESCRIPTION
Chain	PRO_0000081967	1-354	Protein sex-lethal

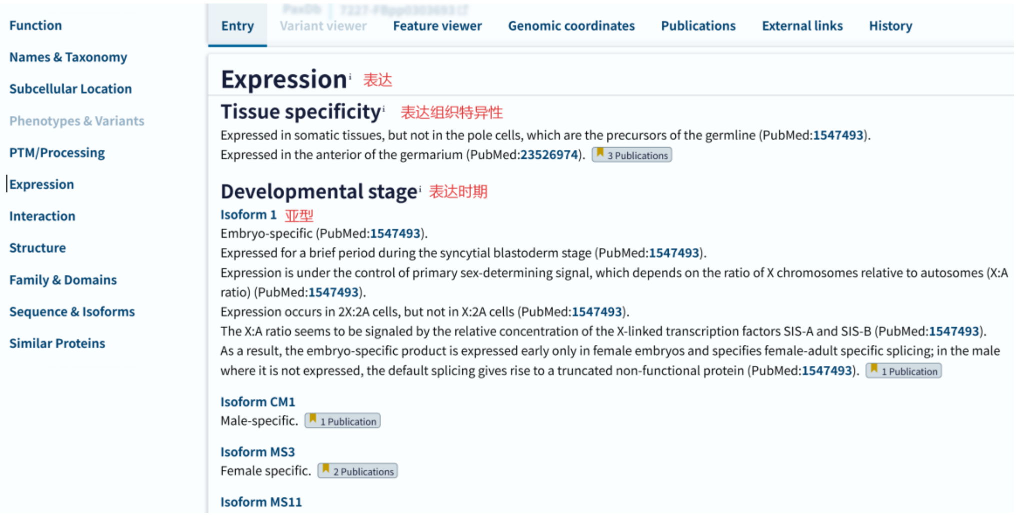
功能描述

点击可以选择blast/复制序列

Proteomic databases

蛋白组数据库

PaxDb	7227-FBpp0303693



Expression 表达

Tissue specificity 表达组织特异性

Expressed in somatic tissues, but not in the pole cells, which are the precursors of the germline (PubMed:**1547493**).
Expressed in the anterior of the germarium (PubMed:**23526974**). 3 Publications

Developmental stage 表达时期

Isoform 1 亚型

Embryo-specific (PubMed:**1547493**).
Expressed for a brief period during the syncytial blastoderm stage (PubMed:**1547493**).
Expression is under the control of primary sex-determining signal, which depends on the ratio of X chromosomes relative to autosomes (X:A ratio) (PubMed:**1547493**).
Expression occurs in 2X:2A cells, but not in X:2A cells (PubMed:**1547493**).
The X:A ratio seems to be signaled by the relative concentration of the X-linked transcription factors SIS-A and SIS-B (PubMed:**1547493**).
As a result, the embryo-specific product is expressed early only in female embryos and specifies female-adult specific splicing; in the male where it is not expressed, the default splicing gives rise to a truncated non-functional protein (PubMed:**1547493**). 1 Publication

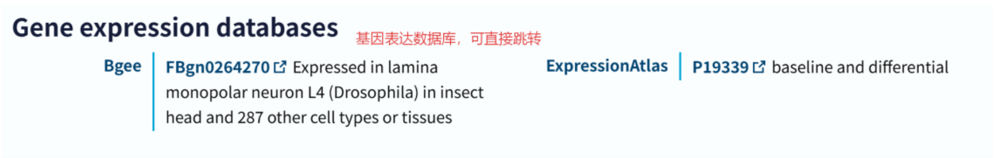
Isoform CM1

Male-specific. 1 Publication

Isoform MS3

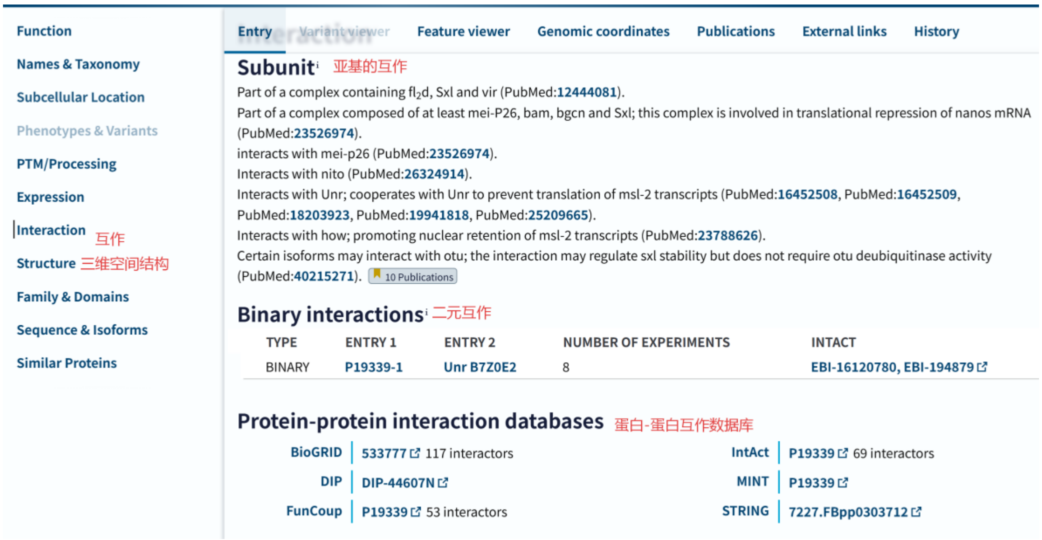
Female specific. 2 Publications

Isoform MS11



Gene expression databases 基因表达数据库，可直接跳转

Bgee	**FBgn0264270** Expressed in lamina monopolar neuron L4 (Drosophila) in insect head and 287 other cell types or tissues	**ExpressionAtlas**	**P19339** baseline and differential



互作

Structure 三维空间结构

Subunit 亚基的互作

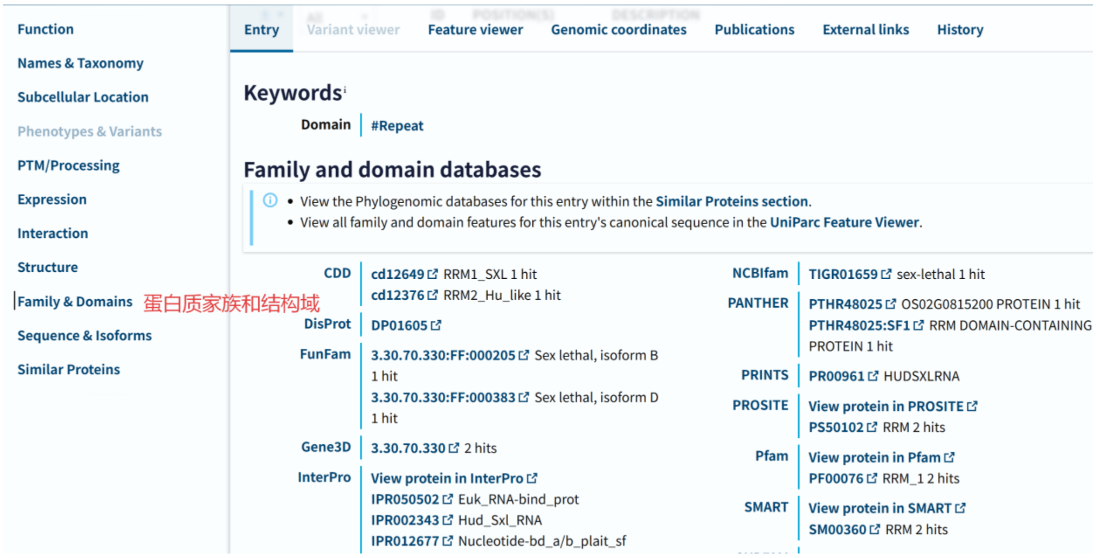
Part of a complex containing fl_2d, Sxl and vir (PubMed:**12444081**).
Part of a complex composed of at least mei-P26, bam, bgcn and Sxl; this complex is involved in translational repression of nanos mRNA (PubMed:**23526974**).
interacts with mei-p26 (PubMed:**23526974**).
Interacts with nito (PubMed:**26324914**).
Interacts with Unr; cooperates with Unr to prevent translation of msl-2 transcripts (PubMed:**16452508**, PubMed:**16452509**, PubMed:**18203923**, PubMed:**19941818**, PubMed:**25209665**).
Interacts with how; promoting nuclear retention of msl-2 transcripts (PubMed:**23788626**).
Certain isoforms may interact with otu; the interaction may regulate sxl stability but does not require otu deubiquitinase activity (PubMed:**40215271**). 10 Publications

Binary interactions 二元互作

TYPE	ENTRY 1	ENTRY 2	NUMBER OF EXPERIMENTS	INTACT
BINARY	P19339-1	Unr B7Z0E2	8	EBI-16120780, EBI-194879

Protein-protein interaction databases 蛋白-蛋白互作数据库

BioGRID	533777 117 interactors	**IntAct**	P19339 69 interactors
DIP	DIP-44607N	**MINT**	P19339
FunCoup	P19339 53 interactors	**STRING**	7227.FBpp0303712



蛋白质家族和结构域

Keywords

Domain | #Repeat

Family and domain databases

- View the Phylogenomic databases for this entry within the **Similar Proteins section**.
- View all family and domain features for this entry's canonical sequence in the **UniParc Feature Viewer**.

CDD	cd12649 RRM1_SXL 1 hit cd12376 RRM2_Hu_like 1 hit	**NCBIfam**	TIGR01659 sex-lethal 1 hit
DisProt	DP01605	**PANTHER**	PTHR48025 OS02G0815200 PROTEIN 1 hit PTHR48025:SF1 RRM DOMAIN-CONTAINING PROTEIN 1 hit
FunFam	3.30.70.330:FF:000205 Sex lethal, isoform B 1 hit 3.30.70.330:FF:000383 Sex lethal, isoform D 1 hit	**PRINTS**	PR00961 HUDSXLRNA
Gene3D	3.30.70.330 2 hits	**PROSITE**	View protein in PROSITE PS50102 RRM 2 hits
InterPro	View protein in InterPro IPR050502 Euk_RNA-bind_prot IPR002343 Hud_Sxl_RNA IPR012677 Nucleotide-bd_a/b_plait_sf	**Pfam**	View protein in Pfam PF00076 RRM_1 2 hits
		SMART	View protein in SMART SM00360 RRM 2 hits

A3 Uniprot 的多序列比对

果蝇的性别致死 Sxl 蛋白目前没有相似序列，这里以 PPF-1 为例。

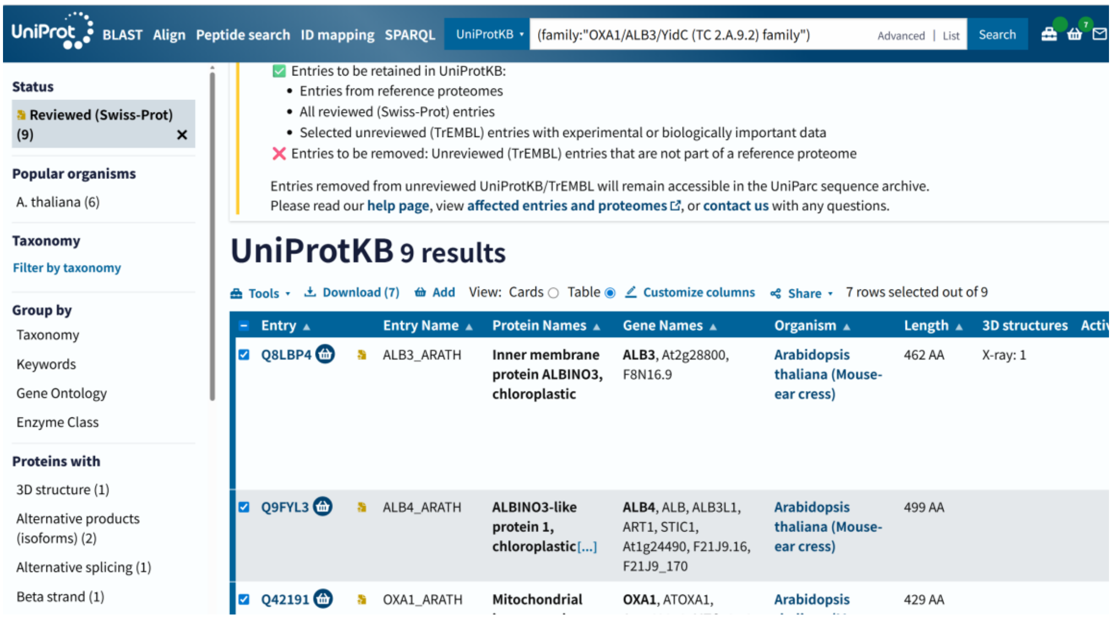
+	Compositional bias	431-442	Basic

Sequence similarities

Belongs to the OXA1/ALB3/YidC (TC 2.A.9.2) family. Curated

Keywords

Domain | #Transit peptide



(family:"OXA1/ALB3/YidC (TC 2.A.9.2) family")

Status

Reviewed (Swiss-Prot) (9) ✕

Popular organisms

A. thaliana (6)

Taxonomy

Filter by taxonomy

Group by

- Taxonomy
- Keywords
- Gene Ontology
- Enzyme Class

Proteins with

- 3D structure (1)
- Alternative products (isoforms) (2)
- Alternative splicing (1)
- Beta strand (1)

☑ Entries to be retained in UniProtKB:

- Entries from reference proteomes
- All reviewed (Swiss-Prot) entries
- Selected unreviewed (TrEMBL) entries with experimental or biologically important data

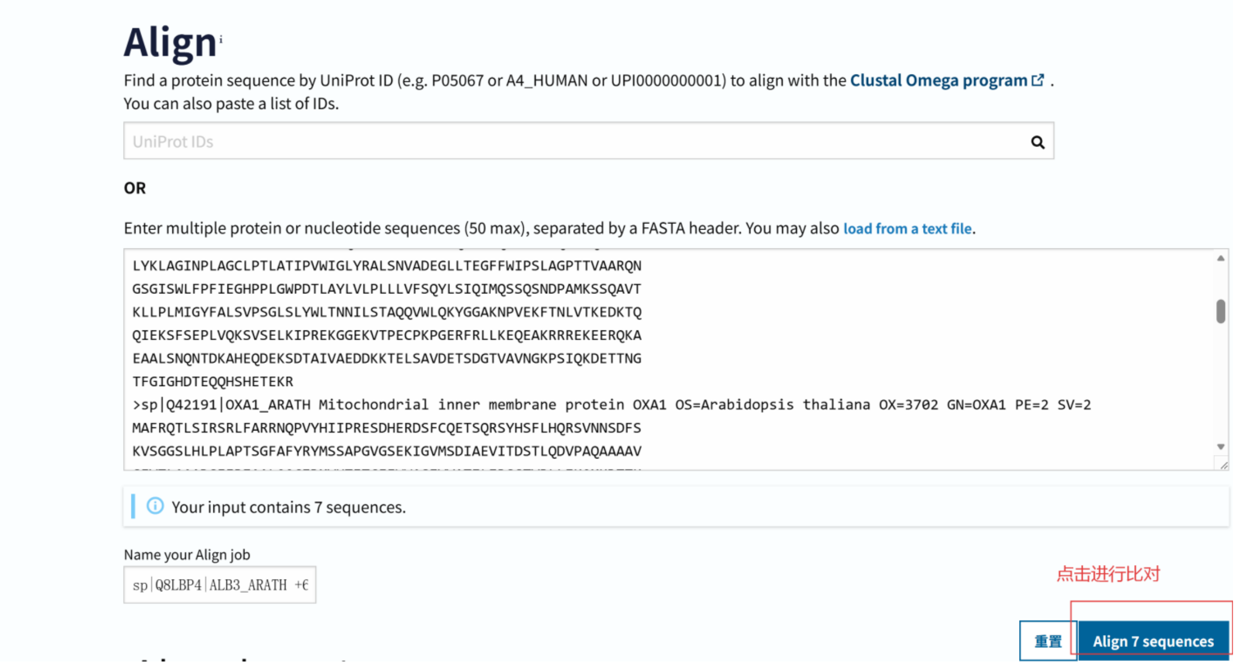
✕ Entries to be removed: Unreviewed (TrEMBL) entries that are not part of a reference proteome

Entries removed from unreviewed UniProtKB/TrEMBL will remain accessible in the UniParc sequence archive. Please read our help page, view affected entries and proteomes, or contact us with any questions.

UniProtKB 9 results

Tools · Download (7) · Add · View: Cards ○ Table ● · Customize columns · Share · 7 rows selected out of 9

	Entry		Entry Name	Protein Names	Gene Names	Organism	Length	3D structures	Acti
☑	Q8LBP4		ALB3_ARATH	**Inner membrane protein ALBINO3, chloroplastic**	**ALB3**, At2g28800, F8N16.9	Arabidopsis thaliana (Mouse-ear cress)	462 AA	X-ray: 1	
☑	Q9FYL3		ALB4_ARATH	**ALBINO3-like protein 1, chloroplastic**[...]	**ALB4**, ALB, ALB3L1, ART1, STIC1, At1g24490, F21J9.16, F21J9_170	Arabidopsis thaliana (Mouse-ear cress)	499 AA		
☑	Q42191		OXA1_ARATH	**Mitochondrial**	**OXA1**, ATOXA1,	Arabidopsis	429 AA		



Align

Find a protein sequence by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001) to align with the Clustal Omega program.
You can also paste a list of IDs.

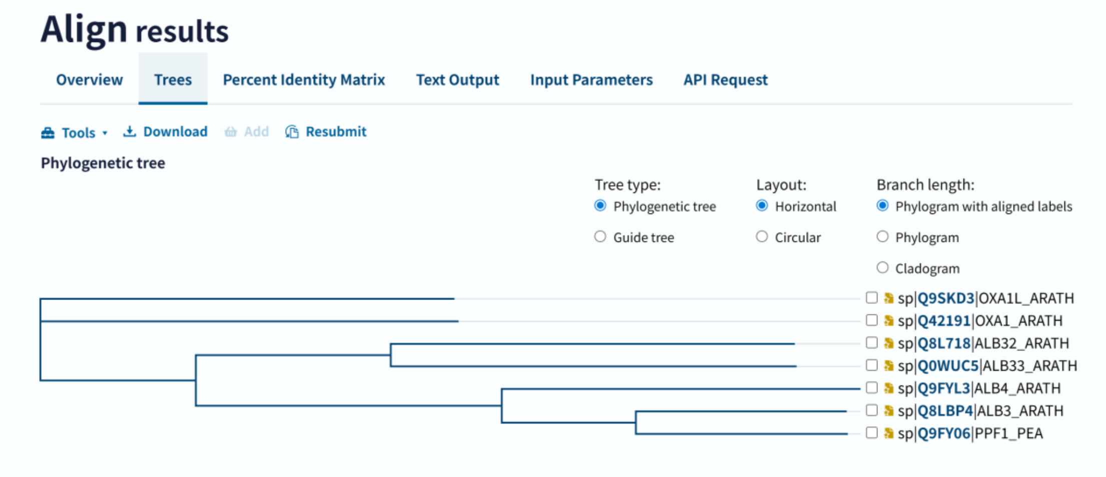
OR

Enter multiple protein or nucleotide sequences (50 max), separated by a FASTA header. You may also load from a text file.

```
LYKLAGINPLAGCLPTLATIPVWIGLYRALSNVADEGLLTEGFFWIPSLAGPTTVAARQN
GSGISWLFPFIEGHPPLGWPDTLAYLVLPLLLVFSQYLSIQIMQSSQSNDPAMKSSQAVT
KLLPLMIGYFALSVPSGLSLYWLTNNILSTAQQVWLQKYGGAKNPVEKFTNLVTKEDKTQ
QIEKSFSEPLVQKSVSELKIPREKGGEKVTPECPKPGERFRLLKEQEAKRRREKEERQKA
EAALSNQNTDKAHEQDEKSDTAIVAEDDKKTELSAVDETSDGTVAVNGKPSIQKDETTNG
TFGIGHDTEQQHSHETEKR
>sp|Q42191|OXA1_ARATH Mitochondrial inner membrane protein OXA1 OS=Arabidopsis thaliana OX=3702 GN=OXA1 PE=2 SV=2
MAFRQTLSIRSRLFARRNQPVYHIIPRESDHERDSFCQETSQRSYHSFLHQRSVNNSDFS
KVSGGSLHLPLAPTSGFAFYRYMSSAPGVGSEKIGVMSDIAEVITDSTLQDVPAQAAAAV
```

ⓘ Your input contains 7 sequences.

Name your Align job: sp|Q8LBP4|ALB3_ARATH +6



Align results

Phylogenetic tree

Tree type: ● Phylogenetic tree ○ Guide tree

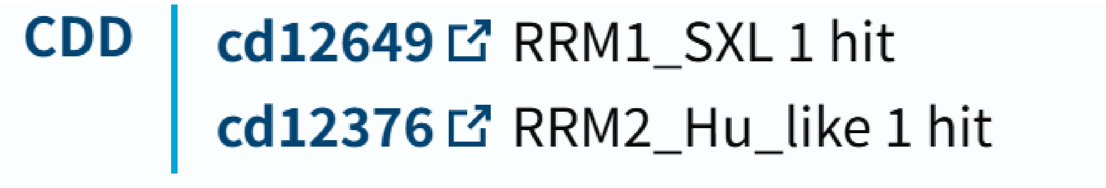
Layout: ● Horizontal ○ Circular

Branch length: ● Phylogram with aligned labels ○ Phylogram ○ Cladogram

B Uniprot 数据库的外部链接

① 查看果蝇的 Sxl 蛋白在 NCBI 蛋白质保守结构域数据库 CDD 中的信息

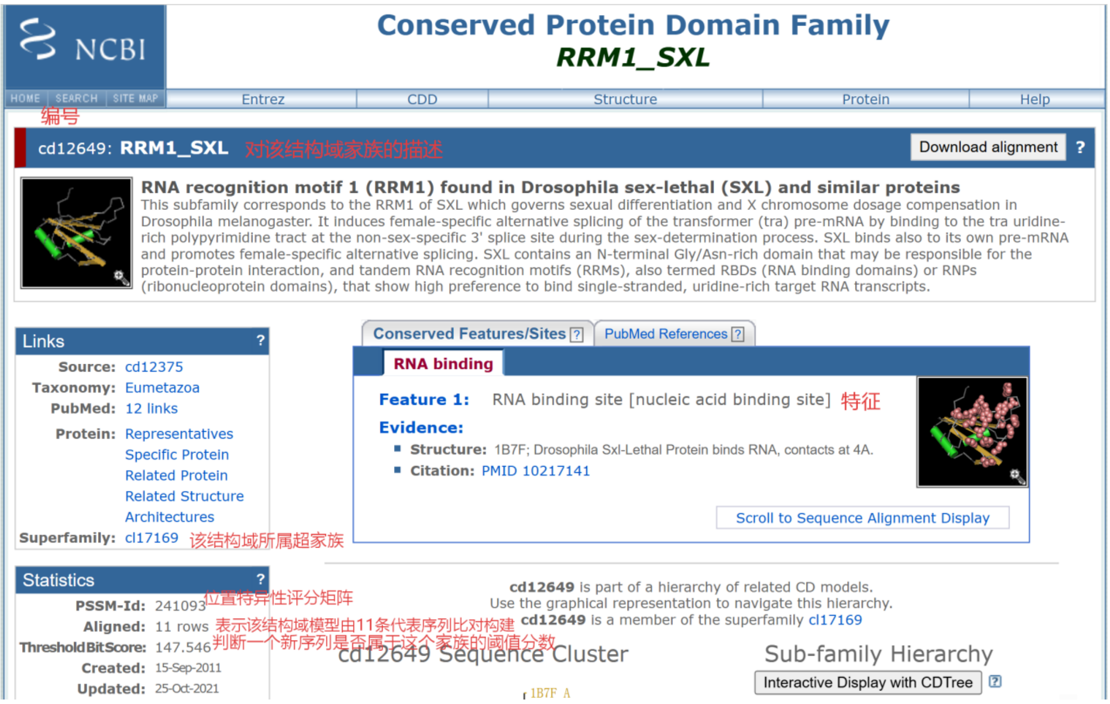
以下分别是 Sxl 蛋白序列中的结构域和他们在蛋白质序列中的个数，RRM1、RRM2 为结构域名称，1hit 表示有一个该结构域，分别可点击跳转至相关界面，以 RRMA1_SXL 为例，如下：



CDD

cd12649 RRM1_SXL 1 hit

cd12376 RRM2_Hu_like 1 hit



Conserved Protein Domain Family

RRM1_SXL

编号

cd12649: **RRM1_SXL** 对该结构域家族的描述

RNA recognition motif 1 (RRM1) found in Drosophila sex-lethal (SXL) and similar proteins

This subfamily corresponds to the RRM1 of SXL which governs sexual differentiation and X chromosome dosage compensation in Drosophila melanogaster. It induces female-specific alternative splicing of the transformer (tra) pre-mRNA by binding to the tra uridine-rich polypyrimidine tract at the non-sex-specific 3' splice site during the sex-determination process. SXL binds also to its own pre-mRNA and promotes female-specific alternative splicing. SXL contains an N-terminal Gly/Asn-rich domain that may be responsible for the protein-protein interaction, and tandem RNA recognition motifs (RRMs), also termed RBDs (RNA binding domains) or RNPs (ribonucleoprotein domains), that show high preference to bind single-stranded, uridine-rich target RNA transcripts.

Links

Source: cd12375
Taxonomy: Eumetazoa
PubMed: 12 links
Protein: Representatives
Specific Protein
Related Protein
Related Structure
Architectures
Superfamily: cl17169 该结构域所属超家族

Conserved Features/Sites | PubMed References

RNA binding

Feature 1: RNA binding site [nucleic acid binding site] 特征

Evidence:

- **Structure:** 1B7F; Drosophila Sxl-Lethal Protein binds RNA, contacts at 4A.
- **Citation:** PMID 10217141

Scroll to Sequence Alignment Display

Statistics

PSSM-Id: 241093 位置特异性评分矩阵
Aligned: 11 rows 表示该结构域模型由11条代表序列比对构建
ThresholdBitScore: 147.546 判断一个新序列是否属于这个家族的阈值分数
Created: 15-Sep-2011
Updated: 25-Oct-2021

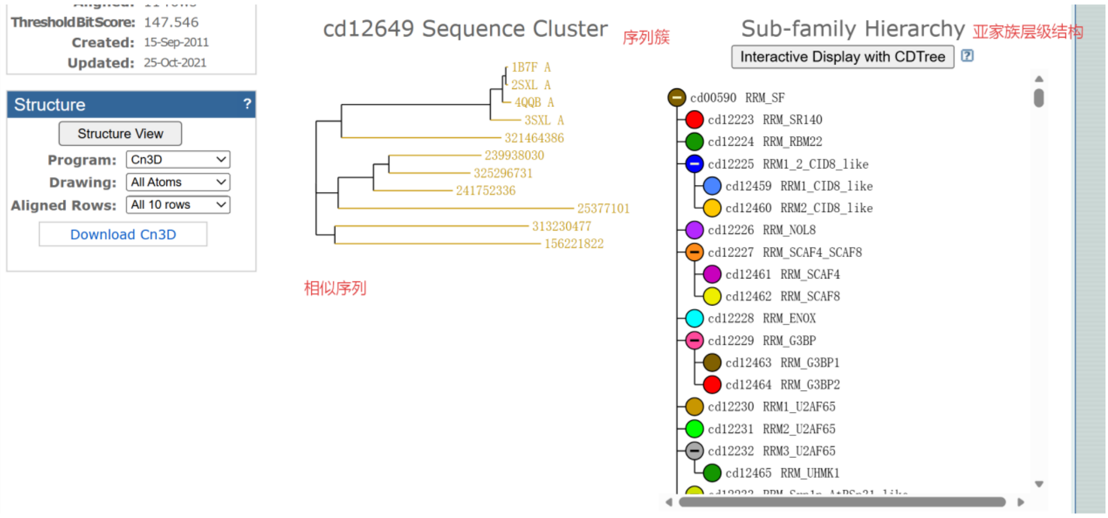
cd12649 is part of a hierarchy of related CD models.
Use the graphical representation to navigate this hierarchy.
cd12649 is a member of the superfamily cl17169

cd12649 Sequence Cluster

Sub-family Hierarchy

Interactive Display with CDTree

1B7F A



ThresholdBitScore: 147.546
Created: 15-Sep-2011
Updated: 25-Oct-2021

Structure

Program: Cn3D
Drawing: All Atoms
Aligned Rows: All 10 rows

Sequence Alignment ☐ include consensus sequence ?

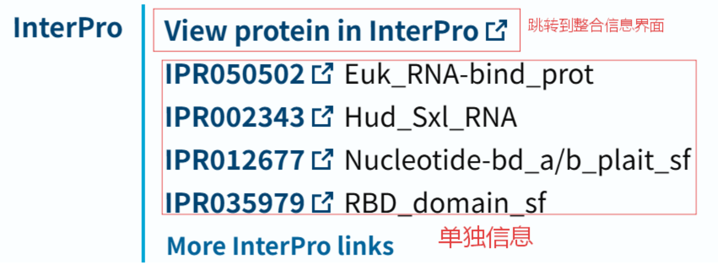
Format: Hypertext Row Display: All 10 rows Color Bits: 2.0 bit Type Selection: top listed sequences

```
Feature 1          # # ##  #                    # ####   ##### #                       ## #   ## #
1B7F_A          4 TNLIVNYLPQdMTDRELYALFRaIGPINTCRIMRDYkTGYSYGYAFVDFTsEMDSQRAIKVLNGITVRNKRLKVSYARPG 83  fruit fly
CBY18692       71 TNLIVNYLPQsFTDQEFFNLFStVGRVNKARIIRHRqTGYSFGYGFIDMAsADDAERAIRKLNQYQIGHKRLKVAYSLPS 150 Oikopleura dioica
EFX75394       27 TNLIINYLPQnLTESELFKMFVtIGTVTNCKIMRDFrTGYSYGFGFVNYQkADDAIRAIQTLNGLQIQNKRIKVSYARPP 106 common water flea
ACS36114       19 TNLIINYLPQtLTDEEFRSMFLaIGPVKASKIVRDKmTGYSYGFGFVDYNlESDASKAIQTLNGLQLQNKRIKVAYSRSG 98  Capitella sp. I...
NP_001191601   18 TNLIINYLPQtLTDEEFRSMFLsVGPVKSSKIVRDKsTGYSYGFGFVDFQhPTDAQRAIETLSGLQLQNKRIKVALARPG 97  California sea ...
XP_002401050   10 TNLIINYLPQtLTDEEFRSLFTsIGPIKSSKIVRHKaTGYSYGFGFVDYQsIEDAARAVGTLNGLQLQNKKIKVAYARPG 89  black-legged tick
ACT35655        7 TNLIINYLPQaFSDDEFRALFGsVAPVKRAKVCRDKrSGYSFGFGFCEFYtAEDASRAVRMLNGLQLHRKTIKVAHARLQ 86  Macrostomum lig...
EDO42673        5 TTLIVNYIPQdMTDQTFRMMFEaVASLNNCKIVRHKpSGWSYGFGFVDYNtTEDAQKAIDKLNGFTIGNKVLKVAFSRPG 84  starlet sea ane...
2SXL_A          4 TNLIVNYLPQdMTDRELYALFRaIGPINTCRIMRDYkTGYSYGYAFVDFTsEMDSQRAIKVLNGITVRNKRLKVSYARPG 83  fruit fly
4QQB_A          7 TNLIVNYLPQdMTDRELYALFRaIGPINTCRIMRDYkTGYSFGYAFVDFTsEMDSQRAIKVLNGITVRNKRLKVSYARPG 86  fruit fly
3SXL_A         15 TNLIVNYLPQdXTDRELYALFRaIGPINTCRIXRDYkTGYSFGYAFVDFTsEXDSQRAIKVLNGITVRNKRLKVSYARPG 94  fruit fly

Feature 1         #
1B7F_A         84 G 84  fruit fly
CBY18692      151 G 151 Oikopleura dioica
EFX75394      107 G 107 common water flea
ACS36114       99 E 99  Capitella sp. I Grassle & Grassle, 1976
NP_001191601   98 G 98  California sea hare
XP_002401050   90 G 90  black-legged tick
ACT35655       87 S 87  Macrostomum lignano
EDO42673       85 G 85  starlet sea anemone
2SXL_A         84 G 84  fruit fly
4QQB_A         87 G 87  fruit fly
3SXL_A         95 G 95  fruit fly
```

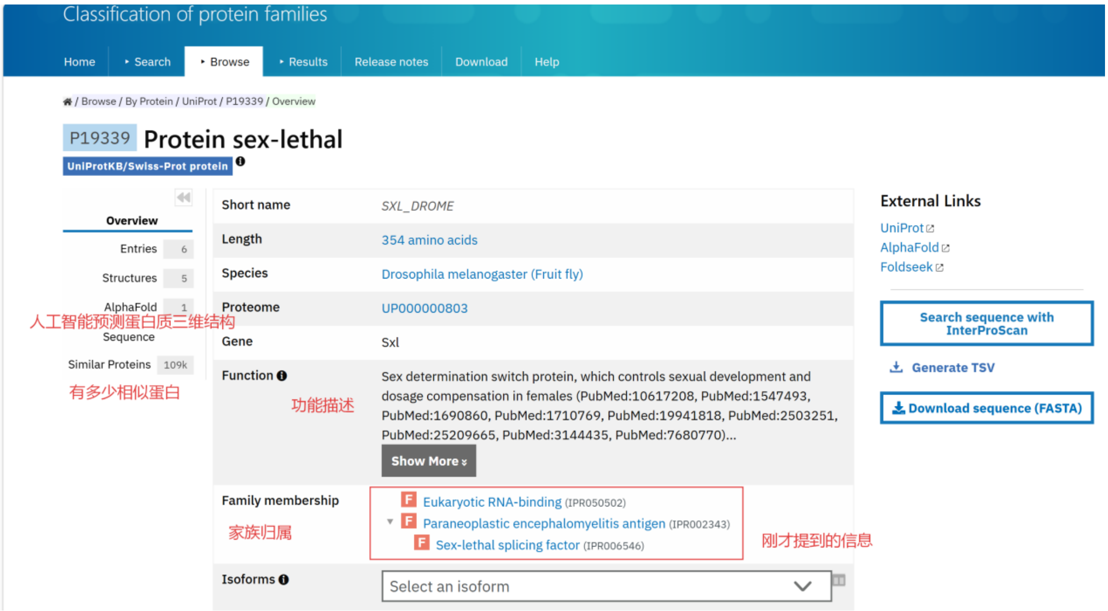
② 查看果蝇的性别致死蛋白 Sxl 在 EBI 蛋白质整合数据库 InterPro 中的信息

InterPro 整合了十多个主流数据库的蛋白质信息，包括结构域和位点、蛋白质家族、序列特征、结构分类等。以果蝇中的 Sxl 为例，IPR050502 告知 Sxl 属于整合 RNA 结合蛋白，IPR002343 告知 Sxl 蛋白的核心工具 Hud,性别致死蛋白家族结合的结构域，IPR012677 是上述工具的立体结构类型，IPR035979 是 RNA 识别基序超家族，总之这些都可以点击跳转到相关信息，而点击 Interpro 则可以跳转到整合有这些所有信息的界面。



InterPro

View protein in InterPro
IPR050502 Euk_RNA-bind_prot
IPR002343 Hud_Sxl_RNA
IPR012677 Nucleotide-bd_a/b_plait_sf
IPR035979 RBD_domain_sf
More InterPro links



/ Browse / By Protein / UniProt / P19339 / Overview

P19339 Protein sex-lethal

Short name	SXL_DROME
Length	354 amino acids
Species	Drosophila melanogaster (Fruit fly)
Proteome	UP000000803
Gene	Sxl
Function	Sex determination switch protein, which controls sexual development and dosage compensation in females (PubMed:10617208, PubMed:1547493, PubMed:1690860, PubMed:1710769, PubMed:19941818, PubMed:2503251, PubMed:25209665, PubMed:3144435, PubMed:7680770)...
Family membership	F Eukaryotic RNA-binding (IPR050502) F Paraneoplastic encephalomyelitis antigen (IPR002343) F Sex-lethal splicing factor (IPR006546)
Isoforms	Select an isoform

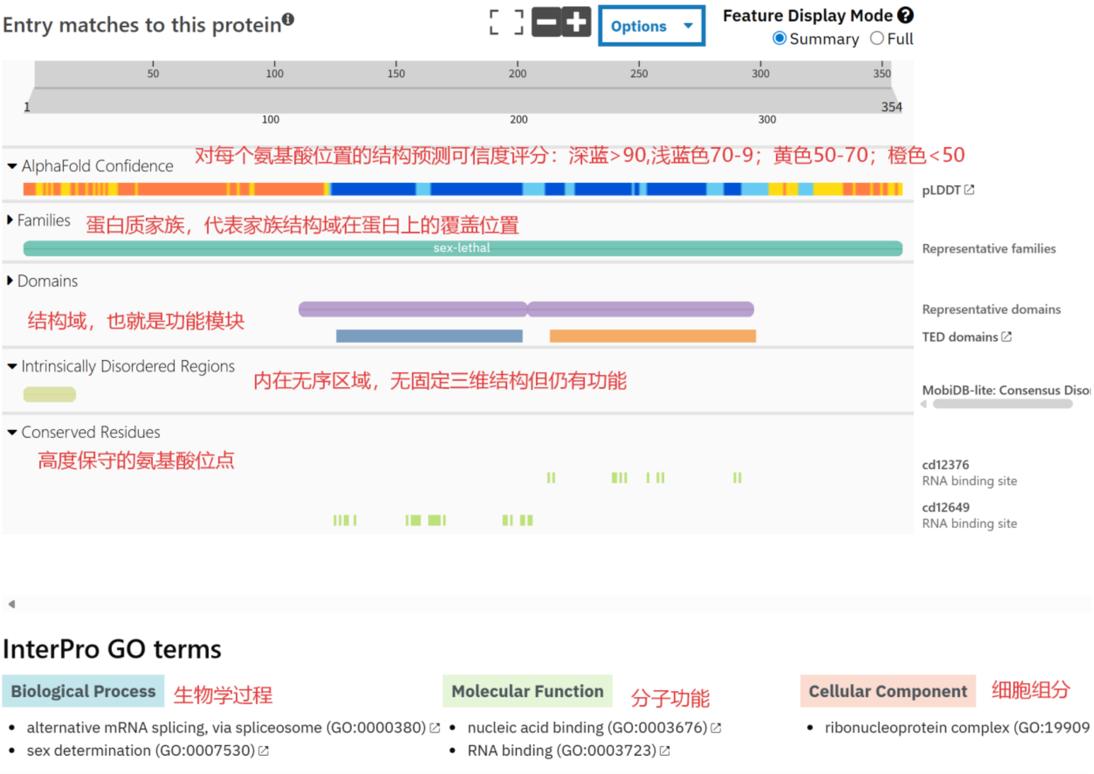
人工智能预测蛋白质三维结构

有多少相似蛋白

功能描述

家族归属

刚才提到的信息



对每个氨基酸位置的结构预测可信度评分：深蓝>90,浅蓝色70-9；黄色50-70；橙色<50

蛋白质家族，代表家族结构域在蛋白上的覆盖位置

结构域，也就是功能模块

内在无序区域，无固定三维结构但仍有功能

高度保守的氨基酸位点

InterPro GO terms

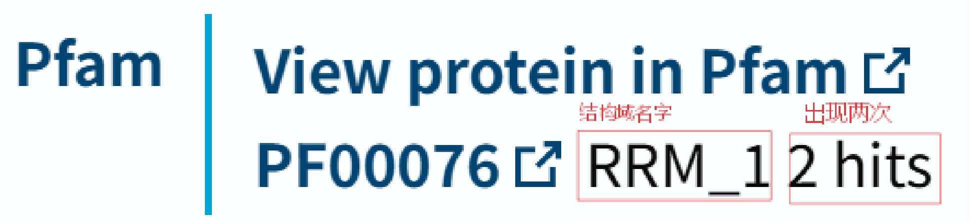
Biological Process 生物学过程
- alternative mRNA splicing, via spliceosome (GO:0000380)
- sex determination (GO:0007530)

Molecular Function 分子功能
- nucleic acid binding (GO:0003676)
- RNA binding (GO:0003723)

Cellular Component 细胞组分
- ribonucleoprotein complex (GO:19909

③ 查看致死蛋白 Sxl 在蛋白质家族数据库 PFam 中的信息

Pfam 是专门收录蛋白质结构域和家族的数据库。

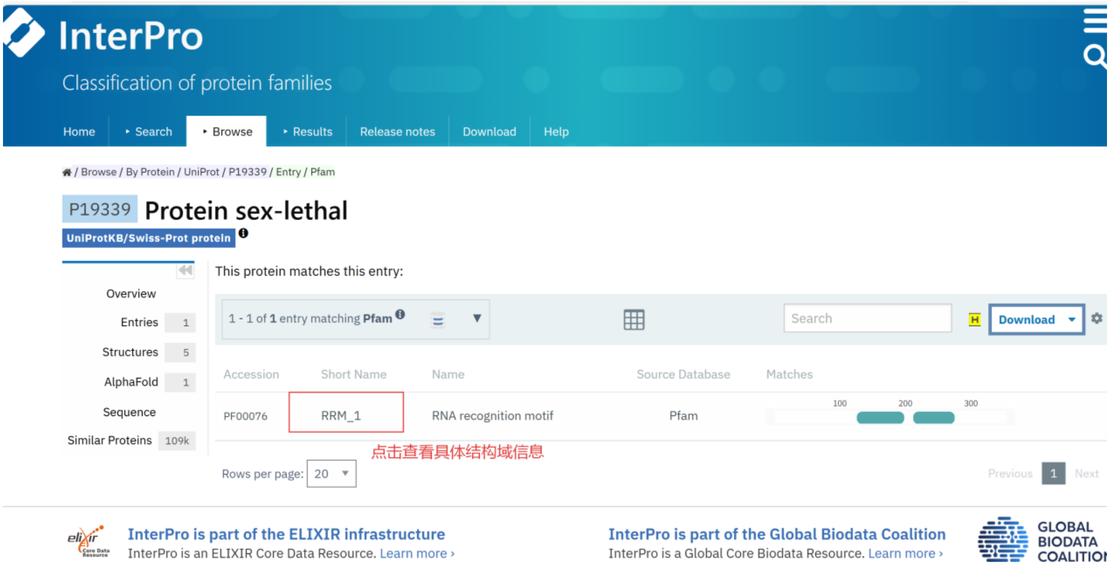

Pfam

View protein in Pfam

PF00076 RRM_1 2 hits

结构域名字

出现两次



/ Browse / By Protein / UniProt / P19339 / Entry / Pfam

P19339 Protein sex-lethal

UniProtKB/Swiss-Prot protein

This protein matches this entry:

1 - 1 of 1 entry matching Pfam

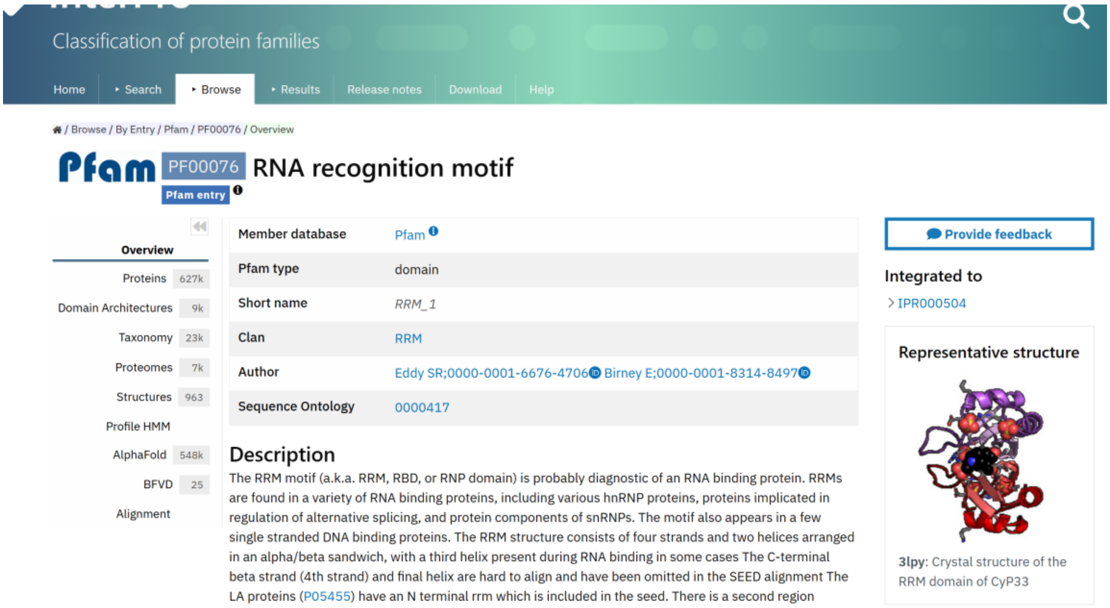
Accession	Short Name	Name	Source Database	Matches
PF00076	RRM_1	RNA recognition motif	Pfam	100 200 300

点击查看具体结构域信息

Rows per page: 20

InterPro is part of the ELIXIR infrastructure
InterPro is an ELIXIR Core Data Resource. Learn more ›

InterPro is part of the Global Biodata Coalition
InterPro is a Global Core Biodata Resource. Learn more ›



/ Browse / By Entry / Pfam / PF00076 / Overview

Pfam PF00076 RNA recognition motif

Pfam entry

Member database	Pfam
Pfam type	domain
Short name	*RRM_1*
Clan	RRM
Author	Eddy SR;0000-0001-6676-4706 Birney E;0000-0001-8314-8497
Sequence Ontology	0000417

Integrated to
IPR000504

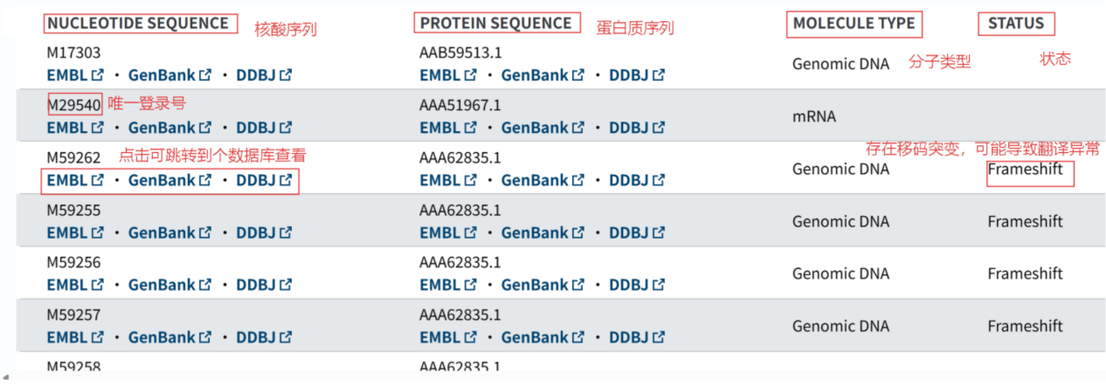
Representative structure

3lpy: Crystal structure of the RRM domain of CyP33

Description
The RRM motif (a.k.a. RRM, RBD, or RNP domain) is probably diagnostic of an RNA binding protein. RRMs are found in a variety of RNA binding proteins, including various hnRNP proteins, proteins implicated in regulation of alternative splicing, and protein components of snRNPs. The motif also appears in a few single stranded DNA binding proteins. The RRM structure consists of four strands and two helices arranged in an alpha/beta sandwich, with a third helix present during RNA binding in some cases The C-terminal beta strand (4th strand) and final helix are hard to align and have been omitted in the SEED alignment The LA proteins (P05455) have an N terminal rrm which is included in the seed. There is a second region

④ 查看核酸序列数据库中数据 DDBJ 中的信息（以 CEAM-6 为例）

多条序列可能是来自不同不同测序版本、不同个体、不同剪接变体或数据库冗余。



NUCLEOTIDE SEQUENCE 核酸序列	PROTEIN SEQUENCE 蛋白序列	MOLECULE TYPE 分子类型	STATUS 状态
M17303 EMBL • GenBank • DDBJ	AAB59513.1 EMBL • GenBank • DDBJ	Genomic DNA	
M29540 唯一登录号 EMBL • GenBank • DDBJ	AAA51967.1 EMBL • GenBank • DDBJ	mRNA	
M59262 点击可跳转到个数据库查看 EMBL • GenBank • DDBJ	AAA62835.1 EMBL • GenBank • DDBJ	Genomic DNA	存在移码突变，可能导致翻译异常 Frameshift
M59255 EMBL • GenBank • DDBJ	AAA62835.1 EMBL • GenBank • DDBJ	Genomic DNA	Frameshift
M59256 EMBL • GenBank • DDBJ	AAA62835.1 EMBL • GenBank • DDBJ	Genomic DNA	Frameshift
M59257 EMBL • GenBank • DDBJ	AAA62835.1 EMBL • GenBank • DDBJ	Genomic DNA	Frameshift
M59258	AAA62835.1		

```
LOCUS       HUMANTCE                3036 bp    DNA     linear   HUM 08-AUG-1995
DEFINITION  Human carcinoembryonic antigen gene, complete cds.
ACCESSION   M17303
VERSION     M17303.1 登录号+版本号.1是原版，修正后是.2
KEYWORDS    carcinoembryonic antigen.
SOURCE      Homo sapiens (human) 来源
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3036)
  AUTHORS   Beauchemin,N.                          文献信息
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2541)
  AUTHORS   Beauchemin,N., Benchimol,S., Cournoyer,D., Fuks,A. and
            Stanners,C.P.
  TITLE     Isolation and characterization of full-length functional cDNA
            clones for human carcinoembryonic antigen
  JOURNAL   Mol. Cell. Biol. 7 (9), 3221-3230 (1987)
   PUBMED   3670312
COMMENT     Original source text: Human colonic adenocarcinoma cell line Ls180,
            cDNA to mRNA, and DNA.
            Draft entry and computer-readable sequence for [2],[1] kindly
            provided by N.Beauchemin, 23-NOV-1987.
            [1]  revises [2].                      序列原始来源
FEATURES 关键特征   Location/Qualifiers
     source          1..3036
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="19q13.2" 定位在人类19号染色体长臂1区3带2亚带
     mRNA            1..2541                    这些数字代表位置
                     /product="CEA mRNA"
     variation       58
                     /note="a in one clone; t in another" a和t两个克隆
                     /replace="t"
     gene            97..2205 基因本身位于97-2205位碱基
                     /gene="CEA"
     CDS             97..2205
                     /gene="CEA" 编码区
                     /note="carcinoembryonic antigen precursor"
                     /codon_start=1
                     /protein_id="AAB59513.1"
                     /db_xref="GDB:G00-119-054"
                     /translation="MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFN
                     VAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIY
                     LSAGATVGIMIGVLVGVALI"
信号肽 sig_peptide     97..198
                     /gene="CEA"
                     /note="carcinoembryonic antigen signal peptide"
成熟肽 mat_peptide     199..2202
                     /gene="CEA"
                     /product="carcinoembryonic antigen"
重复区域repeat_region   2330..2560
                     /note="Alu repeat" Alu重复序列
BASE COUNT 碱基统计  860 a        853 c        614 g        709 t
ORIGIN      477 bp upstream of BglII site; chromosome 19.
        1 cgaccagcag accagacagt cacagcagcc ttgacaaaac gttcctggaa ctcaagcact
       61 tctccacaga ggaggacaga gcagacagca gagaccatgg agtctccctc ggcccctccc
      121 cacagatggt gcatccccctg gcagaggctc ctgctcacag cctcacttct aaccttctgg
      181 aacccgccca ccactgccaa gctcactatt gaatccacgc cgttcaatgt cgcagagggg
      241 aaggaggtgc ttctacttgt ccacaatctg ccccagcatc tttttggcta cagctggtac
```

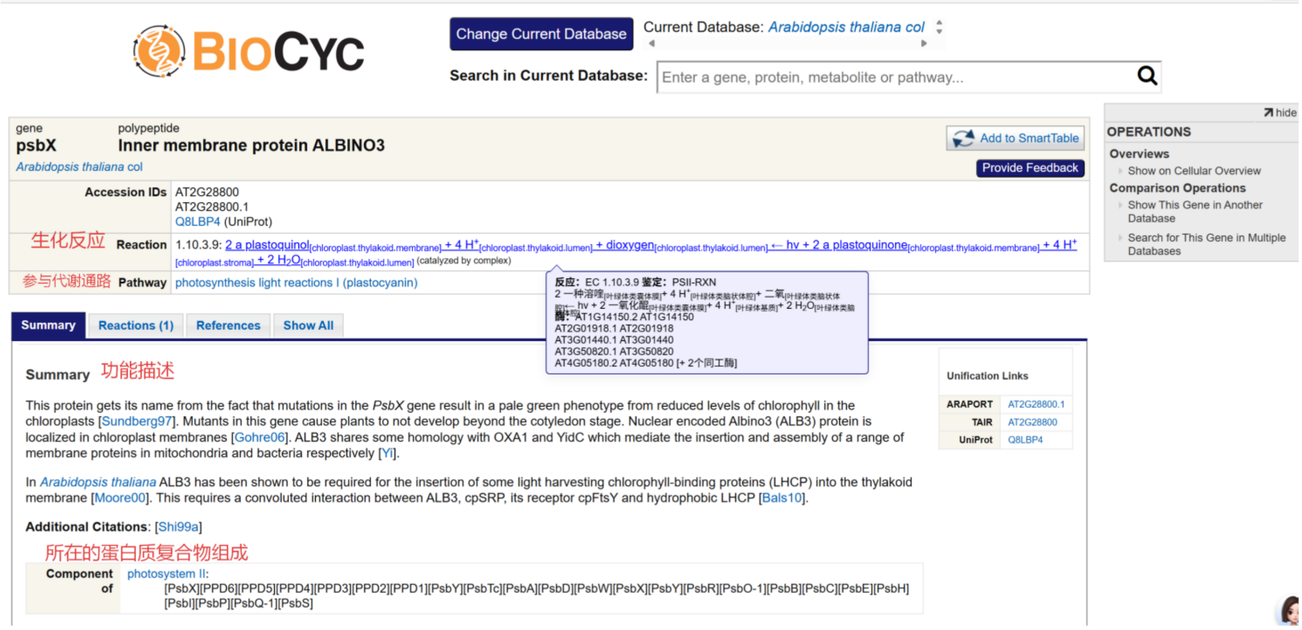
⑤　查看在代谢通路数据库 BioCyc 中的信息（以 ALB3 为例）。

物种名：物种基因编号-该基因编码蛋白是单体

ARA:AT2G28800-MONOMER

MetaCyc:AT2G28800-MONOMER

MetaCyc数据库也有收录



Current Database: *Arabidopsis thaliana col*

gene
psbX
Arabidopsis thaliana col

polypeptide
Inner membrane protein ALBINO3

Accession IDs AT2G28800
AT2G28800.1
Q8LBP4 (UniProt)

生化反应 **Reaction** 1.10.3.9: 2 a plastoquinol[chloroplast.thylakoid.membrane] + 4 H^+[chloroplast.thylakoid.lumen] + dioxygen[chloroplast.thylakoid.lumen] ← hv + 2 a plastoquinone[chloroplast.thylakoid.membrane] + 4 H^+[chloroplast.stroma] + 2 H_2O[chloroplast.thylakoid.lumen] (catalyzed by complex)

参与代谢通路 **Pathway** photosynthesis light reactions I (plastocyanin)

反应： EC 1.10.3.9 **鉴定：** PSII-RXN
2 一种溶喹[叶绿体类囊体膜] + 4 H^+[叶绿体类胎状体腔] + 二氧[叶绿体类胎状体腔] ← hv + 2 一氧化醌[叶绿体类囊体膜] + 4 H^+[叶绿体基质] + 2 H_2O[叶绿体类胎状体腔]
酶： AT1G14150.2 AT1G14150
AT2G01918.1 AT2G01918
AT3G01440.1 AT3G01440
AT3G50820.1 AT3G50820
AT4G05180.2 AT4G05180 [+ 2个同工酶]

Summary 功能描述

This protein gets its name from the fact that mutations in the *PsbX* gene result in a pale green phenotype from reduced levels of chlorophyll in the chloroplasts [Sundberg97]. Mutants in this gene cause plants to not develop beyond the cotyledon stage. Nuclear encoded Albino3 (ALB3) protein is localized in chloroplast membranes [Gohre06]. ALB3 shares some homology with OXA1 and YidC which mediate the insertion and assembly of a range of membrane proteins in mitochondria and bacteria respectively [Yi].

In *Arabidopsis thaliana* ALB3 has been shown to be required for the insertion of some light harvesting chlorophyll-binding proteins (LHCP) into the thylakoid membrane [Moore00]. This requires a convoluted interaction between ALB3, cpSRP, its receptor cpFtsY and hydrophobic LHCP [Bals10].

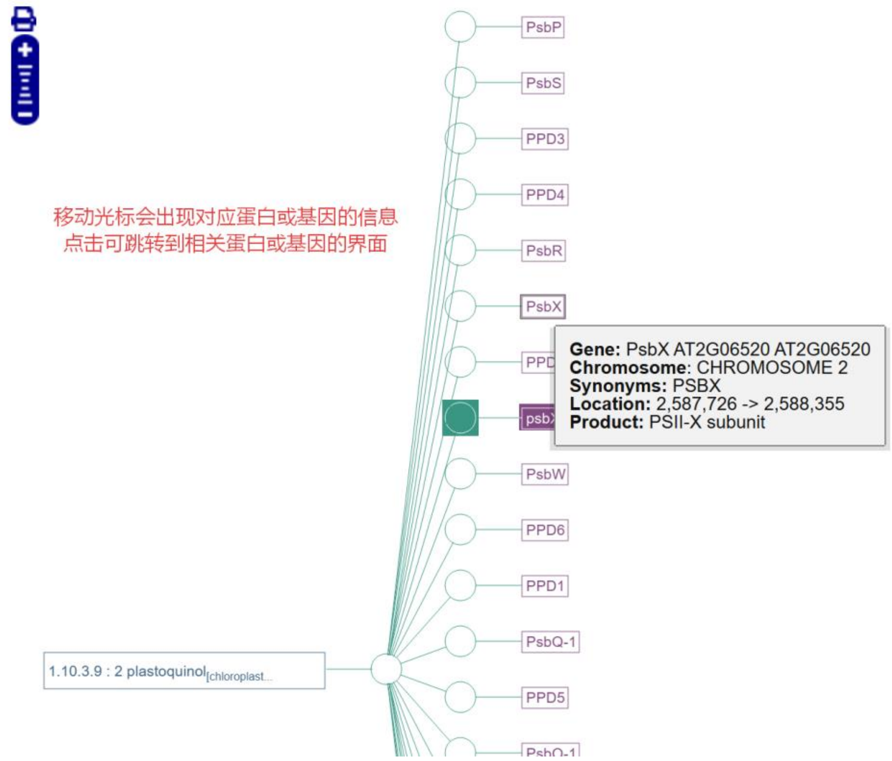
Additional Citations: [Shi99a]

所在的蛋白质复合物组成

Component of photosystem II: [PsbX][PPD6][PPD5][PPD4][PPD3][PPD2][PPD1][PsbY][PsbTc][PsbA][PsbD][PsbW][PsbX][PsbY][PsbR][PsbO-1][PsbB][PsbC][PsbE][PsbH][PsbI][PsbP][PsbQ-1][PsbS]

Unification Links	
ARAPORT	AT2G28800.1
TAIR	AT2G28800
UniProt	Q8LBP4

Gene-Reaction Schematic



移动光标会出现对应蛋白或基因的信息
点击可跳转到相关蛋白或基因的界面

Gene: PsbX AT2G06520 AT2G06520
Chromosome: CHROMOSOME 2
Synonyms: PSBX
Location: 2,587,726 -> 2,588,355
Product: PSII-X subunit

⑥　查看在转运蛋白分类数据库 TCDB 中的信息（以 ALB3 为例）

2.A.9.2.1

Chloroplast protein, ALBINO3 (Alb3). Inserts a subset of light harvesting chlorophyll-binding proteins) (Gerdes *et al.*, 2006). SRP43 (3.A.5.1.2) and the translocase, Alb3, interact directly (Dünschede *et al.*, 2011). SRP43 is an ATP-independent chaperone containing ankyrin repeats required for the biogenesis of the most abundant class of membrane proteins, the light-harvesting chlorophyll a/b-binding proteins (LHCPs) (McAvoy *et al.* 2018).

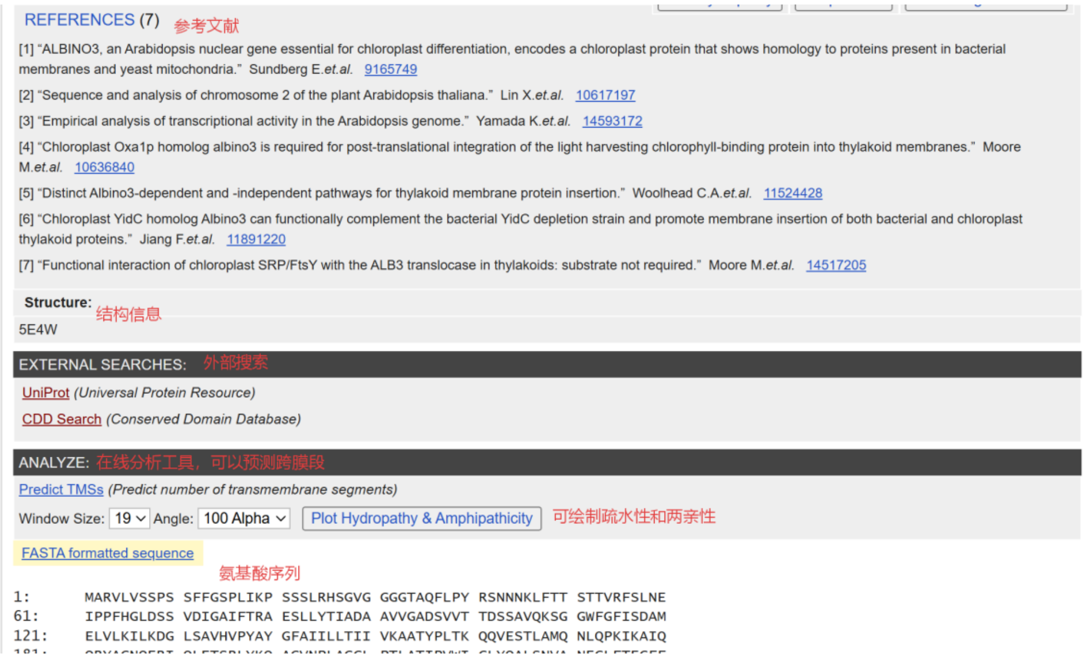
Accession Number:	Q8LBP4
Protein Name:	Inner membrane protein ALBINO3, chloroplastic
Length:	462
Molecular Weight:	50245.00
Species:	[3702]
Number of TMSs:	6
Location^1 / Topology^2 / Orientation^3:	Plastid^1 / Multi-pass membrane protein^2
Substrate	protein polypeptide chain, protein

CROSS DATABASE LINKS:

RefSeq:	NP_180446.1 NP_850125.1
Entrez Gene ID:	817429
Pfam:	PF02096

GENE ONTOLOGY

GO:0009535	C:chloroplast thylakoid membrane
GO:0016021	C:integral to membrane



REFERENCES (7)

[1] “ALBINO3, an Arabidopsis nuclear gene essential for chloroplast differentiation, encodes a chloroplast protein that shows homology to proteins present in bacterial membranes and yeast mitochondria.” Sundberg E.*et.al.* 9165749

[2] “Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.” Lin X.*et.al.* 10617197

[3] “Empirical analysis of transcriptional activity in the Arabidopsis genome.” Yamada K.*et.al.* 14593172

[4] “Chloroplast Oxa1p homolog albino3 is required for post-translational integration of the light harvesting chlorophyll-binding protein into thylakoid membranes.” Moore M.*et.al.* 10636840

[5] “Distinct Albino3-dependent and -independent pathways for thylakoid membrane protein insertion.” Woolhead C.A.*et.al.* 11524428

[6] “Chloroplast YidC homolog Albino3 can functionally complement the bacterial YidC depletion strain and promote membrane insertion of both bacterial and chloroplast thylakoid proteins.” Jiang F.*et.al.* 11891220

[7] “Functional interaction of chloroplast SRP/FtsY with the ALB3 translocase in thylakoids: substrate not required.” Moore M.*et.al.* 14517205

Structure:

5E4W

EXTERNAL SEARCHES:

UniProt *(Universal Protein Resource)*

CDD Search *(Conserved Domain Database)*

ANALYZE:

Predict TMSs *(Predict number of transmembrane segments)*

Window Size: 19 Angle: 100 Alpha [Plot Hydropathy & Amphipathicity]

FASTA formatted sequence

```
1:      MARVLVSSPS SFFGSPLIKP SSSLRHSGVG GGGTAQFLPY RSNNNKLFTT STTVRFSLNE
61:     IPPFHGLDSS VDIGAIFTRA ESLLYTIADA AVVGADSVVT TDSSAVQKSG GWFGFISDAM
121:    ELVLKILKDG LSAVHVPYAY GFAIILLTII VKAATYPLTK QQVESTLAMQ NLQPKIKAIQ
```

⑦　查看果蝇 Sxl 蛋白在蛋白质无序区域数据库 DisProt 中的信息

DP01605 - Protein sex-lethal

Organism Drosophila melanogaster　**Gene** Sxl (Sx1) [CG43770]　**Sequence length** 354　**Disorder content** 12.4% 无序含量

Cross references UniProtKB:P19339, MobiDB:P19339 交叉引用

Dataset(s) RNA-binding proteins RNA结合蛋白数据集

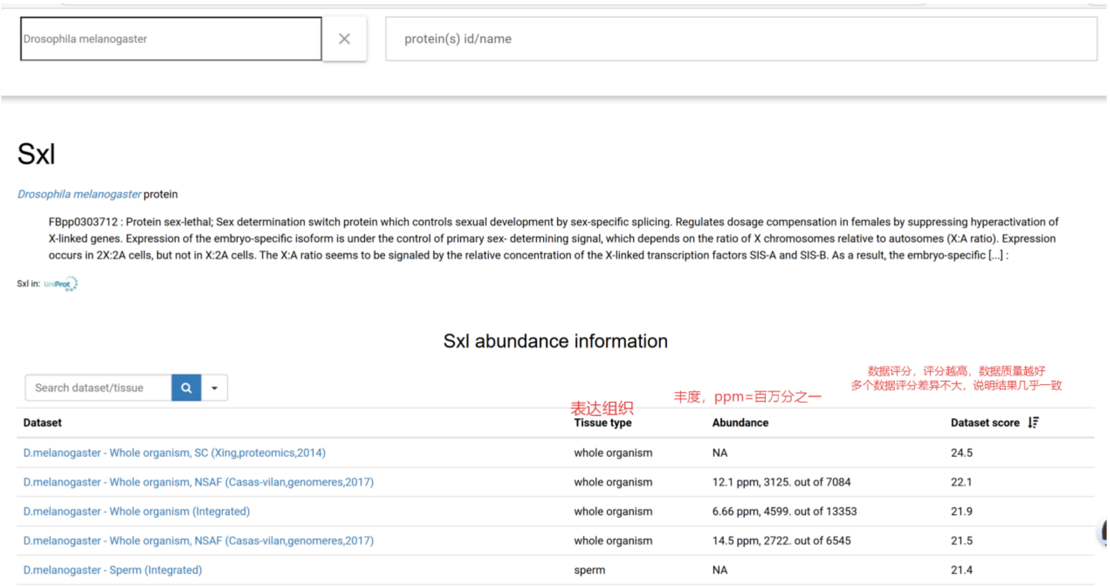
Last update 2023-09-25 最近更新

下拉界面可出现所有右侧项目的具体注释信息
包括实验、参考文献、证据等。
点击搜索某项目可筛选该项目信息

Selected regions **7 / 7**

⑧　查看果蝇的 Sxl 蛋白在蛋白质丰度数据库 PaxDB 中的信息

主要是回答了在果蝇体内，Sxl 蛋白有多少、不同组织含量有什么变化的问题。



Sxl

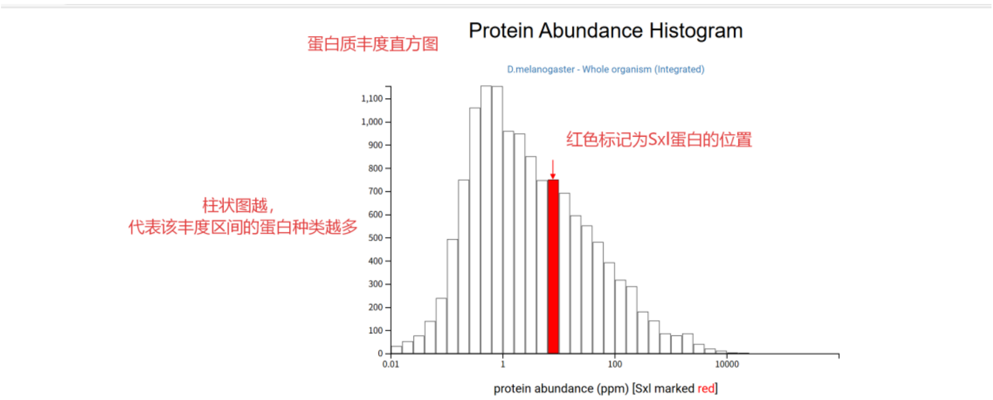
Drosophila melanogaster protein

FBpp0303712 : Protein sex-lethal; Sex determination switch protein which controls sexual development by sex-specific splicing. Regulates dosage compensation in females by suppressing hyperactivation of X-linked genes. Expression of the embryo-specific isoform is under the control of primary sex- determining signal, which depends on the ratio of X chromosomes relative to autosomes (X:A ratio). Expression occurs in 2X:2A cells, but not in X:2A cells. The X:A ratio seems to be signaled by the relative concentration of the X-linked transcription factors SIS-A and SIS-B. As a result, the embryo-specific [...] :

Sxl abundance information

数据评分，评分越高，数据质量越好
多个数据评分差异不大，说明结果几乎一致

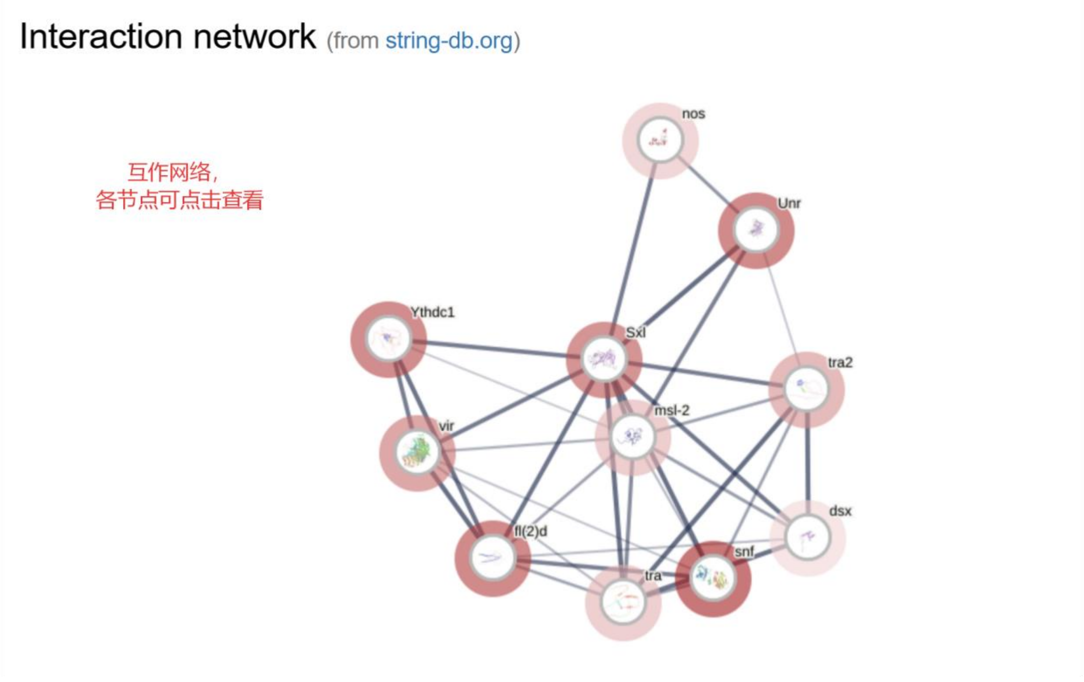
Dataset	表达组织 Tissue type	丰度，ppm=百万分之一 Abundance	Dataset score
D.melanogaster - Whole organism, SC (Xing,proteomics,2014)	whole organism	NA	24.5
D.melanogaster - Whole organism, NSAF (Casas-vilan,genomeres,2017)	whole organism	12.1 ppm, 3125. out of 7084	22.1
D.melanogaster - Whole organism (Integrated)	whole organism	6.66 ppm, 4599. out of 13353	21.9
D.melanogaster - Whole organism, NSAF (Casas-vilan,genomeres,2017)	whole organism	14.5 ppm, 2722. out of 6545	21.5
D.melanogaster - Sperm (Integrated)	sperm	NA	21.4



蛋白质丰度直方图

红色标记为Sxl蛋白的位置

柱状图越，
代表该丰度区间的蛋白种类越多



互作网络，
各节点可点击查看

⑨ 查看拟南芥内膜蛋白 ALB3_ARATH 在蛋白质修饰数据库 iPTMnet 中的信息

蛋白质修饰数据库 iPTMnet，Protein post-translational modifications (PTMs)，iPTMnet 是一个整合型蛋白质翻译后修饰（PTM）知识数据库，由美国特拉华大学与乔治城大学联合开发，专注于蛋白质翻译后修饰位点、修饰酶、底物、相互作用及功能网络的整合与可视化。

iPTMnet Report for Q8LBP4 (ALB3)

Protein Information

UniProt AC / UniProt ID	Q8LBP4 / ALB3_ARATH
Protein Name	Inner membrane protein ALBINO3, chloroplastic precursor
Gene Name	Name: ALB3 OrderedLocusNames:At2g28800;ORFNames:F8N16.9;
Organism	Arabidopsis thaliana (Mouse-ear cress)

该蛋白在 PRO（Protein Ontology，蛋白质本体）数据库的 ID

PRO ID	PR:Q8LBP4
PRO Name	inner membrane protein ALBINO3, chloroplastic (Arabidopsis thaliana)
Definition	A protein that is a translation product of the ALB3 gene in Arabidopsis thaliana.
Short Label	At-ALB3
Category	organism-gene

Interactive Sequence View

交互式序列视图，可视化标注蛋白全长上的所有 PTM 修饰位点

Loading...

Q8LBP4 (ALB3) as Substrate

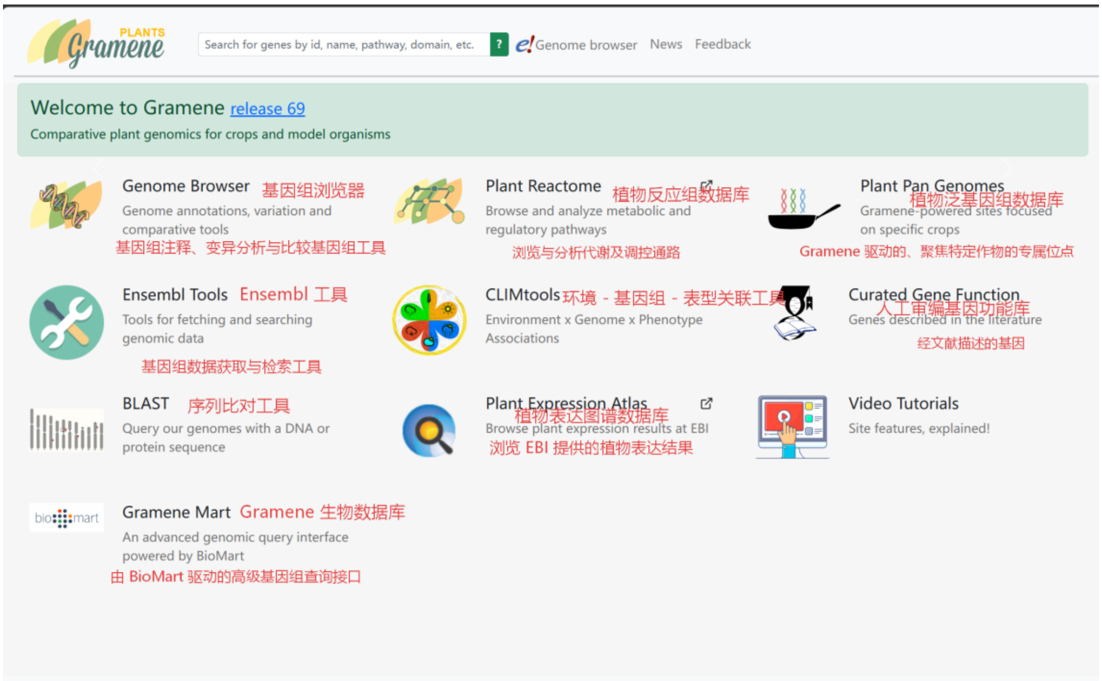
展示该蛋白作为底物的所有 PTM 修饰位点、修饰酶、数据来源、文献证据

Q8LBP4

⑩　拟南芥内膜蛋白 ALB3_ARATH 在拟南芥基因组数据库 GrameneE 中的信息

GrameneE 是 Gramene（植物比较基因组数据库），基于 Ensembl Plant 框架搭建的专业基因组浏览器界面，专门用于查看玉米、水稻、拟南芥等各类植物的基因结构、染色体位置、转录本、蛋白序列及同源信息，是植物基因定位、结构分析和比较基因组研究最常用的可视化工具之一。

Gramene，植物比较基因组数据库：



Welcome to Gramene release 69

Comparative plant genomics for crops and model organisms

Genome Browser 基因组浏览器
Genome annotations, variation and comparative tools
基因组注释、变异分析与比较基因组工具

Plant Reactome 植物反应组数据库
Browse and analyze metabolic and regulatory pathways
浏览与分析代谢及调控通路

Plant Pan Genomes 植物泛基因组数据库
Gramene-powered sites focused on specific crops
Gramene 驱动的、聚焦特定作物的专属位点

Ensembl Tools Ensembl 工具
Tools for fetching and searching genomic data
基因组数据获取与检索工具

CLIMtools 环境 - 基因组 - 表型关联工具
Environment x Genome x Phenotype Associations

Curated Gene Function 人工审编基因功能库
Genes described in the literature
经文献描述的基因

BLAST 序列比对工具
Query our genomes with a DNA or protein sequence

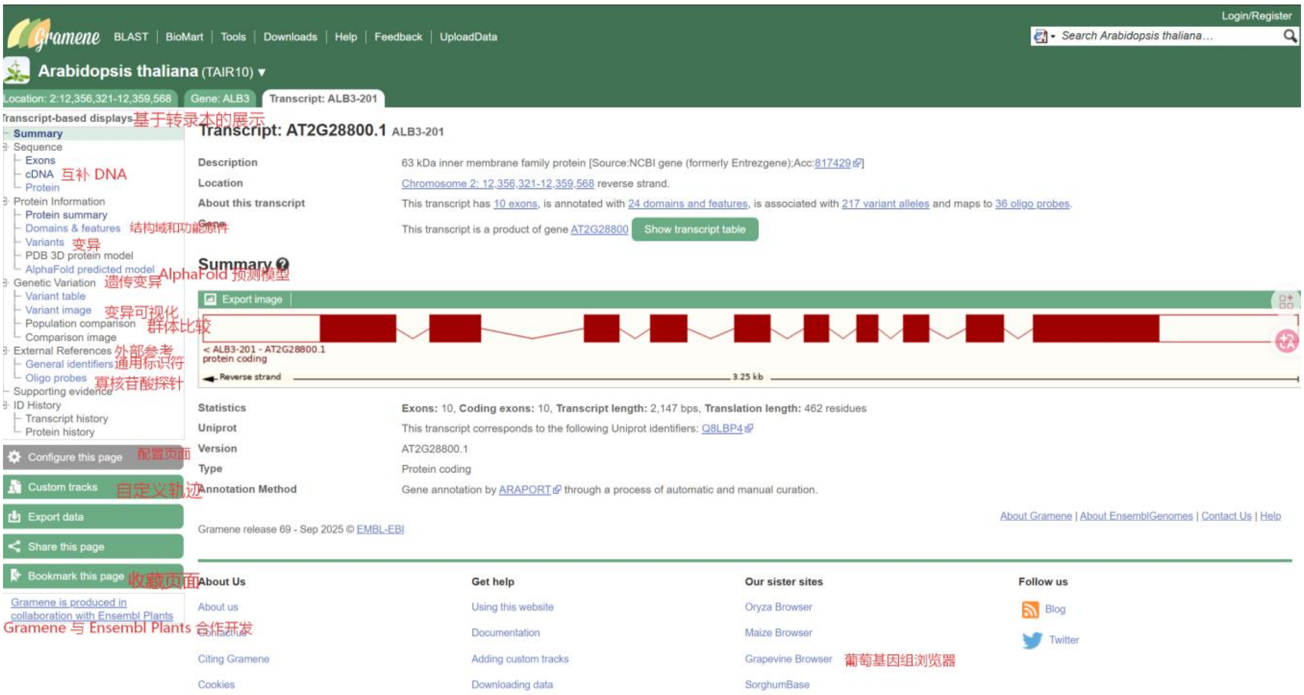
Plant Expression Atlas 植物表达图谱数据库
Browse plant expression results at EBI
浏览 EBI 提供的植物表达结果

Video Tutorials
Site features, explained!

Gramene Mart Gramene 生物数据库
An advanced genomic query interface powered by BioMart
由 BioMart 驱动的高级基因组查询接口

Plant Pan Genomes（植物泛基因组）：指一个植物物种内所有代表性个体/种质的全部基因与非编码序列的总和，包含核心基因组（Core Genome）（所有个体共有的保守基因）、非必需基因组（Dispensable/Accessory Genome）（仅部分个体拥有的可变基因）与特有基因（Unique Genes）（单个/少数个体特有的基因），突破单一参考基因组的局限性，完整呈现物种遗传多样性。

拟南芥内膜蛋白 ALB3_ARATH 在拟南芥基因组数据库 GrameneE 中的信息:

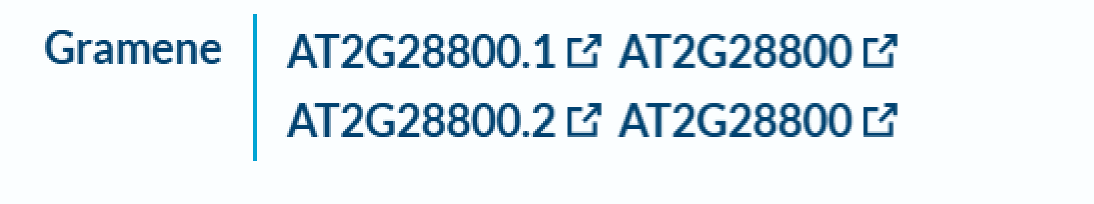


Transcript: AT2G28800.1 ALB3-201

Description	63 kDa inner membrane family protein [Source:NCBI gene (formerly Entrezgene);Acc:817429]
Location	Chromosome 2: 12,356,321-12,359,568 reverse strand.
About this transcript	This transcript has 10 exons, is annotated with 24 domains and features, is associated with 217 variant alleles and maps to 36 oligo probes.
Gene	This transcript is a product of gene AT2G28800
Statistics	**Exons:** 10, **Coding exons:** 10, **Transcript length:** 2,147 bps, **Translation length:** 462 residues
Uniprot	This transcript corresponds to the following Uniprot identifiers: Q8LBP4
Version	AT2G28800.1
Type	Protein coding
Annotation Method	Gene annotation by ARAPORT through a process of automatic and manual curation.

Gramene release 69 - Sep 2025 © EMBL-EBI

AT2G28800 相关 ID 区别与联系：



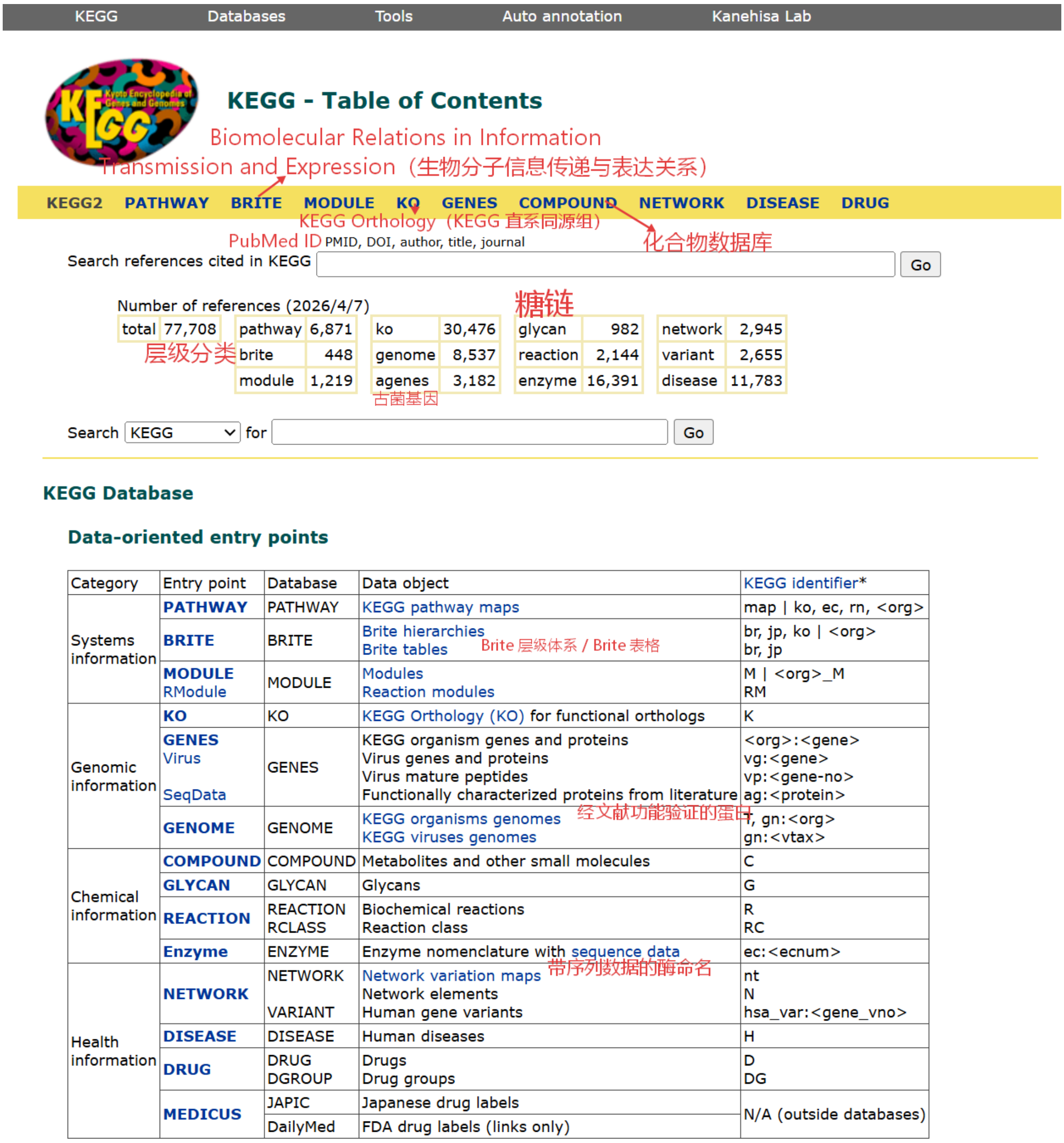
Gramene

AT2G28800.1 AT2G28800

AT2G28800.2 AT2G28800

AT2G28800 是基因座层级代表整个基因，是该基因的唯一、永久标识符；AT2G28800.1，是基因座 AT2G28800 的第 1 号转录本，代表基因的一个特定可变剪接产物；AT2G28800.2 是基因座 AT2G28800 的第 2 号转录本（可变剪接转录本），代表同一基因的另一个可变剪接产物。

11　查看拟南芥内膜蛋白 ALB3_ARATH 在代谢通路数据库 KEGG 中的信息



KEGG - Table of Contents

Biomolecular Relations in Information Transmission and Expression（生物分子信息传递与表达关系）

KEGG Orthology（KEGG 直系同源组）

化合物数据库

PubMed ID PMID, DOI, author, title, journal

Search references cited in KEGG ☐ Go

Number of references (2026/4/7)

糖链

total	77,708	pathway	6,871	ko	30,476	glycan	982	network	2,945
		brite	448	genome	8,537	reaction	2,144	variant	2,655
		module	1,219	agenes	3,182	enzyme	16,391	disease	11,783

层级分类

古菌基因

Search KEGG for ☐ Go

KEGG Database

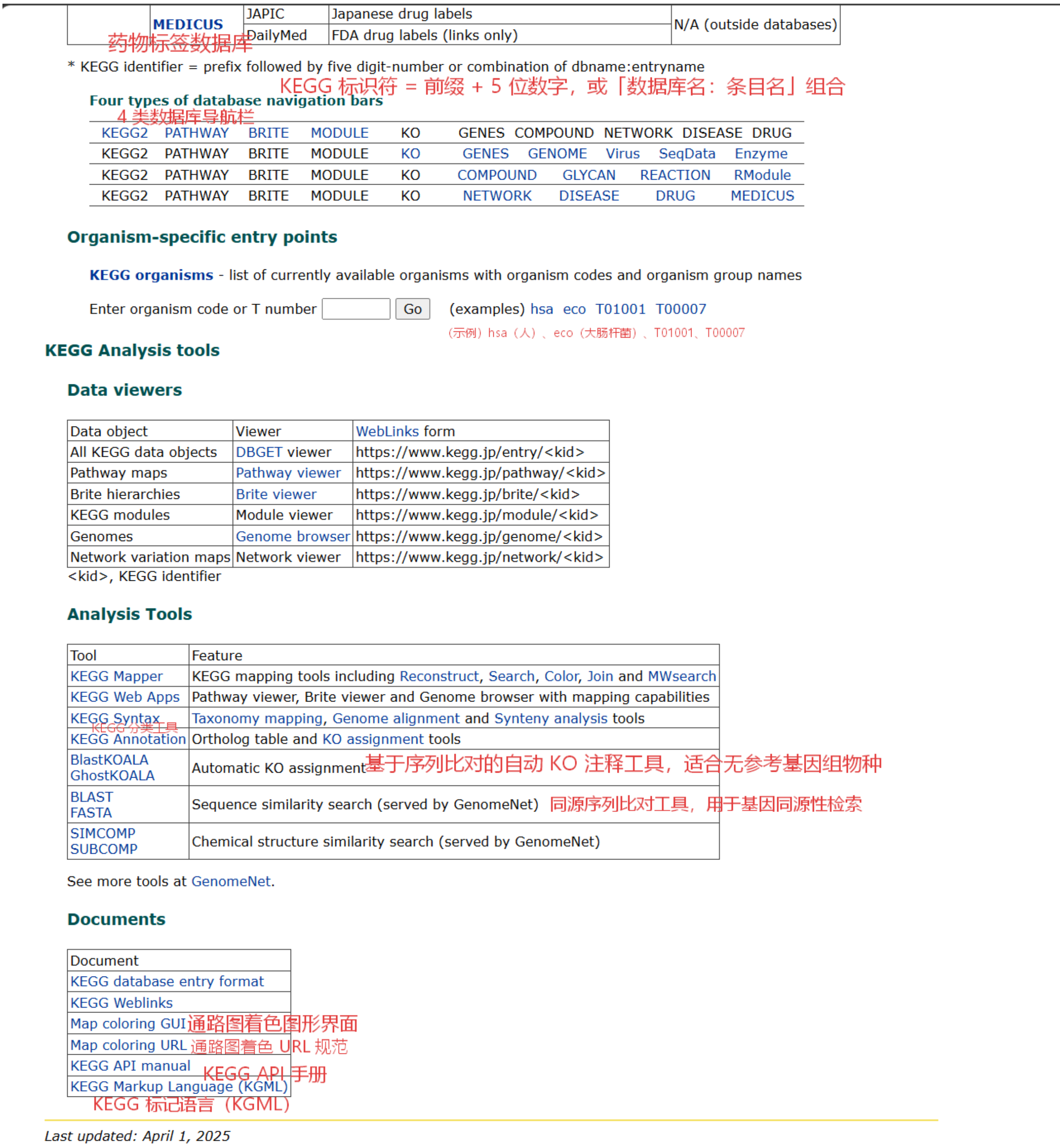
Data-oriented entry points

Category	Entry point	Database	Data object	KEGG identifier*
Systems information	**PATHWAY**	PATHWAY	KEGG pathway maps	map | ko, ec, rn, <org>
	BRITE	BRITE	Brite hierarchies Brite tables	br, jp, ko | <org> br, jp
	MODULE RModule	MODULE	Modules Reaction modules	M | <org>_M RM
Genomic information	**KO**	KO	KEGG Orthology (KO) for functional orthologs	K
	GENES Virus SeqData	GENES	KEGG organism genes and proteins Virus genes and proteins Virus mature peptides Functionally characterized proteins from literature	<org>:<gene> vg:<gene> vp:<gene-no> ag:<protein>
	GENOME	GENOME	KEGG organisms genomes KEGG viruses genomes	T, gn:<org> gn:<vtax>
Chemical information	**COMPOUND**	COMPOUND	Metabolites and other small molecules	C
	GLYCAN	GLYCAN	Glycans	G
	REACTION	REACTION RCLASS	Biochemical reactions Reaction class	R RC
	Enzyme	ENZYME	Enzyme nomenclature with sequence data	ec:<ecnum>
Health information	**NETWORK**	NETWORK VARIANT	Network variation maps Network elements Human gene variants	nt N hsa_var:<gene_vno>
	DISEASE	DISEASE	Human diseases	H
	DRUG	DRUG DGROUP	Drugs Drug groups	D DG
	MEDICUS	JAPIC	Japanese drug labels	N/A (outside databases)
		DailyMed	FDA drug labels (links only)	

Brite 层级体系 / Brite 表格

经文献功能验证的蛋白

带序列数据的酶命名



	MEDICUS	JAPIC	Japanese drug labels	N/A (outside databases)
		DailyMed	FDA drug labels (links only)	

药物标签数据库

* KEGG identifier = prefix followed by five digit-number or combination of dbname:entryname

KEGG 标识符 = 前缀 + 5 位数字，或「数据库名：条目名」组合

Four types of database navigation bars

4 类数据库导航栏

KEGG2	PATHWAY	BRITE	MODULE	KO	GENES	COMPOUND	NETWORK	DISEASE	DRUG
KEGG2	PATHWAY	BRITE	MODULE	KO	GENES	GENOME	Virus	SeqData	Enzyme
KEGG2	PATHWAY	BRITE	MODULE	KO	COMPOUND	GLYCAN	REACTION	RModule	
KEGG2	PATHWAY	BRITE	MODULE	KO	NETWORK	DISEASE	DRUG	MEDICUS	

Organism-specific entry points

KEGG organisms - list of currently available organisms with organism codes and organism group names

Enter organism code or T number [ ] Go (examples) hsa eco T01001 T00007

（示例）hsa（人）、eco（大肠杆菌）、T01001、T00007

KEGG Analysis tools

Data viewers

Data object	Viewer	WebLinks form
All KEGG data objects	DBGET viewer	https://www.kegg.jp/entry/<kid>
Pathway maps	Pathway viewer	https://www.kegg.jp/pathway/<kid>
Brite hierarchies	Brite viewer	https://www.kegg.jp/brite/<kid>
KEGG modules	Module viewer	https://www.kegg.jp/module/<kid>
Genomes	Genome browser	https://www.kegg.jp/genome/<kid>
Network variation maps	Network viewer	https://www.kegg.jp/network/<kid>

<kid>, KEGG identifier

Analysis Tools

Tool	Feature
KEGG Mapper	KEGG mapping tools including Reconstruct, Search, Color, Join and MWsearch
KEGG Web Apps	Pathway viewer, Brite viewer and Genome browser with mapping capabilities
KEGG Syntax	Taxonomy mapping, Genome alignment and Synteny analysis tools
KEGG Annotation	Ortholog table and KO assignment tools
BlastKOALA GhostKOALA	Automatic KO assignment
BLAST FASTA	Sequence similarity search (served by GenomeNet)
SIMCOMP SUBCOMP	Chemical structure similarity search (served by GenomeNet)

KEGG 分类工具

基于序列比对的自动 KO 注释工具，适合无参考基因组物种

同源序列比对工具，用于基因同源性检索

See more tools at GenomeNet.

Documents

Document
KEGG database entry format
KEGG Weblinks
Map coloring GUI 通路图着色图形界面
Map coloring URL 通路图着色 URL 规范
KEGG API manual KEGG API 手册
KEGG Markup Language (KGML)

KEGG 标记语言（KGML）

Last updated: April 1, 2025

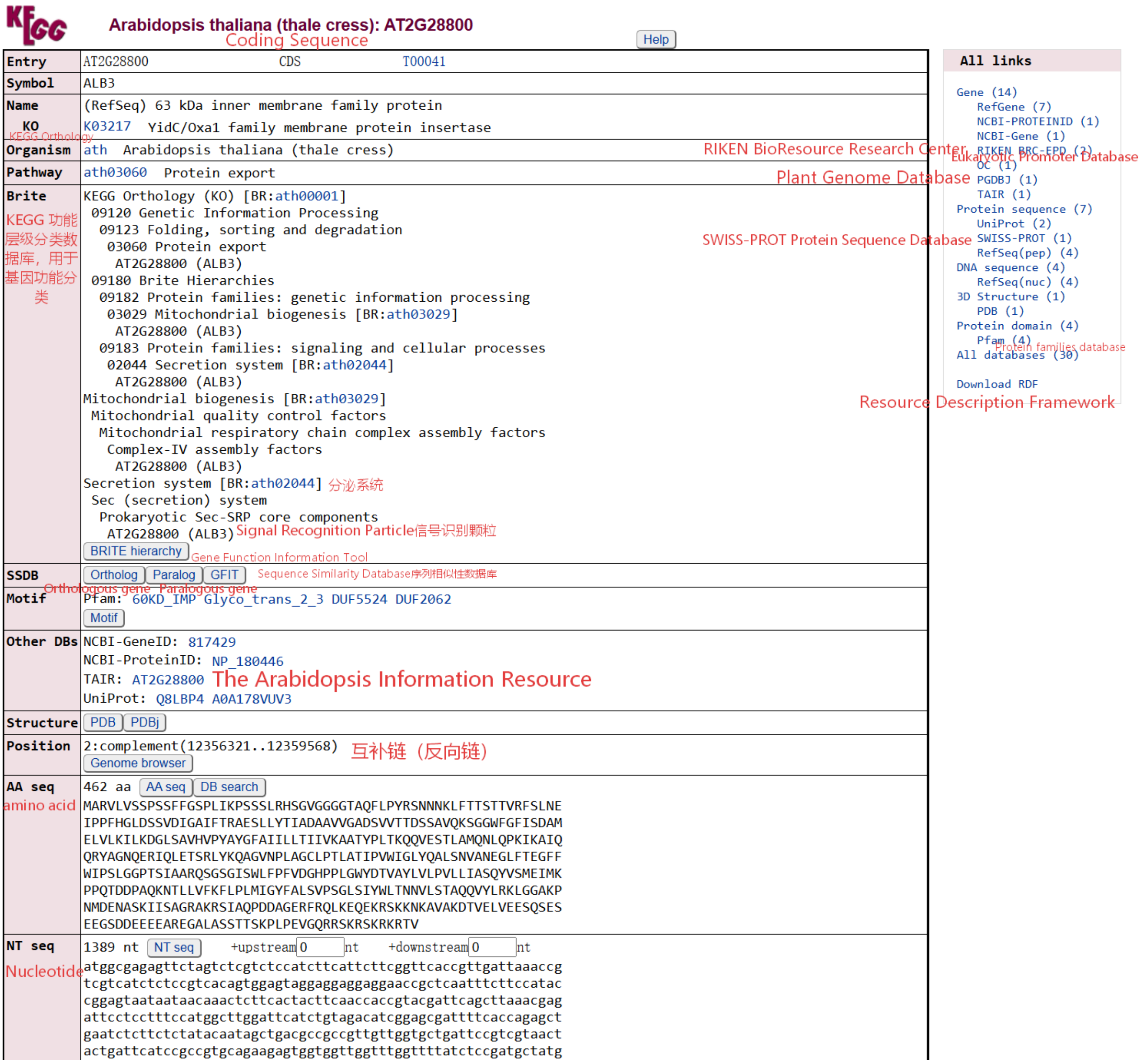
KEGG，Kyoto Encyclopedia of Genes and Genomes，京都基因与基因组百科全书，日本 Kanehisa 实验室（京都大学）开发，是全球最权威的生物信息学综合数据库之一，KEGG 是整合基因组、化学、系统功能与健康信息的一站式生物信息学平台，核心使命是从分子层面（基因、蛋白、代谢物）解析细胞、生物体、生态系统的高级功能，是基因组学、转录组学、代谢组学等多组学研究的核心工具，也是研究动植物基因功能、代谢通路的核心数据库。

URL：Uniform Resource Locator（统一资源定位符），即网址。

API：Application Programming Interface（应用程序编程接口），当你需要批量做富集分析、或用脚本处理数据时，就需要调用 KEGG API（通常需要查看 KEGG API manual）。

GUI：Graphical User Interface（图形用户界面）。

12　查看拟南芥内膜蛋白 ALB3_ARATH 在代谢通路数据库 KEGG 中的信息：



Arabidopsis thaliana (thale cress): AT2G28800

```
Entry      AT2G28800                CDS              T00041
Symbol     ALB3
Name       (RefSeq) 63 kDa inner membrane family protein
  KO       K03217  YidC/Oxa1 family membrane protein insertase
Organism   ath  Arabidopsis thaliana (thale cress)
Pathway    ath03060  Protein export
Brite      KEGG Orthology (KO) [BR:ath00001]
            09120 Genetic Information Processing
             09123 Folding, sorting and degradation
              03060 Protein export
               AT2G28800 (ALB3)
            09180 Brite Hierarchies
             09182 Protein families: genetic information processing
              03029 Mitochondrial biogenesis [BR:ath03029]
               AT2G28800 (ALB3)
             09183 Protein families: signaling and cellular processes
              02044 Secretion system [BR:ath02044]
               AT2G28800 (ALB3)
           Mitochondrial biogenesis [BR:ath03029]
            Mitochondrial quality control factors
             Mitochondrial respiratory chain complex assembly factors
              Complex-IV assembly factors
               AT2G28800 (ALB3)
           Secretion system [BR:ath02044]
            Sec (secretion) system
             Prokaryotic Sec-SRP core components
              AT2G28800 (ALB3)
SSDB
Motif      Pfam: 60KD_IMP Glyco_trans_2_3 DUF5524 DUF2062
Other DBs  NCBI-GeneID: 817429
           NCBI-ProteinID: NP_180446
           TAIR: AT2G28800
           UniProt: Q8LBP4 A0A178VUV3
Structure
Position   2:complement(12356321..12359568)
AA seq     462 aa
           MARVLVSSPSSFFGSPLIKPSSSLRHSGVGGGGTAQFLPYRSNNNKLFTTSTTVRFSLNE
           IPPFHGLDSSVDIGAIFTRAESLLYTIADAAVVGADSVVTTDSSAVQKSGGWFGFISDAM
           ELVLKILKDGLSAVHVPYAYGFAIILLTIIVKAATYPLTKQQVESTLAMQNLQPKIKAIQ
           QRYAGNQERIQLETSRLYKQAGVNPLAGCLPTLATIPVWIGLYQALSNVANEGLFTEGFF
           WIPSLGGPTSIAARQSGSGISWLFPFVDGHPPLGWYDTVAYLVLPVLLIASQYVSMEIMK
           PPQTDDPAQKNTLLVFKFLPLMIGYFALSVPSGLSIYWLTNNVLSTAQQVYLRKLGGAKP
           NMDENASKIISAGRAKRSIAQPDDAGERFRQLKEQEKRSKKNKAVAKDTVELVEESQSES
           EEGSDDEEEEAREGALASSTTSKPLPEVGQRRSKRSKRKRTV
NT seq     1389 nt        +upstream 0   nt    +downstream 0   nt
           atggcgagagttctagtctcgtctccatcttcattcttcggttcaccgttgattaaaccg
           tcgtcatctctccgtcacagtggagtaggaggaggaggaaccgctcaatttcttccatac
           cggagtaataataacaaactcttcactacttcaaccaccgtacgattcagcttaaacgag
           attcctcctttccatggcttggattcatctgtagacatcggagcgattttcaccagagct
           gaatctcttctctatacaatagctgacgccgccgttgttggtgctgattccgtcgtaact
           actgattcatccgccgtgcagaagagtggtggttggtttggttttatctccgatgctatg
```

All links

- Gene (14)
  - RefGene (7)
  - NCBI-PROTEINID (1)
  - NCBI-Gene (1)
  - RIKEN BRC-EPD (2)
  - OC (1)
  - PGDBJ (1)
  - TAIR (1)
- Protein sequence (7)
  - UniProt (2)
  - SWISS-PROT (1)
  - RefSeq(pep) (4)
- DNA sequence (4)
  - RefSeq(nuc) (4)
- 3D Structure (1)
  - PDB (1)
- Protein domain (4)
  - Pfam (4)
- All databases (30)

Download RDF

13　查看拟南芥内膜蛋白 ALB3_ARATH 在蛋白质结构数据库 PDB 中的信息

PDBe is a founding member of the Worldwide Protein Data Bank (wwPDB) which collects, organises and disseminates data on biological macromolecular structures. wwPDB Partners: RCSB PDB, PDBj, BMRB, EMDB. Read more about PDBe.

Services

Latest archive statistics

Total PDB entries in the archive: 251801

Latest Data:

379 Latest Entries

62 Latest Chemistry

0 Latest Biology

Featured

Antimicrobial resistance: The global burden

2025-11-03

A post-antibiotic era looms

Latest news

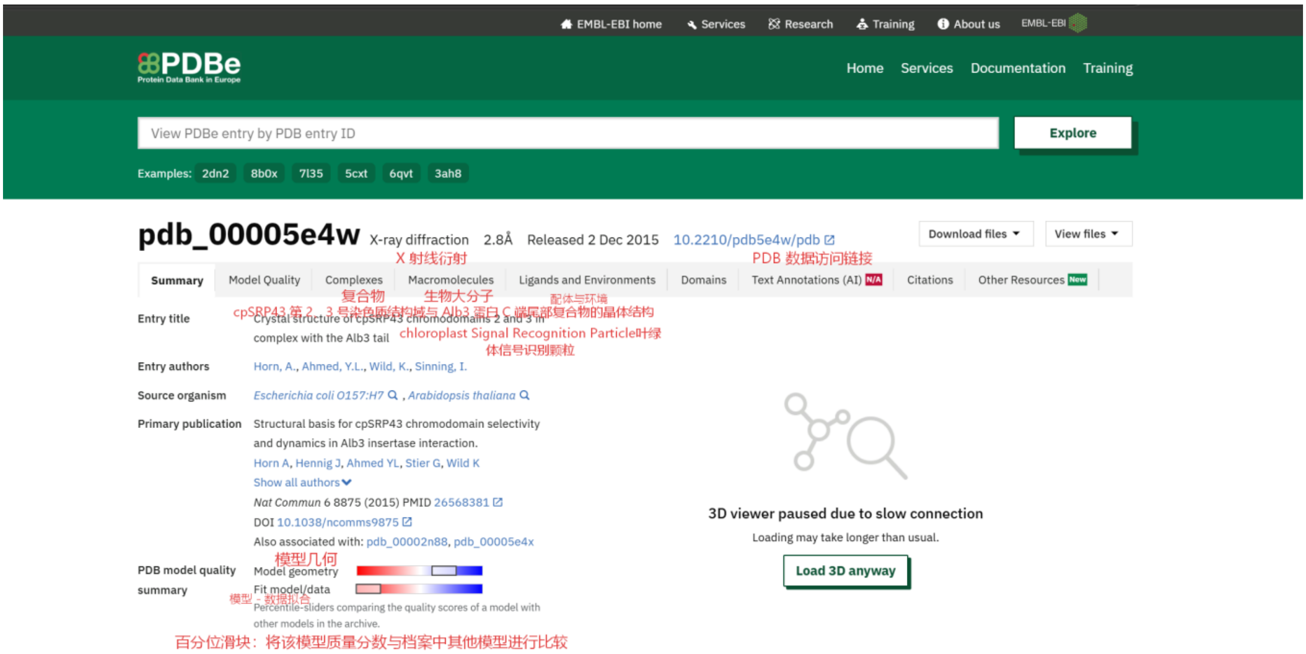
Poster Prize Awarded at ASBMB

2026-03-31

The wwPDB Foundation made an award for an outstanding student presentation at the 2026 meeting...

Release of the PDBe mmCIF

Protein Data Bank in Europe，欧洲蛋白质数据库，属欧洲分子生物学实验室-欧洲生物信息研究所（EMBL-EBI），全球蛋白质数据银行（wwPDB, Worldwide Protein Data Bank）的创始成员之一，与美国 RCSB PDB、日本 PDBj、BMRB 共同维护全球唯一的生物大分子结构档案（PDB），旨在收集、整理、分发全球所有生物大分子（蛋白质、核酸、复合物等）的实验解析三维结构数据。



pdb_00005e4w X-ray diffraction 2.8Å Released 2 Dec 2015 10.2210/pdb5e4w/pdb

Entry title: Crystal structure of cpSRP43 chromodomains 2 and 3 in complex with the Alb3 tail

cpSRP43 第 2、3 号染色质结构域与 Alb3 蛋白 C 端尾部复合物的晶体结构

chloroplast Signal Recognition Particle叶绿体信号识别颗粒

Entry authors: Horn, A., Ahmed, Y.L., Wild, K., Sinning, I.

Source organism: *Escherichia coli O157:H7*, *Arabidopsis thaliana*

Primary publication: Structural basis for cpSRP43 chromodomain selectivity and dynamics in Alb3 insertase interaction.

Horn A, Hennig J, Ahmed YL, Stier G, Wild K

Show all authors

Nat Commun 6 8875 (2015) PMID 26568381

DOI 10.1038/ncomms9875

Also associated with: pdb_00002n88, pdb_00005e4x

PDB model quality summary: Model geometry; Fit model/data

Percentile-sliders comparing the quality scores of a model with other models in the archive.

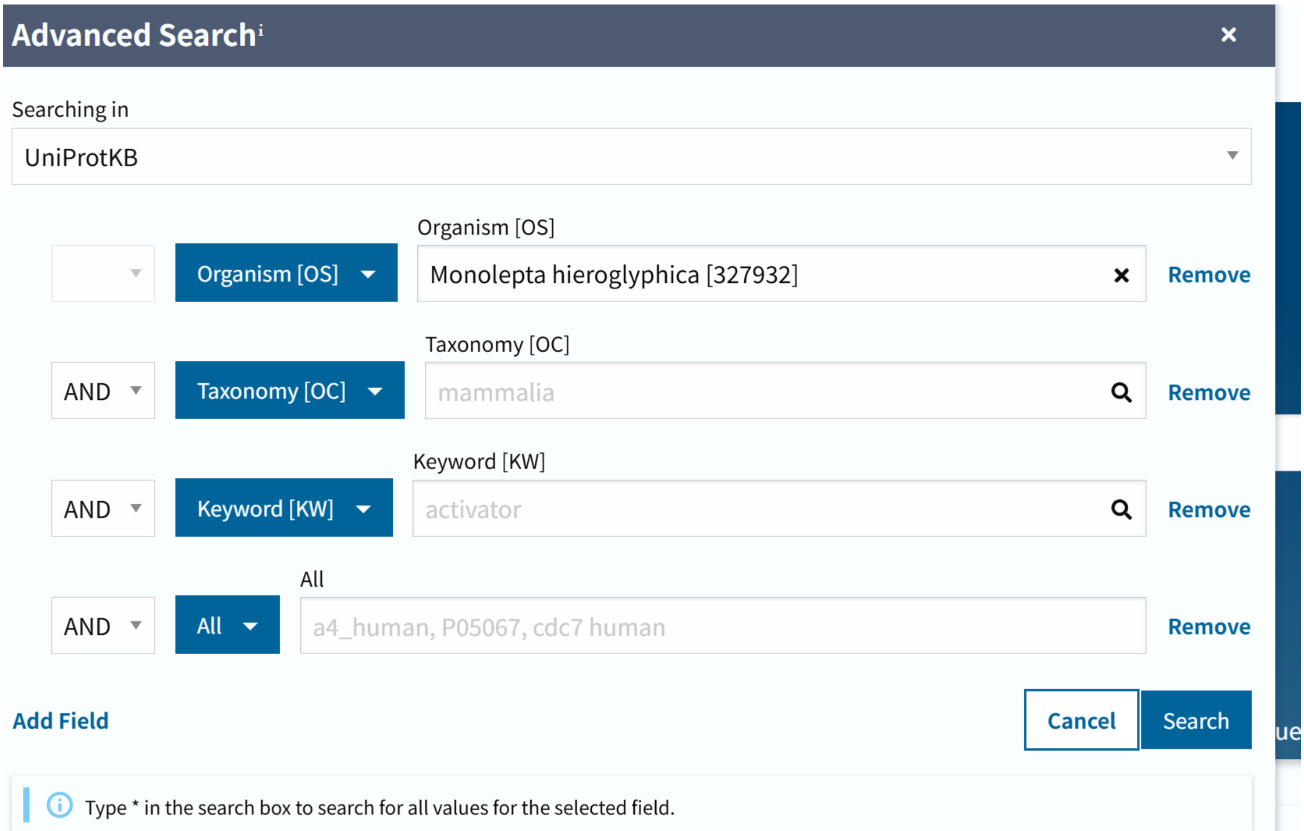
百分位滑块：将该模型质量分数与档案中其他模型进行比较

3D viewer paused due to slow connection

Loading may take longer than usual.

14　查看双斑长跗萤叶甲细胞色素 c 氧化酶亚基 I A0A891T1X2_9CUCU 在 NCBI 参考序列数据库 RefSeq 中的核酸和蛋白质序列信息（以 A0A891T1X2_9CUCU 基因为例）

打开 Uniprot，advance 搜索，下拉选项为 organism[OS]，输入物种名称双斑长跗萤叶甲拉丁文 *Monolepta hieroglyphica*，点击 Search，可搜索物种的 entries。


Searching in: UniProtKB

Organism [OS]: Monolepta hieroglyphica [327932]

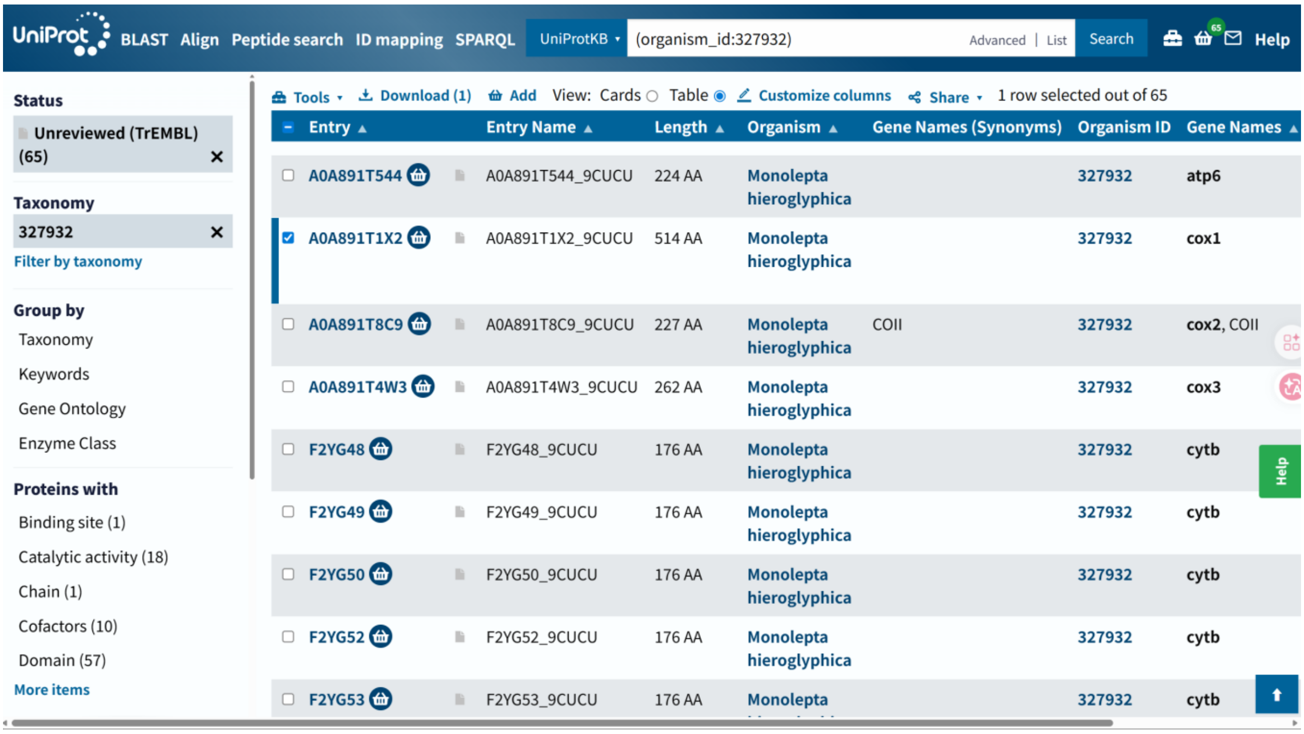
AND Taxonomy [OC]: mammalia

AND Keyword [KW]: activator

AND All: a4_human, P05067, cdc7 human

Type * in the search box to search for all values for the selected field.

找到双斑长跗萤叶甲（*Monolepta hieroglyphica*）的 A0A891T1X2_9CUCU，细胞色素 c 氧化酶亚基 I（cox1/COI）。



(organism_id:327932)

View: Cards ○ Table ●

1 row selected out of 65

Status

Unreviewed (TrEMBL) (65)

Taxonomy

327932

Group by

Taxonomy

Keywords

Gene Ontology

Enzyme Class

Proteins with

Binding site (1)

Catalytic activity (18)

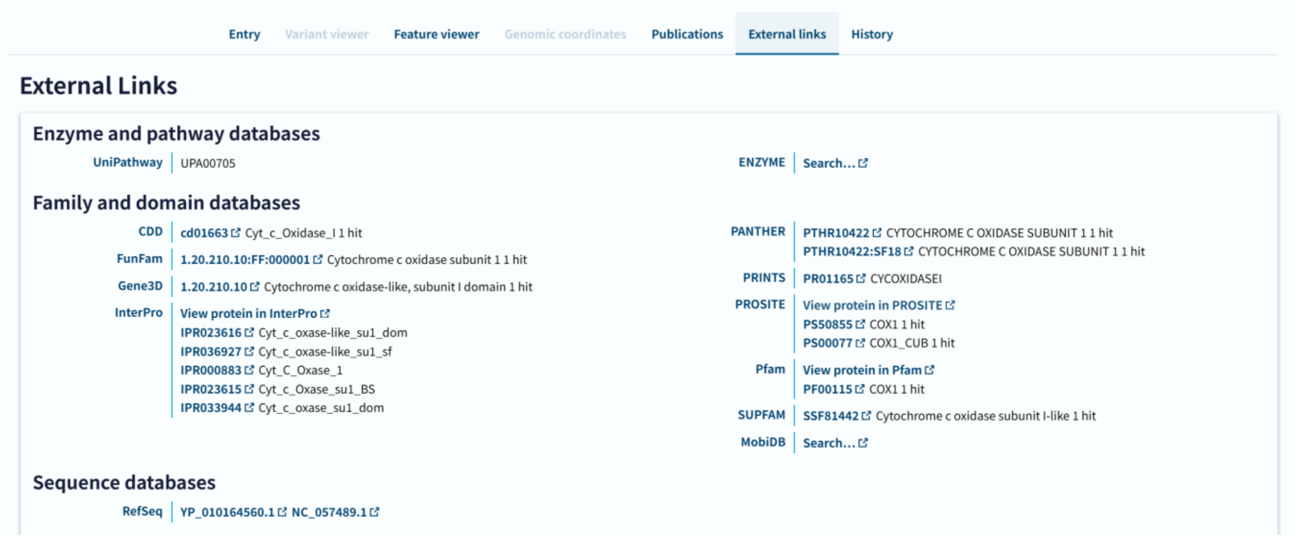
Chain (1)

Cofactors (10)

Domain (57)

	Entry	Entry Name	Length	Organism	Gene Names (Synonyms)	Organism ID	Gene Names
☐	A0A891T544	A0A891T544_9CUCU	224 AA	Monolepta hieroglyphica		327932	atp6
☑	A0A891T1X2	A0A891T1X2_9CUCU	514 AA	Monolepta hieroglyphica		327932	cox1
☐	A0A891T8C9	A0A891T8C9_9CUCU	227 AA	Monolepta hieroglyphica	COII	327932	cox2, COII
☐	A0A891T4W3	A0A891T4W3_9CUCU	262 AA	Monolepta hieroglyphica		327932	cox3
☐	F2YG48	F2YG48_9CUCU	176 AA	Monolepta hieroglyphica		327932	cytb
☐	F2YG49	F2YG49_9CUCU	176 AA	Monolepta hieroglyphica		327932	cytb
☐	F2YG50	F2YG50_9CUCU	176 AA	Monolepta hieroglyphica		327932	cytb
☐	F2YG52	F2YG52_9CUCU	176 AA	Monolepta hieroglyphica		327932	cytb
☐	F2YG53	F2YG53_9CUCU	176 AA	Monolepta		327932	cytb

点击 External Links，可链接到其他数据库。



External Links

Enzyme and pathway databases

UniPathway | UPA00705

ENZYME | Search...

Family and domain databases

CDD | cd01663 Cyt_c_Oxidase_I 1 hit

FunFam | 1.20.210.10:FF:000001 Cytochrome c oxidase subunit 1 1 hit

Gene3D | 1.20.210.10 Cytochrome c oxidase-like, subunit I domain 1 hit

InterPro | View protein in InterPro
IPR023616 Cyt_c_oxase-like_su1_dom
IPR036927 Cyt_c_oxase-like_su1_sf
IPR000883 Cyt_C_Oxase_1
IPR023615 Cyt_c_Oxase_su1_BS
IPR033944 Cyt_c_oxase_su1_dom

PANTHER | PTHR10422 CYTOCHROME C OXIDASE SUBUNIT 1 1 hit
PTHR10422:SF18 CYTOCHROME C OXIDASE SUBUNIT 1 1 hit

PRINTS | PR01165 CYCOXIDASEI

PROSITE | View protein in PROSITE
PS50855 COX1 1 hit
PS00077 COX1_CUB 1 hit

Pfam | View protein in Pfam
PF00115 COX1 1 hit

SUPFAM | SSF81442 Cytochrome c oxidase subunit I-like 1 hit

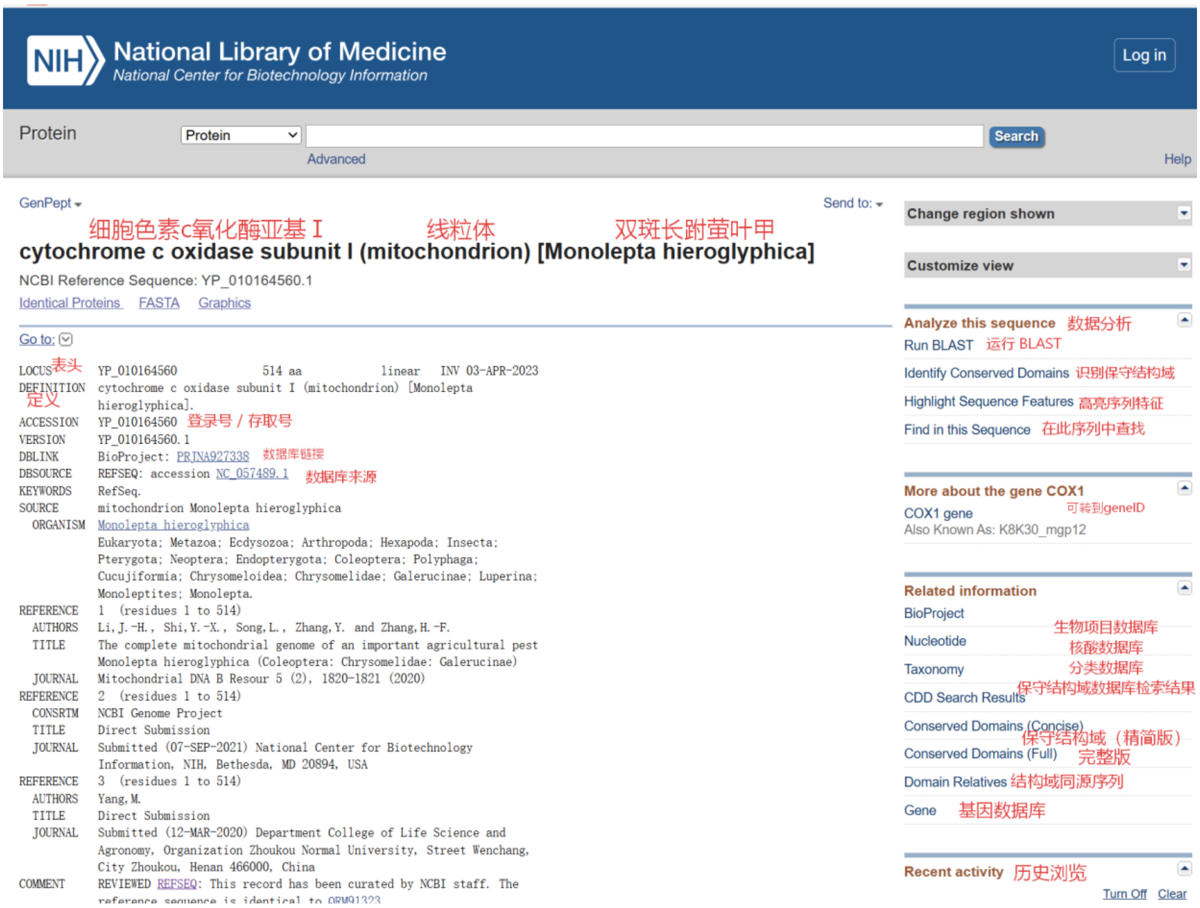
MobiDB | Search...

Sequence databases

RefSeq | YP_010164560.1 NC_057489.1

RefSeq | YP_010164560.1 NC_057489.1

点击 YP_010164560.1，可链接到 RefSeq 的蛋白质数据库；

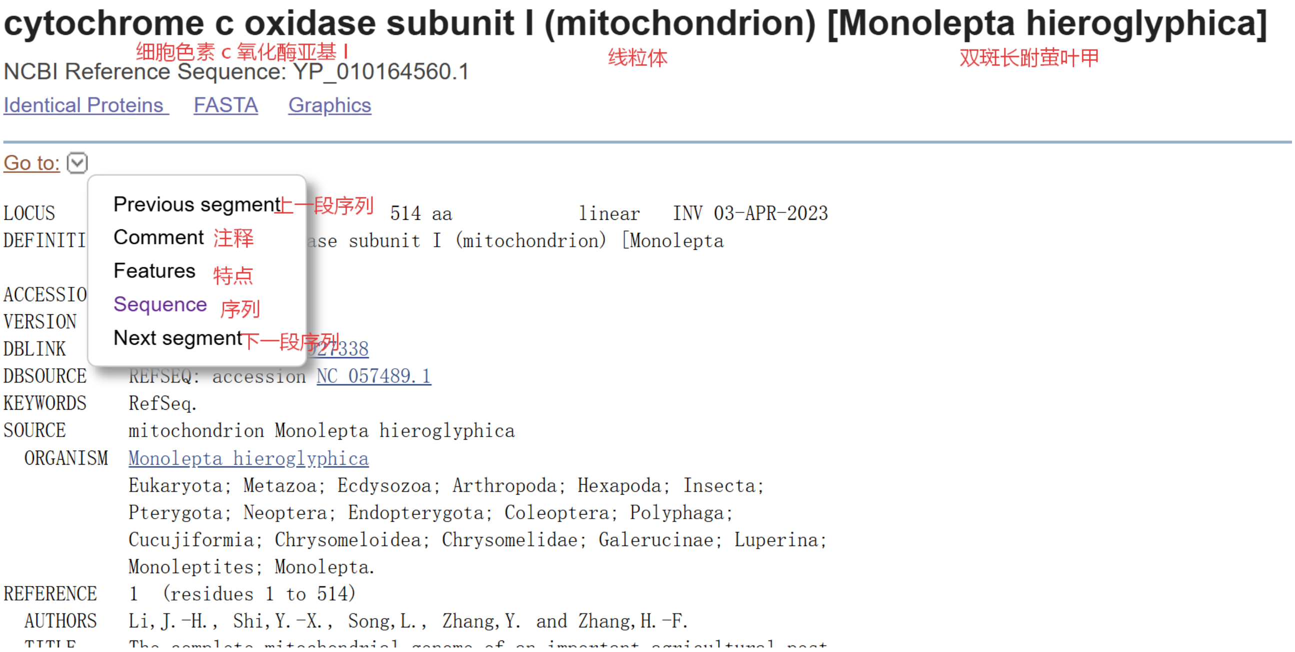


cytochrome c oxidase subunit I (mitochondrion) [Monolepta hieroglyphica]

NCBI Reference Sequence: YP_010164560.1

```
LOCUS       YP_010164560             514 aa            linear   INV 03-APR-2023
DEFINITION  cytochrome c oxidase subunit I (mitochondrion) [Monolepta
            hieroglyphica].
ACCESSION   YP_010164560
VERSION     YP_010164560.1
DBLINK      BioProject: PRJNA927338
DBSOURCE    REFSEQ: accession NC_057489.1
KEYWORDS    RefSeq.
SOURCE      mitochondrion Monolepta hieroglyphica
  ORGANISM  Monolepta hieroglyphica
            Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
            Cucujiformia; Chrysomeloidea; Chrysomelidae; Galerucinae; Luperina;
            Monoleptites; Monolepta.
REFERENCE   1  (residues 1 to 514)
  AUTHORS   Li,J.-H., Shi,Y.-X., Song,L., Zhang,Y. and Zhang,H.-F.
  TITLE     The complete mitochondrial genome of an important agricultural pest
            Monolepta hieroglyphica (Coleoptera: Chrysomelidae: Galerucinae)
  JOURNAL   Mitochondrial DNA B Resour 5 (2), 1820-1821 (2020)
REFERENCE   2  (residues 1 to 514)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2021) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 514)
  AUTHORS   Yang,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2020) Department College of Life Science and
            Agronomy, Organization Zhoukou Normal University, Street Wenchang,
            City Zhoukou, Henan 466000, China
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence is identical to QRM91323
```

页面左边信息包含：基础信息如序列条目、定义、登录号、版本、数据库链接、数据库来源、参考文献等，上一段序列，注释，结构特点，序列，下一段序列；页面右边信息，整合了序列分析、基因信息、相关信息与浏览历史四大模块：提供 BLAST、保守结构域分析等序列分析工具；



cytochrome c oxidase subunit I (mitochondrion) [Monolepta hieroglyphica]

NCBI Reference Sequence: YP_010164560.1

```
LOCUS                    514 aa            linear   INV 03-APR-2023
DEFINITI             ase subunit I (mitochondrion) [Monolepta

ACCESSIO
VERSION
DBLINK                    27338
DBSOURCE    REFSEQ: accession NC_057489.1
KEYWORDS    RefSeq.
SOURCE      mitochondrion Monolepta hieroglyphica
  ORGANISM  Monolepta hieroglyphica
            Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
            Cucujiformia; Chrysomeloidea; Chrysomelidae; Galerucinae; Luperina;
            Monoleptites; Monolepta.
REFERENCE   1  (residues 1 to 514)
  AUTHORS   Li,J.-H., Shi,Y.-X., Song,L., Zhang,Y. and Zhang,H.-F.
  TITLE     The complete mitochondrial genome of an important agricultural pest
```

点击 NC_057489.1 则可链接到核酸数据库；页面整合了基因组的完整元数据（包括物种分类、参考文献、序列属性等），左侧 Go to 菜单可快速跳转至序列、注释、特征，序列等内容，右侧提供 BLAST 比对、引物设计、特征标注等分析工具，并关联了蛋白、分类学、基因等跨数据库资源。

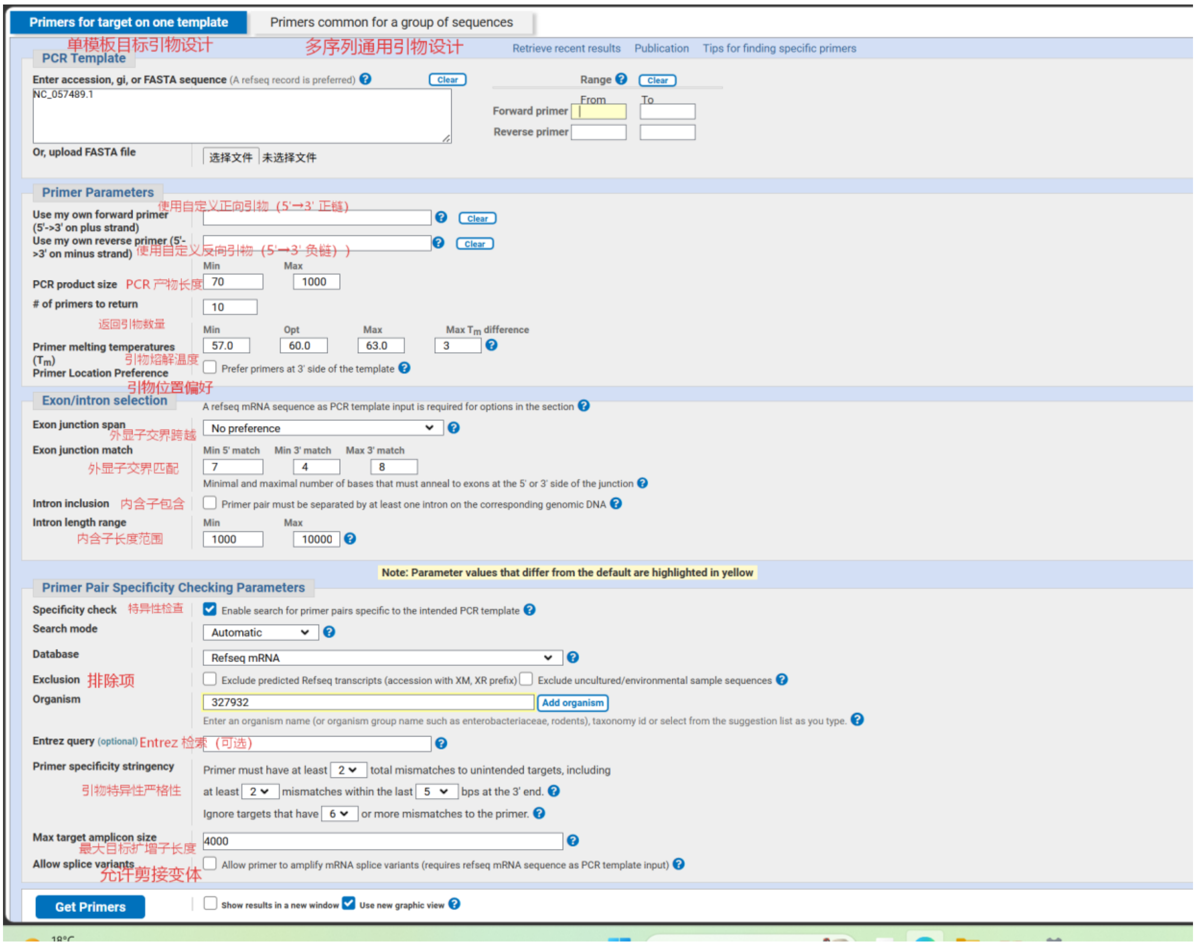
Monolepta hieroglyphica mitochondrion, complete genome

NCBI Reference Sequence: NC_057489.1

```
LOCUS       NC_057489               15963 bp    DNA     circular INV 03-APR-2023
DEFINITION  Monolepta hieroglyphica mitochondrion, complete genome.
ACCESSION   NC_057489
VERSION     NC_057489.1
DBLINK      BioProject: PRJNA927338
KEYWORDS    RefSeq.
SOURCE      mitochondrion Monolepta hieroglyphica
  ORGANISM  Monolepta hieroglyphica
            Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
            Cucujiformia; Chrysomeloidea; Chrysomelidae; Galerucinae; Luperina;
            Monoleptites; Monolepta.
REFERENCE   1  (bases 1 to 15963)
  AUTHORS   Li,J.-H., Shi,Y.-X., Song,L., Zhang,Y. and Zhang,H.-F.
  TITLE     The complete mitochondrial genome of an important agricultural pest
            Monolepta hieroglyphica (Coleoptera: Chrysomelidae: Galerucinae)
  JOURNAL   Mitochondrial DNA B Resour 5 (2), 1820-1821 (2020)
REFERENCE   2  (bases 1 to 15963)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2021) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (bases 1 to 15963)
  AUTHORS   Yang,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2020) Department College of Life Science and
            Agronomy, Organization Zhoukou Normal University, Street Wenchang,
            City Zhoukou, Henan 466000, China
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence is identical to MT192098.

            ##Assembly-Data-START##
            Assembly Method       :: A5-miseq v. v50150522; SPAdes v. v3.9.0
```

点击 Pick Primers 可设计引物。


Primers for target on one template 单模板目标引物设计 | Primers common for a group of sequences 多序列通用引物设计

PCR Template

Enter accession, gi, or FASTA sequence (A refseq record is preferred): NC_057489.1

Or, upload FASTA file: 选择文件 未选择文件

Range — Forward primer: From / To; Reverse primer: From / To

Primer Parameters

Use my own forward primer (5'->3' on plus strand) 使用自定义正向引物（5'→3' 正链）

Use my own reverse primer (5'->3' on minus strand) 使用自定义反向引物（5'→3' 负链））

PCR product size PCR 产物长度: Min 70, Max 1000

# of primers to return 返回引物数量: 10

Primer melting temperatures (Tm) 引物熔解温度: Min 57.0, Opt 60.0, Max 63.0, Max Tm difference 3

Primer Location Preference 引物位置偏好: ☐ Prefer primers at 3' side of the template

Exon/intron selection — A refseq mRNA sequence as PCR template input is required for options in the section

Exon junction span 外显子交界跨越: No preference

Exon junction match 外显子交界匹配: Min 5' match 7, Min 3' match 4, Max 3' match 8 — Minimal and maximal number of bases that must anneal to exons at the 5' or 3' side of the junction

Intron inclusion 内含子包含: ☐ Primer pair must be separated by at least one intron on the corresponding genomic DNA

Intron length range 内含子长度范围: Min 1000, Max 10000

Note: Parameter values that differ from the default are highlighted in yellow

Primer Pair Specificity Checking Parameters

Specificity check 特异性检查: ☑ Enable search for primer pairs specific to the intended PCR template

Search mode: Automatic

Database: Refseq mRNA

Exclusion 排除项: ☐ Exclude predicted Refseq transcripts (accession with XM, XR prefix) ☐ Exclude uncultured/environmental sample sequences

Organism: 327932 — Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy id or select from the suggestion list as you type.

Entrez query (optional) Entrez 检索（可选）

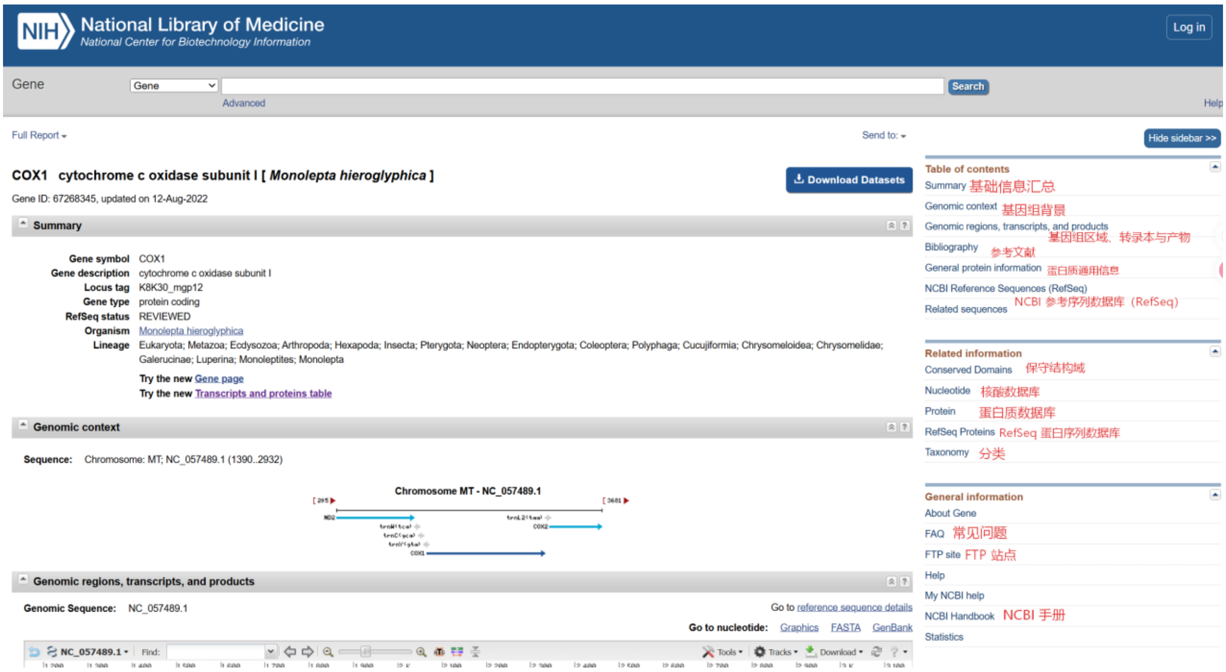
Primer specificity stringency 引物特异性严格性: Primer must have at least 2 total mismatches to unintended targets, including at least 2 mismatches within the last 5 bps at the 3' end. Ignore targets that have 6 or more mismatches to the primer.

Max target amplicon size 最大目标扩增子长度: 4000

Allow splice variants 允许剪接变体: ☐ Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input)

Get Primers ☐ Show results in a new window ☑ Use new graphic view

15　查看双斑长跗萤叶甲细胞色素 c 氧化酶亚基 I A0A891T1X2_9CUCU 在 NCBI 基因数据库 GeneID 中的信息


National Library of Medicine
National Center for Biotechnology Information

COX1 cytochrome c oxidase subunit I [*Monolepta hieroglyphica*]

Gene ID: 67268345, updated on 12-Aug-2022

Summary

Gene symbol COX1
Gene description cytochrome c oxidase subunit I
Locus tag K8K30_mgp12
Gene type protein coding
RefSeq status REVIEWED
Organism Monolepta hieroglyphica
Lineage Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Chrysomeloidea; Chrysomelidae; Galerucinae; Luperina; Monoleptites; Monolepta
Try the new Gene page
Try the new Transcripts and proteins table

Genomic context

Sequence: Chromosome: MT; NC_057489.1 (1390..2932)

Genomic regions, transcripts, and products

Genomic Sequence: NC_057489.1

Go to reference sequence details
Go to nucleotide: Graphics FASTA GenBank

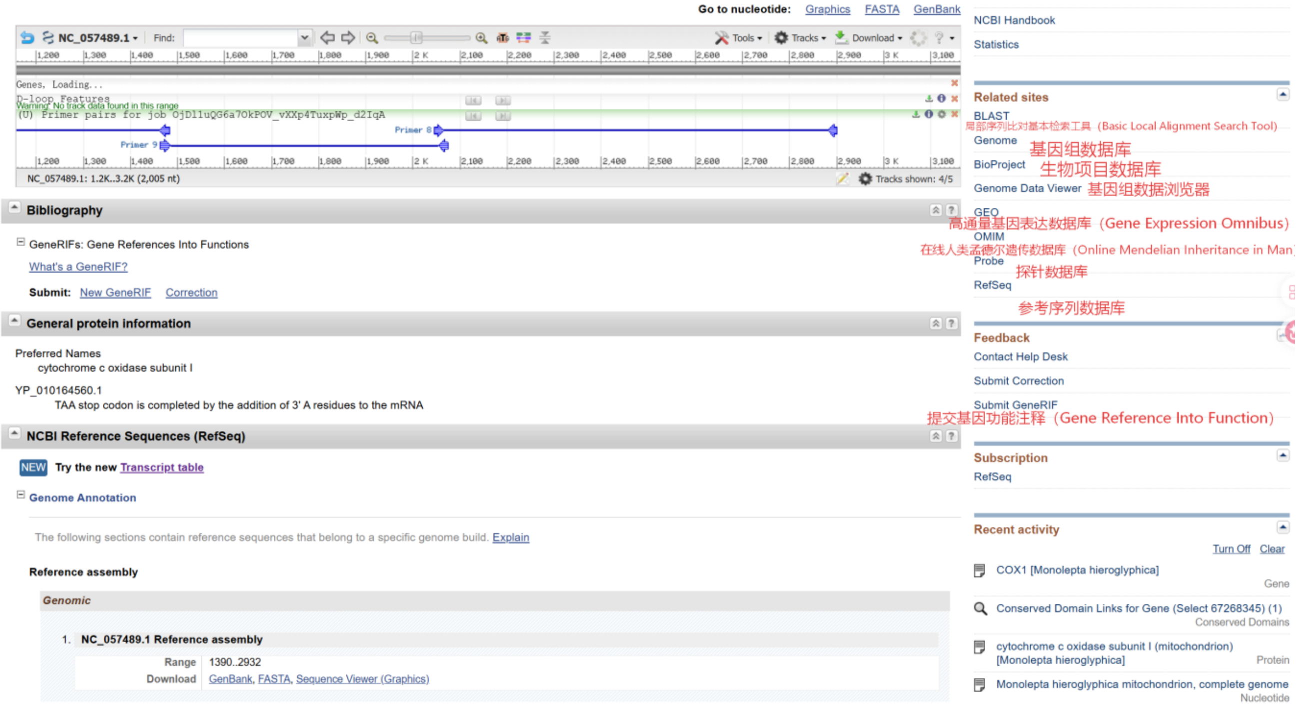
Table of contents: Summary 基础信息汇总; Genomic context 基因组背景; Genomic regions, transcripts, and products 基因组区域、转录本与产物; Bibliography 参考文献; General protein information 蛋白质通用信息; NCBI Reference Sequences (RefSeq) NCBI 参考序列数据库（RefSeq）; Related sequences

Related information: Conserved Domains 保守结构域; Nucleotide 核酸数据库; Protein 蛋白质数据库; RefSeq Proteins RefSeq 蛋白序列数据库; Taxonomy 分类

General information: About Gene; FAQ 常见问题; FTP site FTP 站点; Help; My NCBI help; NCBI Handbook NCBI 手册; Statistics

页面整合了 COX1 的基础信息（基因符号、功能描述、分类谱系、RefSeq

审核状态等），Genomic context 描述基因在其所在基因组上的位置、相邻基因、上下游序列、结构排布等全部周边环境信息，Genomic regions, transcripts, and products 指转录本与蛋白对应关系，参考文献，保守结构域等核心信息，右侧提供 BLAST、序列下载、跨数据库跳转（核酸、蛋白、分类学等）等工具。



Bibliography

GeneRIFs: Gene References Into Functions

What's a GeneRIF?

Submit: New GeneRIF Correction

General protein information

Preferred Names
cytochrome c oxidase subunit I

YP_010164560.1
TAA stop codon is completed by the addition of 3' A residues to the mRNA

NCBI Reference Sequences (RefSeq)

NEW Try the new Transcript table

Genome Annotation

The following sections contain reference sequences that belong to a specific genome build. Explain

Reference assembly

Genomic

1. **NC_057489.1 Reference assembly**

Range	1390..2932
Download	GenBank, FASTA, Sequence Viewer (Graphics)

COX1 [Monolepta hieroglyphica] — Gene

Conserved Domain Links for Gene (Select 67268345) (1) — Conserved Domains

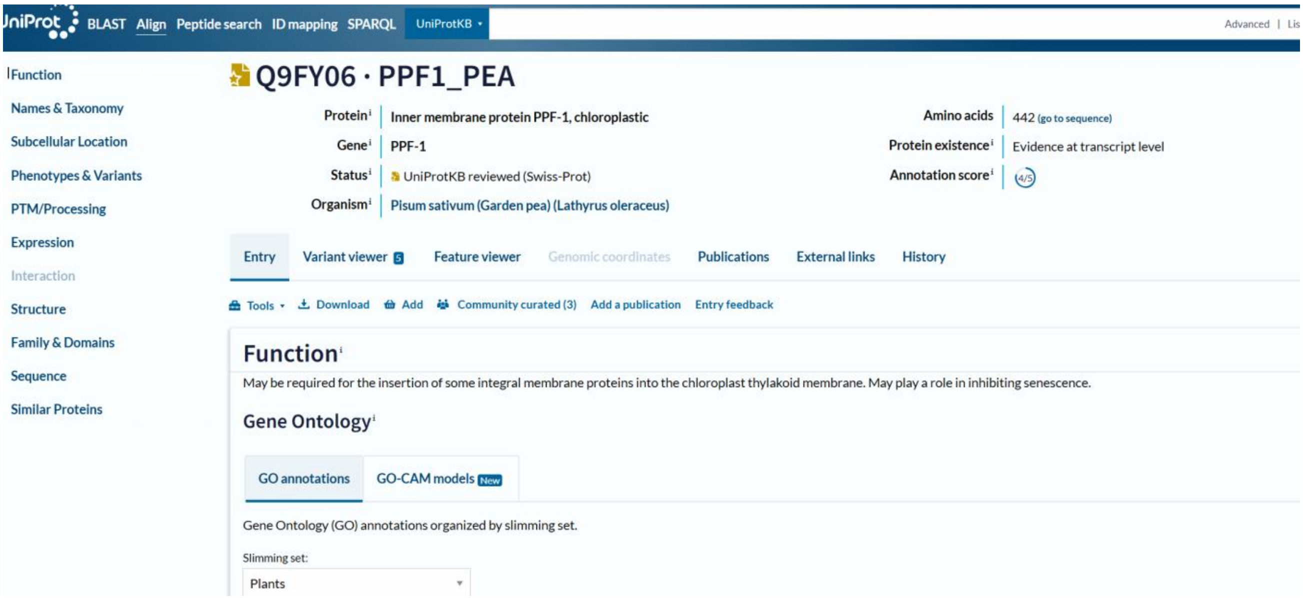
cytochrome c oxidase subunit I (mitochondrion) [Monolepta hieroglyphica] — Protein

Monolepta hieroglyphica mitochondrion, complete genome — Nucleotide

C UniProt 文献信息

（1）UniProt 查看文献信息

打开序列条目：PPF1_PEA



Q9FY06 · PPF1_PEA

Protein	Inner membrane protein PPF-1, chloroplastic	Amino acids	442 (go to sequence)
Gene	PPF-1	Protein existence	Evidence at transcript level
Status	UniProtKB reviewed (Swiss-Prot)	Annotation score	4/5
Organism	Pisum sativum (Garden pea) (Lathyrus oleraceus)		

Function

May be required for the insertion of some integral membrane proteins into the chloroplast thylakoid membrane. May play a role in inhibiting senescence.

Gene Ontology

Gene Ontology (GO) annotations organized by slimming set.

Slimming set: Plants

点击 Publication 查看文献信息：Publications

Publications for Q9FY06

PPF-1, a post-floral-specific gene expressed in short-day-grown G2 pea, may be important for its never-senescing phenotype.

Zhu Y., Zhang Y., Luo J., Davies P.J., Ho D.T.-H.

View abstract

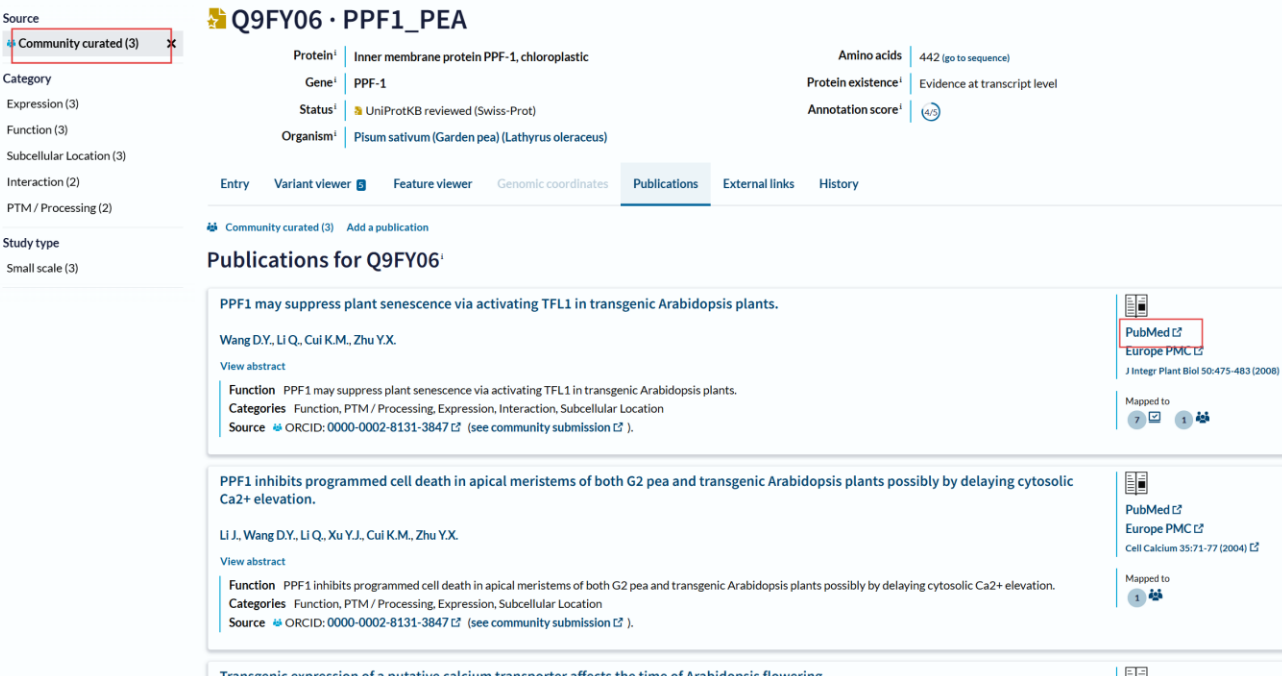
Cited for NUCLEOTIDE SEQUENCE [MRNA]
Strain cv. G2
Tissue Apical bud
Category Sequences
Source UniProtKB reviewed (Swiss-Prot)

Cloning and sequencing of the homologous gene of PPF-1 from Alaska pea.

Xu Y., Wang M., Li Q., Ji X., Zhu Y.

Cited for NUCLEOTIDE SEQUENCE [MRNA], VARIANTS GLY-107; GLY-127; PRO-150; PRO-250 AND 326-GLN-GLN-327 DELINS HIS-LYS
Strain cv. Alaska
Category Sequences

查看文献文献来源，找出用户审编的 3 篇文献：Community curated



Q9FY06 · PPF1_PEA

Protein Inner membrane protein PPF-1, chloroplastic
Gene PPF-1
Status UniProtKB reviewed (Swiss-Prot)
Organism Pisum sativum (Garden pea) (Lathyrus oleraceus)
Amino acids 442 (go to sequence)
Protein existence Evidence at transcript level
Annotation score 4/5

Publications for Q9FY06

PPF1 may suppress plant senescence via activating TFL1 in transgenic Arabidopsis plants.

Wang D.Y., Li Q., Cui K.M., Zhu Y.X.

View abstract

Function PPF1 may suppress plant senescence via activating TFL1 in transgenic Arabidopsis plants.
Categories Function, PTM / Processing, Expression, Interaction, Subcellular Location
Source ORCID: 0000-0002-8131-3847 (see community submission).

PubMed
Europe PMC
J Integr Plant Biol 50:475-483 (2008)
Mapped to
7 1

PPF1 inhibits programmed cell death in apical meristems of both G2 pea and transgenic Arabidopsis plants possibly by delaying cytosolic Ca2+ elevation.

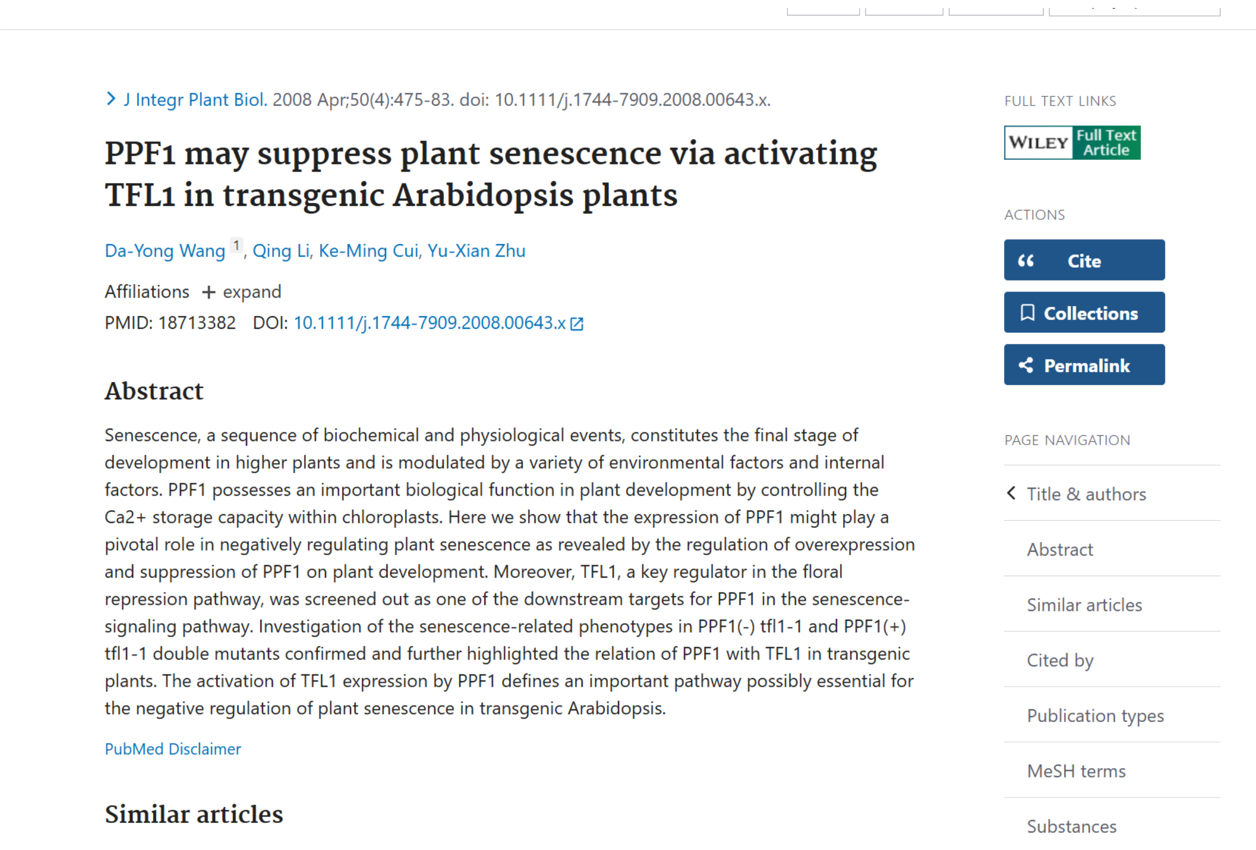
Li J., Wang D.Y., Li Q., Xu Y.J., Cui K.M., Zhu Y.X.

View abstract

Function PPF1 inhibits programmed cell death in apical meristems of both G2 pea and transgenic Arabidopsis plants possibly by delaying cytosolic Ca2+ elevation.
Categories Function, PTM / Processing, Expression, Subcellular Location
Source ORCID: 0000-0002-8131-3847 (see community submission).

PubMed
Europe PMC
Cell Calcium 35:71-77 (2004)
Mapped to
1

点击页面右侧第 1 篇文献链接，打开 PubMed 数据库收录的该文献：PubMed



J Integr Plant Biol. 2008 Apr;50(4):475-83. doi: 10.1111/j.1744-7909.2008.00643.x.

PPF1 may suppress plant senescence via activating TFL1 in transgenic Arabidopsis plants

Da-Yong Wang [1], Qing Li, Ke-Ming Cui, Yu-Xian Zhu

Affiliations + expand

PMID: 18713382 DOI: 10.1111/j.1744-7909.2008.00643.x

Abstract

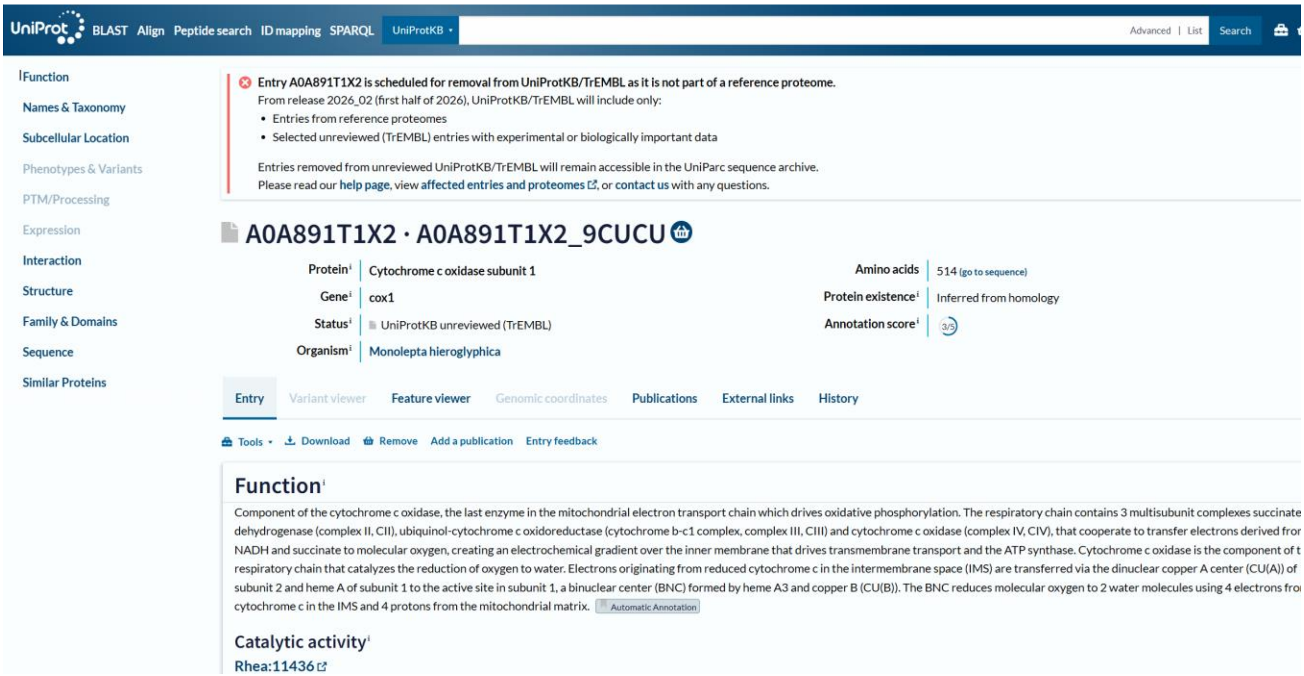
Senescence, a sequence of biochemical and physiological events, constitutes the final stage of development in higher plants and is modulated by a variety of environmental factors and internal factors. PPF1 possesses an important biological function in plant development by controlling the Ca2+ storage capacity within chloroplasts. Here we show that the expression of PPF1 might play a pivotal role in negatively regulating plant senescence as revealed by the regulation of overexpression and suppression of PPF1 on plant development. Moreover, TFL1, a key regulator in the floral repression pathway, was screened out as one of the downstream targets for PPF1 in the senescence-signaling pathway. Investigation of the senescence-related phenotypes in PPF1(-) tfl1-1 and PPF1(+) tfl1-1 double mutants confirmed and further highlighted the relation of PPF1 with TFL1 in transgenic plants. The activation of TFL1 expression by PPF1 defines an important pathway possibly essential for the negative regulation of plant senescence in transgenic Arabidopsis.

PubMed Disclaimer

Similar articles

（以 A0A891T1X2_9CUCU 为例）

打开序列条目：**A0A891T1X2_9CUCU**



Entry A0A891T1X2 is scheduled for removal from UniProtKB/TrEMBL as it is not part of a reference proteome.
From release 2026_02 (first half of 2026), UniProtKB/TrEMBL will include only:

- Entries from reference proteomes
- Selected unreviewed (TrEMBL) entries with experimental or biologically important data

Entries removed from unreviewed UniProtKB/TrEMBL will remain accessible in the UniParc sequence archive.
Please read our help page, view affected entries and proteomes, or contact us with any questions.

A0A891T1X2 · A0A891T1X2_9CUCU

Protein: Cytochrome c oxidase subunit 1
Gene: cox1
Status: UniProtKB unreviewed (TrEMBL)
Organism: Monolepta hieroglyphica
Amino acids: 514 (go to sequence)
Protein existence: Inferred from homology
Annotation score: 3/5

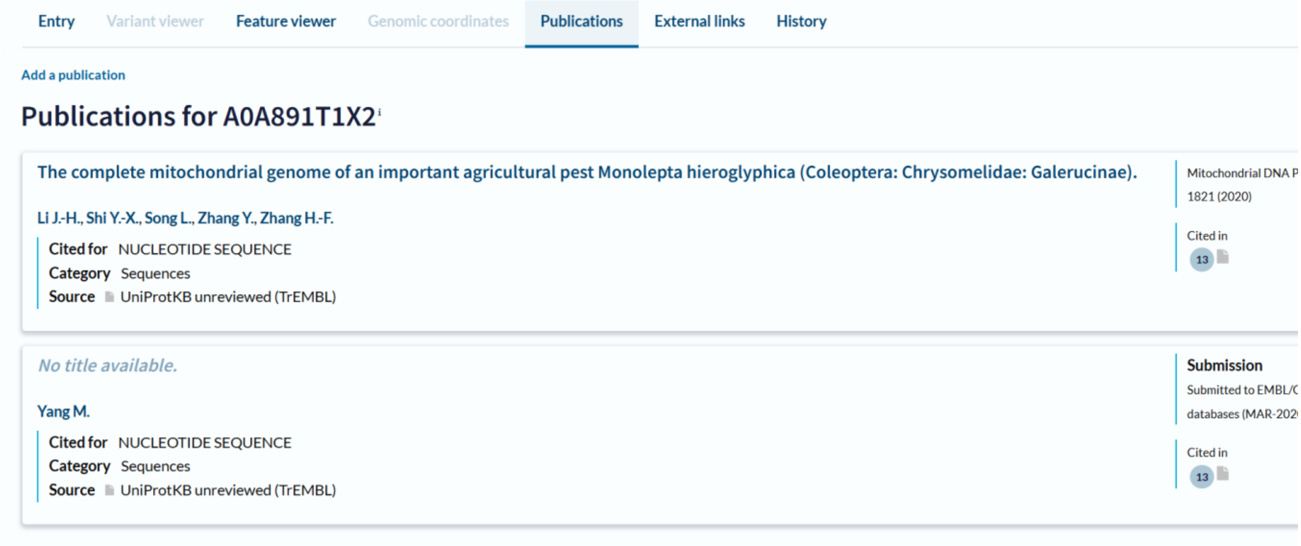
Function

Component of the cytochrome c oxidase, the last enzyme in the mitochondrial electron transport chain which drives oxidative phosphorylation. The respiratory chain contains 3 multisubunit complexes succinate dehydrogenase (complex II, CII), ubiquinol-cytochrome c oxidoreductase (cytochrome b-c1 complex, complex III, CIII) and cytochrome c oxidase (complex IV, CIV), that cooperate to transfer electrons derived fror NADH and succinate to molecular oxygen, creating an electrochemical gradient over the inner membrane that drives transmembrane transport and the ATP synthase. Cytochrome c oxidase is the component of t respiratory chain that catalyzes the reduction of oxygen to water. Electrons originating from reduced cytochrome c in the intermembrane space (IMS) are transferred via the dinuclear copper A center (CU(A)) of subunit 2 and heme A of subunit 1 to the active site in subunit 1, a binuclear center (BNC) formed by heme A3 and copper B (CU(B)). The BNC reduces molecular oxygen to 2 water molecules using 4 electrons fro cytochrome c in the IMS and 4 protons from the mitochondrial matrix.

Catalytic activity

Rhea:11436

点击 Publication 查看文献信息：Publications



Publications for A0A891T1X2

The complete mitochondrial genome of an important agricultural pest Monolepta hieroglyphica (Coleoptera: Chrysomelidae: Galerucinae).

Li J.-H., Shi Y.-X., Song L., Zhang Y., Zhang H.-F.

Cited for NUCLEOTIDE SEQUENCE
Category Sequences
Source UniProtKB unreviewed (TrEMBL)

Mitochondrial DNA P
1821 (2020)

Cited in 13

No title available.

Yang M.

Cited for NUCLEOTIDE SEQUENCE
Category Sequences
Source UniProtKB unreviewed (TrEMBL)

Submission
Submitted to EMBL/C
databases (MAR-202

Cited in 13

（2）添加文献信息（以 A0A891T1X2_9CUCU 为例）

PubMed 上查找相关文献

(cox1) AND (monolepta hieroglyphica)

Found 1 result for *(cox1) AND (monolepta hieroglyphica)*

Mitochondrial DNA B Resour. 2021 Apr 26;6(4):1541-1543. doi: 10.1080/23802359.2021.1914522.

Mitochondrial genome of *Monolepta hieroglyphica* (Coleoptera: Chrysomeloidea: Chrysomelidae) and phylogenetic analysis

Wei Li [1 2 3], Shuo Shen [1 2], Hongyu Chen [1 2 3]

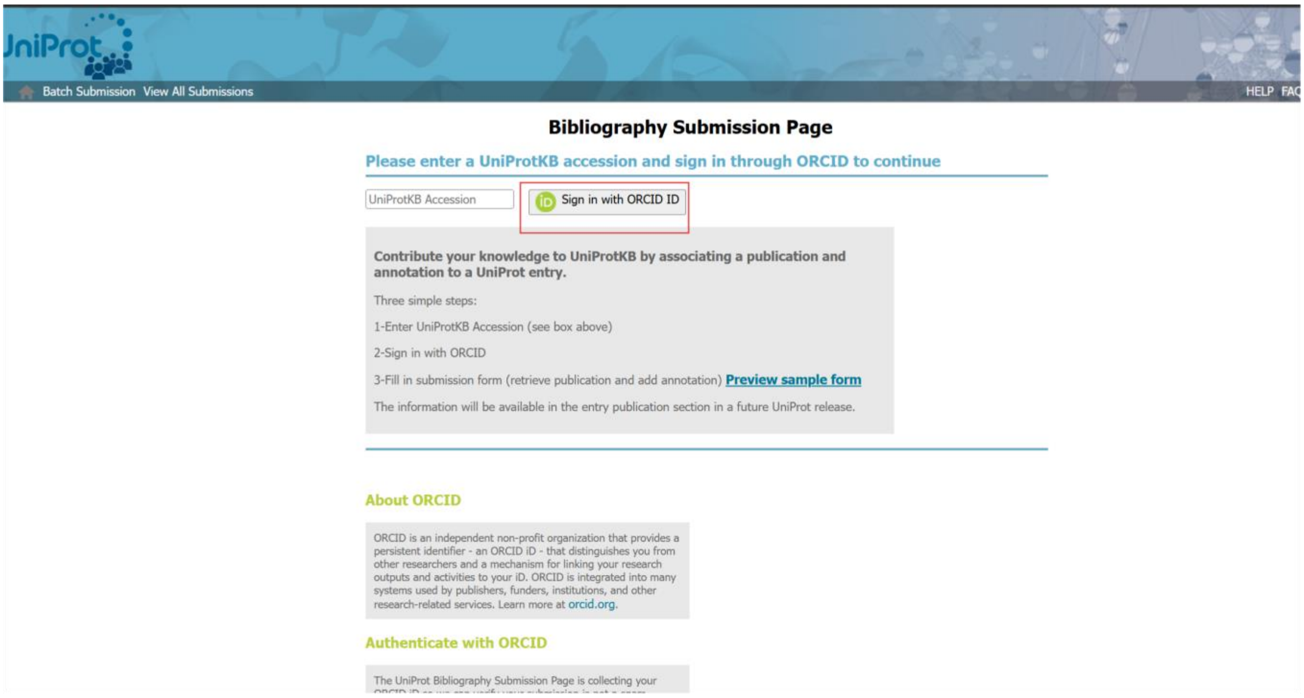
Affiliations + expand

PMID: 33969213 PMCID: PMC8079056 DOI: 10.1080/23802359.2021.1914522

Abstract

Monolepta hieroglyphica (Motschulsky, 1858) is a major pest of potato, maize, cotton and sorghum in China. In this study, we sequenced and analyzed the complete mitochondrial genome (mitogenome) of *M. hieroglyphica*. This mitogenome was 15,761 bp long and encoded 13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNAs) and two ribosomal RNA unit genes (rRNAs). Gene order was conserved and identical to most other previously sequenced Galerucinae Most PCGs of *M. hieroglyphica* have the conventional start codons ATN (six ATT, five ATG and one ATC), with the exception of *nad1* (TTG). Except for three genes (*cox1*, *nad4* and *nad5*) end with the incomplete stop

回到 Uniprot 界面，点击 add a publication



Bibliography Submission Page

Please enter a UniProtKB accession and sign in through ORCID to continue

UniProtKB Accession

Contribute your knowledge to UniProtKB by associating a publication and annotation to a UniProt entry.

Three simple steps:

1-Enter UniProtKB Accession (see box above)

2-Sign in with ORCID

3-Fill in submission form (retrieve publication and add annotation) Preview sample form

The information will be available in the entry publication section in a future UniProt release.

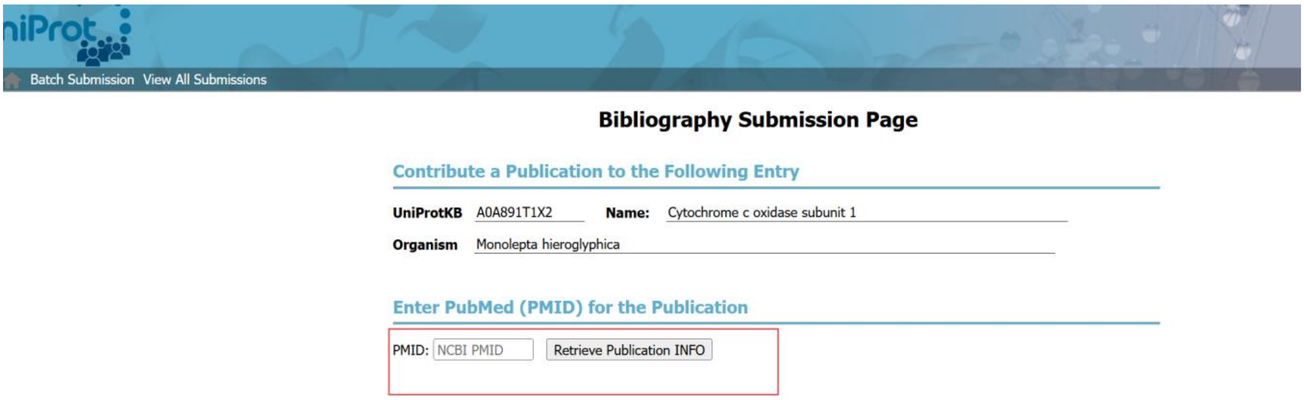
About ORCID

ORCID is an independent non-profit organization that provides a persistent identifier - an ORCID iD - that distinguishes you from other researchers and a mechanism for linking your research outputs and activities to your iD. ORCID is integrated into many systems used by publishers, funders, institutions, and other research-related services. Learn more at orcid.org.

Authenticate with ORCID

The UniProt Bibliography Submission Page is collecting your

登录 ORCID ID



Bibliography Submission Page

Contribute a Publication to the Following Entry

UniProtKB A0A891T1X2 **Name:** Cytochrome c oxidase subunit 1

Organism Monolepta hieroglyphica

Enter PubMed (PMID) for the Publication

PMID: NCBI PMID

输入要添加文献的 PMID，后点击 Retrieve publication INFO(检索出版物信息)

Bibliography Submission Page

Contribute a Publication to the Following Entry

UniProtKB A0A891T1X2 **Name:** Cytochrome c oxidase subunit 1

Organism Monolepta hieroglyphica

Enter PubMed (PMID) for the Publication

PMID: 33969213

Publication information for PMID:33969213

"Mitochondrial genome of Monolepta hieroglyphica (Coleoptera: Chrysomeloidea: Chrysomelidae) and phylogenetic analysis."

Li W, Shen S, Chen H.

Mitochondrial DNA B Resour 6:1541-1543 (2021 Apr 26) [PubMed] [Europe PMC]

Is this the publication you want to submit?

● Yes, this is the publication. ○ No, it's not.

Is this publication about the protein in the organism indicated above?

● Yes, it is. ○ No, it is not.

Annotation Section (optional)

Contribute your knowledge.

Does this publication describe any of these properties about the gene/protein?

Please select all that apply.

☐ Function (includes molecular function, processes/pathway, sequence features important for function)

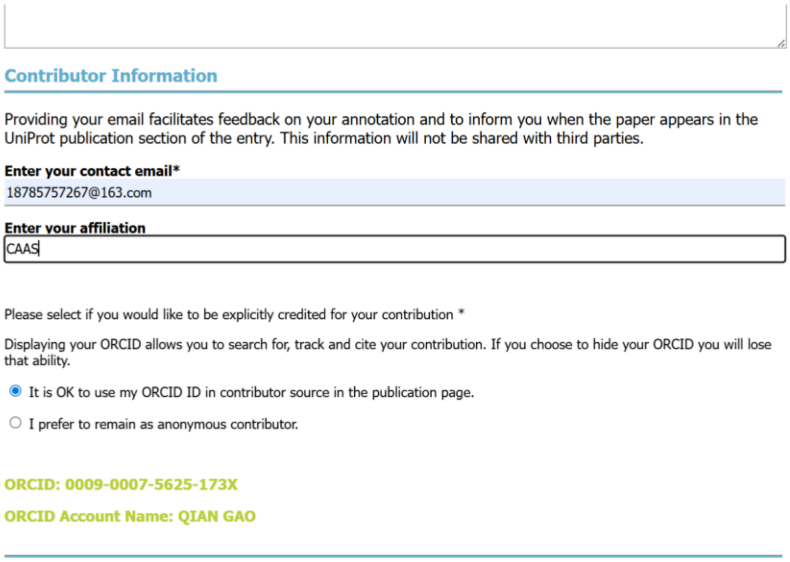
☐ Subcellular location (information on the location and/or the topology of the mature protein in the cell)

☐ Diseases/Phenotypes & Variants (includes disease, phenotype, variants, mutagenesis, biotechnology use)

☐ PTM/Processing (post-translational information, protein cleavage)

☐ Expression (information on natural expression of the RNA/protein in this organism: e.g. tissue, developmental stage)

填完相关信息后点击最后的 Submit Bibliography



Contributor Information

Providing your email facilitates feedback on your annotation and to inform you when the paper appears in the UniProt publication section of the entry. This information will not be shared with third parties.

Enter your contact email*
18785757267@163.com

Enter your affiliation
CAAS

Please select if you would like to be explicitly credited for your contribution *

Displaying your ORCID allows you to search for, track and cite your contribution. If you choose to hide your ORCID you will lose that ability.

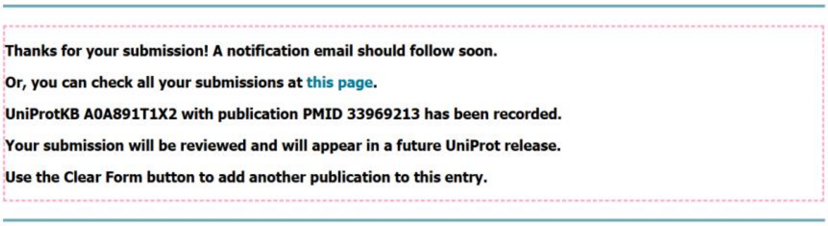
● It is OK to use my ORCID ID in contributor source in the publication page.

○ I prefer to remain as anonymous contributor.

ORCID: 0009-0007-5625-173X

ORCID Account Name: QIAN GAO

收到邮箱即提交成功



Thanks for your submission! A notification email should follow soon.

Or, you can check all your submissions at this page.

UniProtKB A0A891T1X2 with publication PMID 33969213 has been recorded.

Your submission will be reviewed and will appear in a future UniProt release.

Use the Clear Form button to add another publication to this entry.

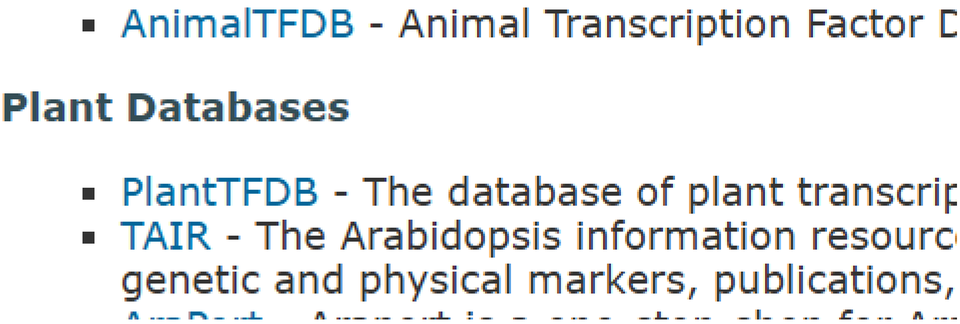
网易邮箱大师

UniProt Bibliography Submission Notice
Bibliography Submission Confirmation 标为已读

D PlantTFDB 序列注释信息

阅读植物转录因子数据库 PlantTFDB 中 SBP 家族基本介绍 SBP Family Introduction 了解该转录因子家族 SQUAMOSA promoter Binding Protein (SBP)名称来源、DNA 结合结构域序列长度和特异保守位点等基本信息

abc 网站中打开 PlantTFDB；

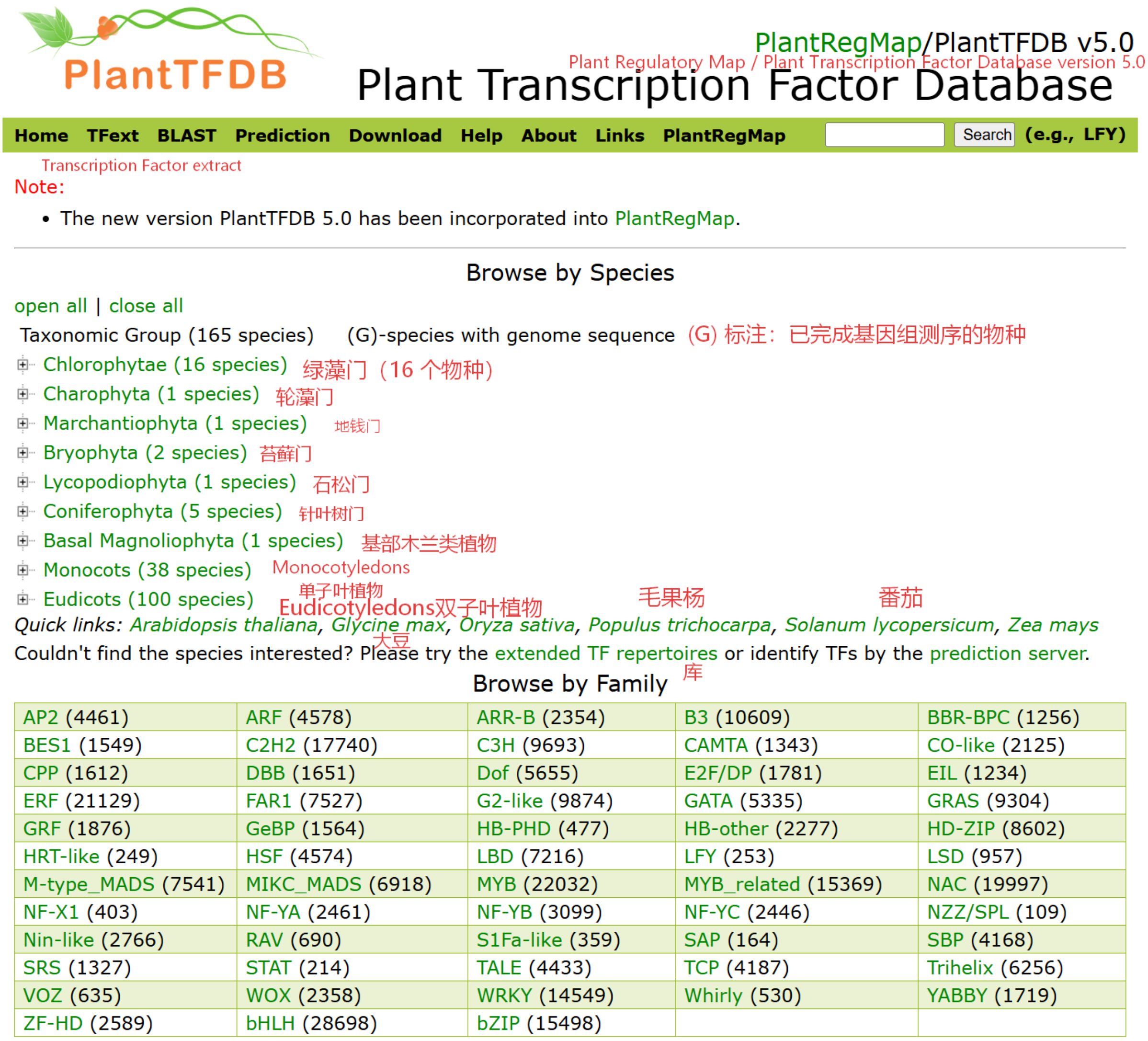


- AnimalTFDB - Animal Transcription Factor D

Plant Databases

- PlantTFDB - The database of plant transcrip
- TAIR - The Arabidopsis information resourc
genetic and physical markers, publications,

界面如下：



PlantRegMap/PlantTFDB v5.0

Plant Regulatory Map / Plant Transcription Factor Database version 5.0

Plant Transcription Factor Database

Transcription Factor extract

Note:

- The new version PlantTFDB 5.0 has been incorporated into PlantRegMap.

Browse by Species

open all | close all

Taxonomic Group (165 species) (G)-species with genome sequence (G) 标注：已完成基因组测序的物种

- ⊞ Chlorophytae (16 species) 绿藻门（16 个物种）
- ⊞ Charophyta (1 species) 轮藻门
- ⊞ Marchantiophyta (1 species) 地钱门
- ⊞ Bryophyta (2 species) 苔藓门
- ⊞ Lycopodiophyta (1 species) 石松门
- ⊞ Coniferophyta (5 species) 针叶树门
- ⊞ Basal Magnoliophyta (1 species) 基部木兰类植物
- ⊞ Monocots (38 species) Monocotyledons 单子叶植物
- ⊞ Eudicots (100 species) Eudicotyledons双子叶植物

毛果杨 番茄

Quick links: Arabidopsis thaliana, *Glycine max*, *Oryza sativa*, *Populus trichocarpa*, *Solanum lycopersicum*, *Zea mays*

大豆

Couldn't find the species interested? Please try the extended TF repertoires or identify TFs by the prediction server.

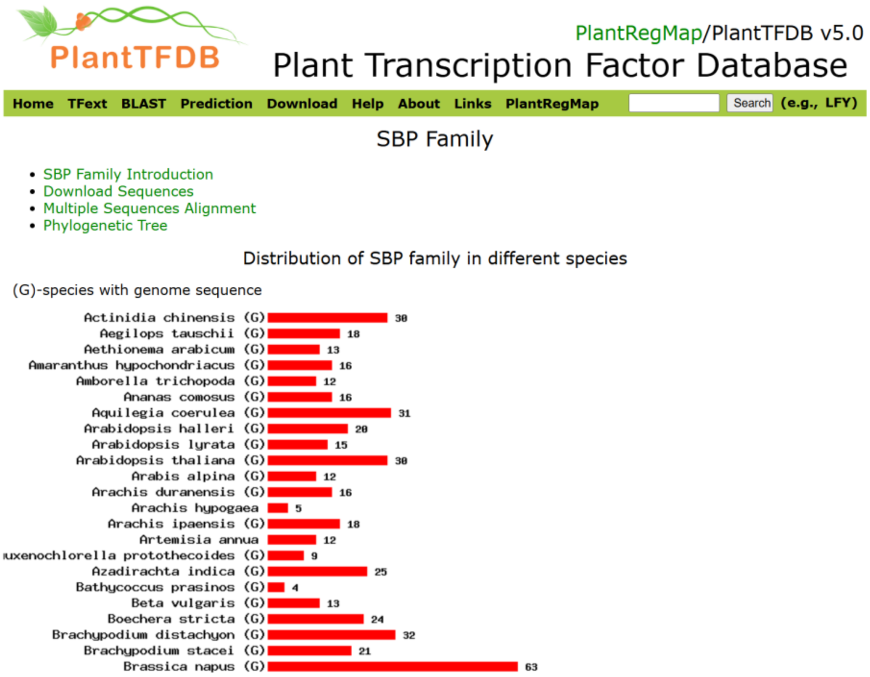
库

Browse by Family

AP2 (4461)	ARF (4578)	ARR-B (2354)	B3 (10609)	BBR-BPC (1256)
BES1 (1549)	C2H2 (17740)	C3H (9693)	CAMTA (1343)	CO-like (2125)
CPP (1612)	DBB (1651)	Dof (5655)	E2F/DP (1781)	EIL (1234)
ERF (21129)	FAR1 (7527)	G2-like (9874)	GATA (5335)	GRAS (9304)
GRF (1876)	GeBP (1564)	HB-PHD (477)	HB-other (2277)	HD-ZIP (8602)
HRT-like (249)	HSF (4574)	LBD (7216)	LFY (253)	LSD (957)
M-type_MADS (7541)	MIKC_MADS (6918)	MYB (22032)	MYB_related (15369)	NAC (19997)
NF-X1 (403)	NF-YA (2461)	NF-YB (3099)	NF-YC (2446)	NZZ/SPL (109)
Nin-like (2766)	RAV (690)	S1Fa-like (359)	SAP (164)	SBP (4168)
SRS (1327)	STAT (214)	TALE (4433)	TCP (4187)	Trihelix (6256)
VOZ (635)	WOX (2358)	WRKY (14549)	Whirly (530)	YABBY (1719)
ZF-HD (2589)	bHLH (28698)	bZIP (15498)		

PlantTFDB（Plant Transcription Factor Database，植物转录因子数据库）v5.0 是全球权威的植物转录因子综合数据库，已整合至 PlantRegMap 平台。该数据库收录了涵盖绿藻、苔藓、蕨类、裸子植物到被子植物共 165 个植物物种的转录因子信息，其中带(G)标记的物种为已完成基因组测序的物种，支持按物种分类浏览与按转录因子家族分类检索。数据库系统整理了 AP2、bHLH、MYB、NAC 等超 60 个转录因子家族的完整注释信息，同时提供 TFext（转录因子提取）、BLAST 序列比对、转录因子预测、数据下载等实用分析工具，为植物基因表达调控、分子育种、功能基因组学研究提供了全面、权威的转录因子资源与分析平台。

SBP Family



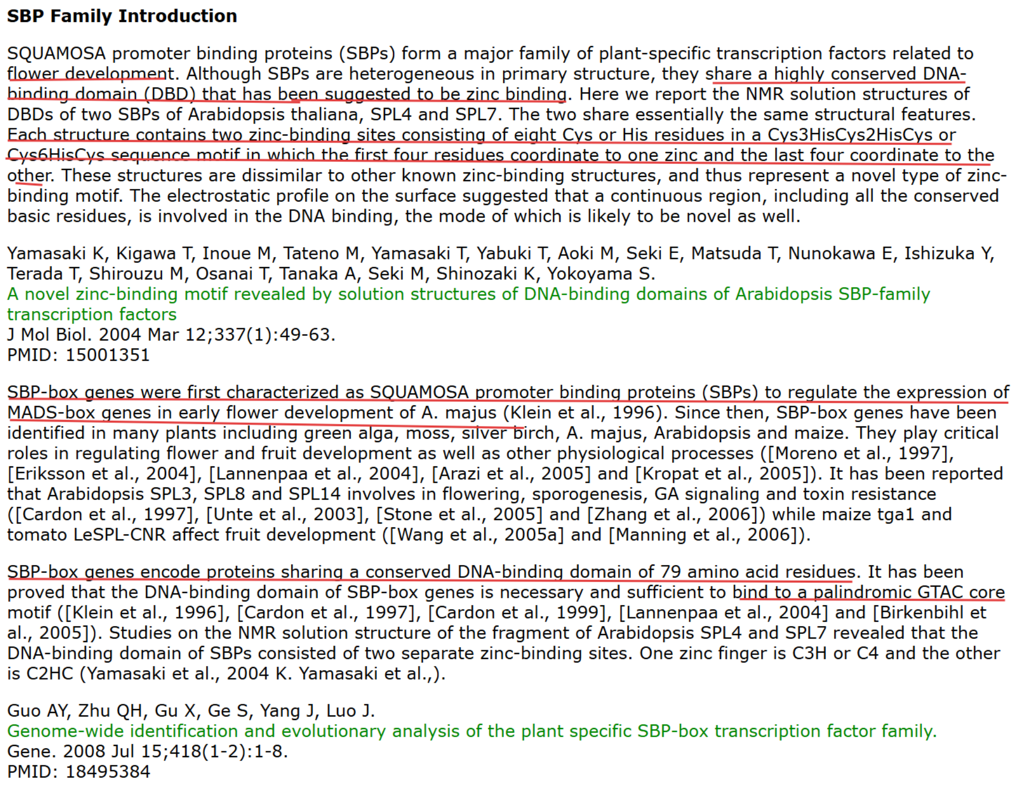
SBP Family

- SBP Family Introduction
- Download Sequences
- Multiple Sequences Alignment
- Phylogenetic Tree

Distribution of SBP family in different species

(G)-species with genome sequence

SQUAMOSA promoter Binding Protein（SBP，SBP-box）家族名称来源于其最初被鉴定为可结合金鱼草（*Antirrhinum majus*）SQUAMOSA 基因启动子的转录因子，该家族是植物特有的转录因子家族，核心功能与花发育调控相关。其 DNA 结合结构域（DBD）为高度保守的 79 个氨基酸残基的保守区域，该结构域包含 2 个独立的锌结合位点，对应 Cys3HisCys2HisCys 或 Cys6HisCys 的序列基序（其中前 4 个残基结合 1 个锌离子，后 4 个残基结合另 1 个锌离子），可特异性识别回文 GTAC 核心基序；同时该结构域表面的连续碱性保守残基区域参与 DNA 结合，是该家族发挥转录调控功能的关键结构基础。



SBP Family Introduction

SQUAMOSA promoter binding proteins (SBPs) form a major family of plant-specific transcription factors related to flower development. Although SBPs are heterogeneous in primary structure, they share a highly conserved DNA-binding domain (DBD) that has been suggested to be zinc binding. Here we report the NMR solution structures of DBDs of two SBPs of Arabidopsis thaliana, SPL4 and SPL7. The two share essentially the same structural features. Each structure contains two zinc-binding sites consisting of eight Cys or His residues in a Cys3HisCys2HisCys or Cys6HisCys sequence motif in which the first four residues coordinate to one zinc and the last four coordinate to the other. These structures are dissimilar to other known zinc-binding structures, and thus represent a novel type of zinc-binding motif. The electrostatic profile on the surface suggested that a continuous region, including all the conserved basic residues, is involved in the DNA binding, the mode of which is likely to be novel as well.

Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Nunokawa E, Ishizuka Y, Terada T, Shirouzu M, Osanai T, Tanaka A, Seki M, Shinozaki K, Yokoyama S.
A novel zinc-binding motif revealed by solution structures of DNA-binding domains of Arabidopsis SBP-family transcription factors
J Mol Biol. 2004 Mar 12;337(1):49-63.
PMID: 15001351

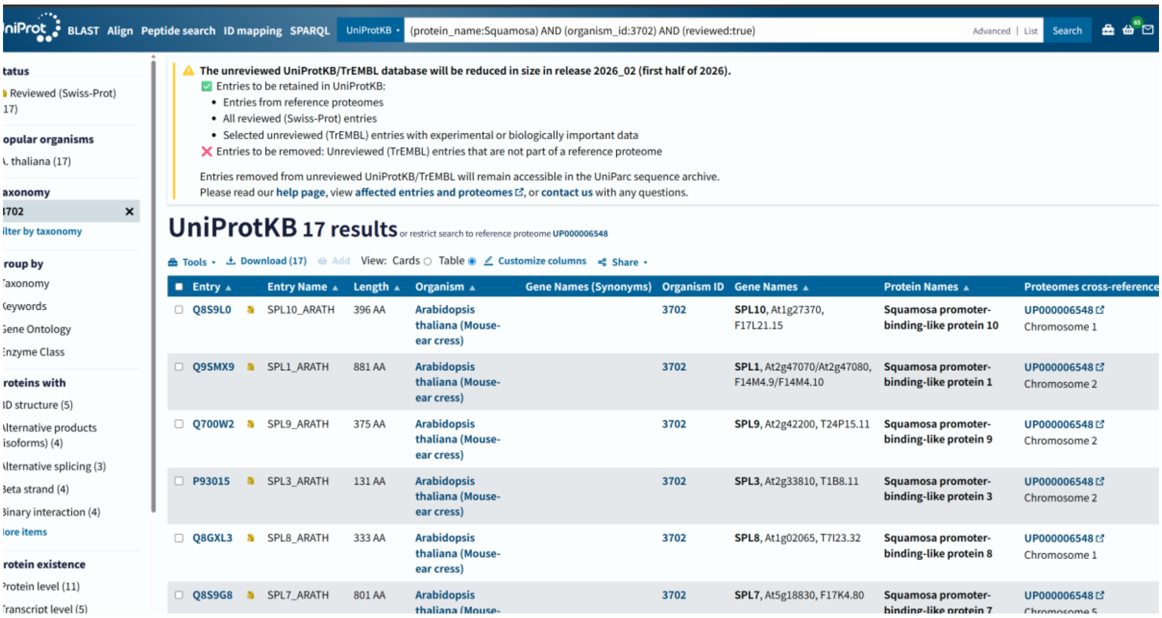
SBP-box genes were first characterized as SQUAMOSA promoter binding proteins (SBPs) to regulate the expression of MADS-box genes in early flower development of A. majus (Klein et al., 1996). Since then, SBP-box genes have been identified in many plants including green alga, moss, silver birch, A. majus, Arabidopsis and maize. They play critical roles in regulating flower and fruit development as well as other physiological processes ([Moreno et al., 1997], [Eriksson et al., 2004], [Lannenpaa et al., 2004], [Arazi et al., 2005] and [Kropat et al., 2005]). It has been reported that Arabidopsis SPL3, SPL8 and SPL14 involves in flowering, sporogenesis, GA signaling and toxin resistance ([Cardon et al., 1997], [Unte et al., 2003], [Stone et al., 2005] and [Zhang et al., 2006]) while maize tga1 and tomato LeSPL-CNR affect fruit development ([Wang et al., 2005a] and [Manning et al., 2006]).

SBP-box genes encode proteins sharing a conserved DNA-binding domain of 79 amino acid residues. It has been proved that the DNA-binding domain of SBP-box genes is necessary and sufficient to bind to a palindromic GTAC core motif ([Klein et al., 1996], [Cardon et al., 1997], [Cardon et al., 1999], [Lannenpaa et al., 2004] and [Birkenbihl et al., 2005]). Studies on the NMR solution structure of the fragment of Arabidopsis SPL4 and SPL7 revealed that the DNA-binding domain of SBPs consisted of two separate zinc-binding sites. One zinc finger is C3H or C4 and the other is C2HC (Yamasaki et al., 2004 K. Yamasaki et al.,).

Guo AY, Zhu QH, Gu X, Ge S, Yang J, Luo J.
Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family.
Gene. 2008 Jul 15;418(1-2):1-8.
PMID: 18495384

利用 UniProtKB 高级检索功能，在下拉菜单蛋白质名称 Protein name 栏目中输入 Squamosa，在下拉菜单物质名称栏目 Orgamism 中输入拟南芥拉丁名 Arabidopsis thaliana，在下拉菜单审阅状况 Reviewed 中选择 Yes，点击搜索按钮，检索框显示以下信息：

(protein_name:SQUAMOSA) AND (organism_id:3702) AND (reviewed:true)



(protein_name:Squamosa) AND (organism_id:3702) AND (reviewed:true)

The unreviewed UniProtKB/TrEMBL database will be reduced in size in release 2026_02 (first half of 2026).

☑ Entries to be retained in UniProtKB:
- Entries from reference proteomes
- All reviewed (Swiss-Prot) entries
- Selected unreviewed (TrEMBL) entries with experimental or biologically important data

☒ Entries to be removed: Unreviewed (TrEMBL) entries that are not part of a reference proteome

Entries removed from unreviewed UniProtKB/TrEMBL will remain accessible in the UniParc sequence archive.
Please read our **help page**, view **affected entries and proteomes**, or **contact us** with any questions.

UniProtKB 17 results or restrict search to reference proteome **UP000006548**

☐	Entry	Entry Name	Length	Organism	Gene Names (Synonyms)	Organism ID	Gene Names	Protein Names	Proteomes cross-reference
☐	Q8S9L0	SPL10_ARATH	396 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL10**, At1g27370, F17L21.15	**Squamosa promoter-binding-like protein 10**	UP000006548 Chromosome 1
☐	Q9SMX9	SPL1_ARATH	881 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL1**, At2g47070/At2g47080, F14M4.9/F14M4.10	**Squamosa promoter-binding-like protein 1**	UP000006548 Chromosome 2
☐	Q700W2	SPL9_ARATH	375 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL9**, At2g42200, T24P15.11	**Squamosa promoter-binding-like protein 9**	UP000006548 Chromosome 2
☐	P93015	SPL3_ARATH	131 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL3**, At2g33810, T1B8.11	**Squamosa promoter-binding-like protein 3**	UP000006548 Chromosome 2
☐	Q8GXL3	SPL8_ARATH	333 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL8**, At1g02065, T7I23.32	**Squamosa promoter-binding-like protein 8**	UP000006548 Chromosome 1
☐	Q8S9G8	SPL7_ARATH	801 AA	Arabidopsis thaliana (Mouse-		3702	**SPL7**, At5g18830, F17K4.80	**Squamosa promoter-binding-like protein 7**	UP000006548 Chromosome 5

检索结果显示人工审阅的 17 个拟南芥 SBP 家族成员 Squamosa

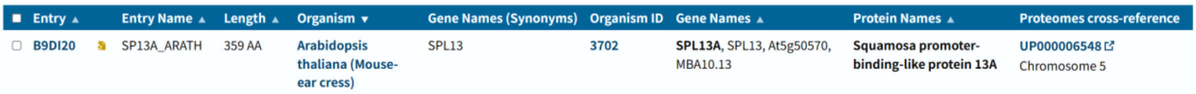
promoter-binding-like protein (SPL)

按序列长度排序，找出最短和最长的序列

最短：

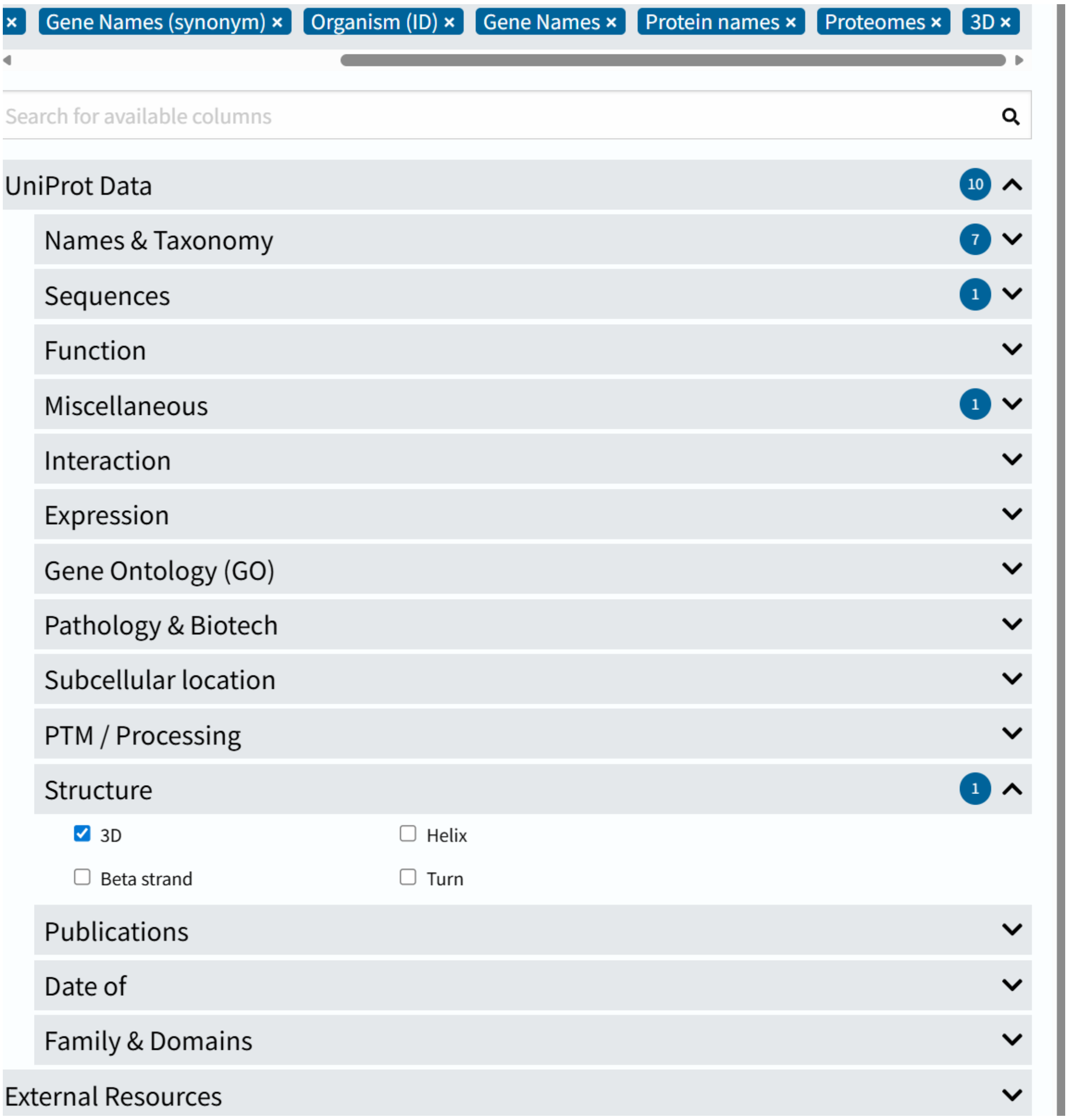
■ Entry ▲		Entry Name ▲	Length ▲	Organism ▲	Gene Names (Synonyms)	Organism ID	Gene Names ▲	Protein Names ▲	Proteomes cross-reference
☐ P93015		SPL3_ARATH	131 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL3**, At2g33810, T1B8.11	**Squamosa promoter-binding-like protein 3**	UP000006548 Chromosome 2

最长；



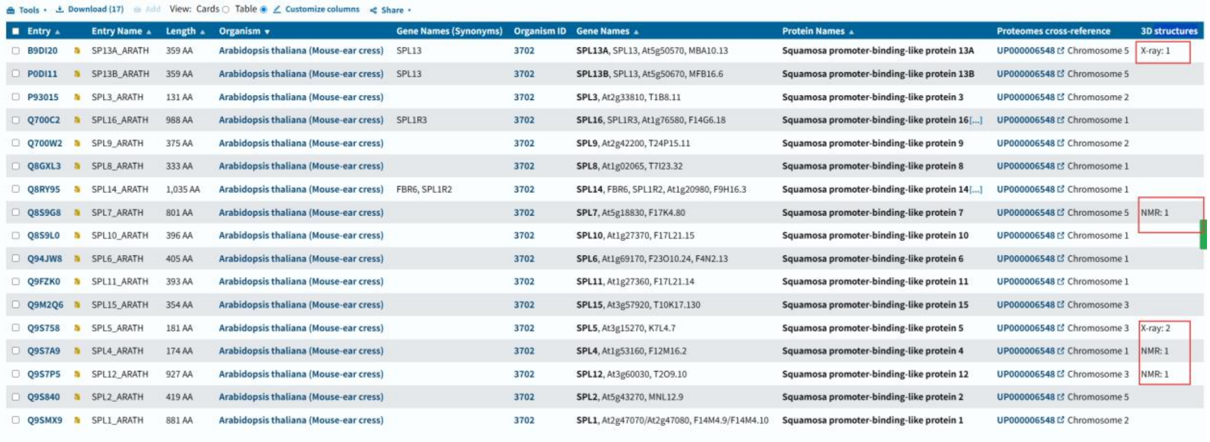
■ Entry ▲		Entry Name ▲	Length ▲	Organism ▼	Gene Names (Synonyms)	Organism ID	Gene Names ▲	Protein Names ▲	Proteomes cross-reference
☐ B9DI20		SP13A_ARATH	359 AA	Arabidopsis thaliana (Mouse-ear cress)	SPL13	3702	**SPL13A**, SPL13, At5g50570, MBA10.13	**Squamosa promoter-binding-like protein 13A**	UP000006548 Chromosome 5

点击定制显示字段 Customize column，在弹出窗口中点击结构 Structue,选择三维结构 3D.点击保存按钮 Save,查看哪些条目结构已经测定



☑ 3D
☐ Helix
☐ Beta strand
☐ Turn

3D Structure 已测定



Tools • Download (17) Add View: Cards ○ Table ● Customize columns Share •

■ Entry ▲	Entry Name ▲	Length ▲	Organism ▼	Gene Names (Synonyms)	Organism ID	Gene Names ▲	Protein Names ▲	Proteomes cross-reference	3D structures
☐ B9DI20	SP13A_ARATH	359 AA	Arabidopsis thaliana (Mouse-ear cress)	SPL13	3702	**SPL13A**, SPL13, At5g50570, MBA10.13	**Squamosa promoter-binding-like protein 13A**	UP000006548 Chromosome 5	X-ray: 1
☐ P0DI11	SP13B_ARATH	359 AA	Arabidopsis thaliana (Mouse-ear cress)	SPL13	3702	**SPL13B**, SPL13, At5g50670, MFB16.6	**Squamosa promoter-binding-like protein 13B**	UP000006548 Chromosome 5	
☐ P93015	SPL3_ARATH	131 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL3**, At2g33810, T1B8.11	**Squamosa promoter-binding-like protein 3**	UP000006548 Chromosome 2	
☐ Q700C2	SPL16_ARATH	988 AA	Arabidopsis thaliana (Mouse-ear cress)	SPL1R3	3702	**SPL16**, SPL1R3, At1g76580, F14G6.18	**Squamosa promoter-binding-like protein 16**[...]	UP000006548 Chromosome 1	
☐ Q700W2	SPL9_ARATH	375 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL9**, At2g42200, T24P15.11	**Squamosa promoter-binding-like protein 9**	UP000006548 Chromosome 2	
☐ Q8GXL3	SPL8_ARATH	333 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL8**, At1g02065, T7I23.32	**Squamosa promoter-binding-like protein 8**	UP000006548 Chromosome 1	
☐ Q8RY95	SPL14_ARATH	1,035 AA	Arabidopsis thaliana (Mouse-ear cress)	FBR6, SPL1R2	3702	**SPL14**, FBR6, SPL1R2, At1g20980, F9H16.3	**Squamosa promoter-binding-like protein 14**[...]	UP000006548 Chromosome 1	
☐ Q8S9G8	SPL7_ARATH	801 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL7**, At5g18830, F17K4.80	**Squamosa promoter-binding-like protein 7**	UP000006548 Chromosome 5	NMR: 1
☐ Q8S9L0	SPL10_ARATH	396 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL10**, At1g27370, F17L21.15	**Squamosa promoter-binding-like protein 10**	UP000006548 Chromosome 1	
☐ Q94JW8	SPL6_ARATH	405 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL6**, At1g69170, F23O10.24, F4N2.13	**Squamosa promoter-binding-like protein 6**	UP000006548 Chromosome 1	
☐ Q9FZK0	SPL11_ARATH	393 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL11**, At1g27360, F17L21.14	**Squamosa promoter-binding-like protein 11**	UP000006548 Chromosome 1	
☐ Q9M2Q6	SPL15_ARATH	354 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL15**, At3g57920, T10K17.130	**Squamosa promoter-binding-like protein 15**	UP000006548 Chromosome 3	
☐ Q9S758	SPL5_ARATH	181 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL5**, At3g15270, K7L4.7	**Squamosa promoter-binding-like protein 5**	UP000006548 Chromosome 3	X-ray: 2
☐ Q9S7A9	SPL4_ARATH	174 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL4**, At1g53160, F12M16.2	**Squamosa promoter-binding-like protein 4**	UP000006548 Chromosome 1	NMR: 1
☐ Q9S7P5	SPL12_ARATH	927 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL12**, At3g60030, T2O9.10	**Squamosa promoter-binding-like protein 12**	UP000006548 Chromosome 3	NMR: 1
☐ Q9S840	SPL2_ARATH	419 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL2**, At5g43270, MNL12.9	**Squamosa promoter-binding-like protein 2**	UP000006548 Chromosome 5	
☐ Q9SMX9	SPL1_ARATH	881 AA	Arabidopsis thaliana (Mouse-ear cress)		3702	**SPL1**, At2g47070/At2g47080, F14M4.9/F14M4.10	**Squamosa promoter-binding-like protein 1**	UP000006548 Chromosome 2	

D 个人总结

G2A 边汉青

在学过 Uniprot 数据库，发现 Uniprot 数据库可以查找自己领域内相关的蛋白质，同时也可以通过相关蛋白质的信息页面中的 publication 找到其引用的文章，更加详细的了解和确认是否是自己所要找的蛋白质，同时有许多外连接，可以非常方便的跳转至基因或蛋白（IEMBL、GenBank、DDBJ）库，查找其相关的家族数据库，可以通过 CDD 查找出蛋白质保守结构域数据库中的信息、通过 Pfam 查找到其所在的家族和其他家族蛋白，也可以通过 AlphaFoldDB 查看蛋白质结构，另一方面，也可以利用 uniprot 进行蛋白质多序列比对，方便而快捷。

总的来说，Unirpot 不仅作为一个蛋白质数据库，同时他也是蛋白质相关数据链接的综合性网站，对于研究某个基因的蛋白质功能、查询其相关结构和相似蛋白或相似基因蛋白质是非常方便的。

只是对于 Unirpot 的运用还需要多加练习，部分网站初次接触，仍需时间慢慢熟悉，另一方面，可以借助 ai 快速熟悉各个网站的功能，在实际课题应用时，利用 ai 指导会加快学习和对网站的应用，找出自己需要的数据信息。

G2B 刘奇

以一条核酸序列为起点，系统演练了从序列到功能的数据库交叉检索流程，对果蝇 SXL 蛋白形成了完整认知。

SXL 蛋白信息：SXL 是果蝇性别决定的主开关（UniProt P19339），全长 354 个氨基酸。结构上含两个 RRM 结构域（CDD 家族 cd12375），偏好结合富含 U 的 RNA 序列，通过调控 tra 和自身 Sxl 的选择性剪接控制性别分化；另有 12.4% 固有无序区域（DisProt 共识），分布于 N 端及结构域连接区，可在结合 RNA 时发生无序→有序转变，有 NMR 和 X 射线证据支持（PDB：1B7F、2SXL、3SXL）。功能上，SXL 在雌性中抑制 msl-2 实现 X 染色体剂量补偿。表达丰度约 6-15 ppm（PaxDb），属低丰度调控蛋白。互作网络包括 snf、tra2、vir 等剪接因子（STRING）。

一个 UniProt 登录号可串联各数据库。不同数据库解决不同问题：CDD 找结构域，DisProt 看无序区域，TCDB 排除跨膜特征，STRING 查互作网络，PaxDb 看表达丰度。单一数据库只能提供片面信息，交叉验证才能构建完整认知。此外，PDB 结构号与文献 PMID 是验证结论可靠性的重要依据，序列比对中的保守残基往往揭示功能关键位点。这些方法可迁移至任何目标蛋白的分析中。

G2C 高倩

过去三周，我简要学习了关于蛋白、核酸、基因组、反应和通路、家族等各类数据库，重点掌握了 UniProt 蛋白质数据库的基本框架、检索、序列下载、外部链接等，并结合自己课题相关蛋白质练习 UniProt 蛋白质数据库的使用。

一是 UniProt 蛋白质数据库的学习和练习。UniProt 蛋白质数据库由 UniProtKB 知识库、UniParc 归档库、UniRef 参考数据集三大核心模块组成，了解了 UniProtKB 中 Swiss-Prot（人工审编、高质量）与 TrEMBL（自动注释、大规模）的分工差异；掌握了数据库统计报表的核心信息维度，包括常规注释（General Annotation）、序列特征注释（Sequence Annotation）的核心内容，进一步熟悉了数据库与 NCBI、PDB、KEGG 等多类数据库的交叉链接。围绕 UniProt 完成了四项练习。第一项以人 α-血红蛋白（HBA_HUMAN）为对象，系统练习了蛋白功能、名称分类、亚细胞定位、相关疾病、翻译后修饰（PTM）、表达、相互作用、三维结构、蛋白家族、序列、相似蛋白等全维度注释信息的检索；第

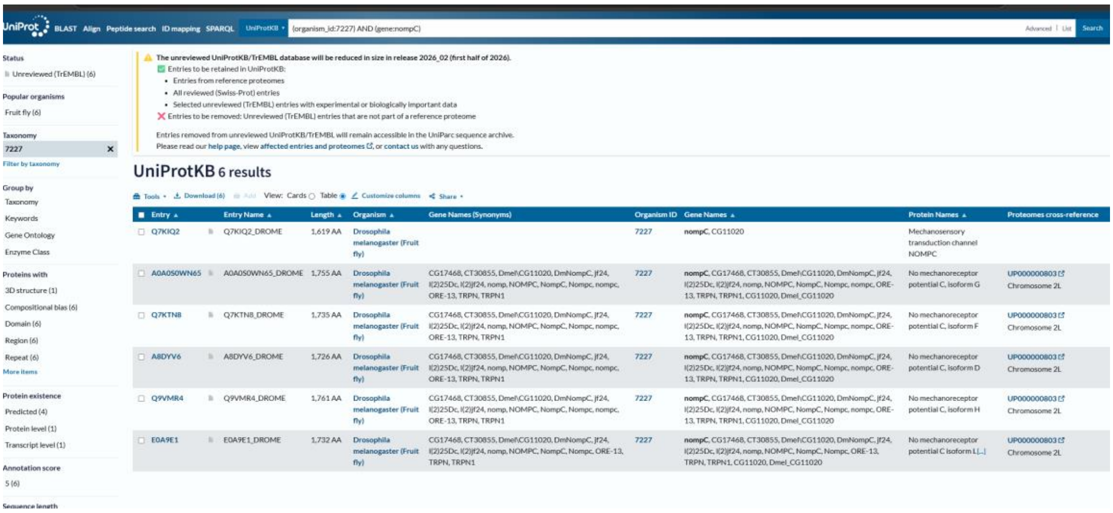
二项以豌豆内膜蛋白（PPF1_PEA）、拟南芥 Alb3 蛋白（ALB3_ARATH）为对象，完成了蛋白功能、基因信息、亚细胞定位、三维结构、家族同源蛋白的检索，学习了蛋白在 CDD、Pfam、InterPro、KEGG、PDB 等外部数据库的交叉链接，练习了文献信息提取、用户审编状态查看、PubMed 文献溯源及在 UniProt 中点击添加文献；第三项针对拟南芥 SBP 转录因子家族，运用 UniProtKB 高级检索功能，通过设定蛋白名、物种、审编状态等条件，检索得到 17 个人工审编的 SBP 家族成员，完成了序列长度排序、最短/最长序列筛选、三维结构条目筛选等批量数据处理操作，掌握了 Customize column 等个性化检索技能。同时，结合自己课题相关物种双斑长跗萤叶甲 *Monolepta hieroglyphica* 的相关蛋白细胞色素 c 氧化酶亚基ⅠCOX1，进行 Uniprot 高级检索，External Links 加深对该蛋白的学习。

二是各数据库的简要学习。我解了综合性数据库，包括 Database Commons、NAR Database Issue，JBDC 等生物信息学核心数据库，三个综合数据库整合了几乎所有的数据库；整合蛋白序列信息的 UniProt、参考序列库 RefSeq、国际核酸序列库 GenBank、日本 DNA 数据库 DDBJ，以及蛋白质结构相关的 PDBj、RCSB（Research Collaboratory for Structural Bioinformatics）、PDBe，NCBI 的保守结构域库 CDD；同时学习了家族数据库，包括蛋白家族库 Pfam、RNA 家族库 RFam、重复序列库 DFam 和进化树数据库 TreeFam。此外还有互作分析与通路数据库，如通路数据库 REACTOME 及其植物版本 Plant REACTOME；基因组相关资源，包括 KEGG、欧洲真核基因组库 ENSEMBL、植物基因组分析平台 Gramene，以及 NCBI 的基因组浏览器与基因组资源入口；系统学习了 PlantTFDB 植物转录因子数据库，了解其覆盖 165 个植物物种；阅读了 SBP（SQUAMOSA promoter Binding Protein）转录因子家族的名称来源、DNA 结合结构域特征、保守位点等基础信息，明确了植物转录因子在基因表达调控中的核心作用；以上数据库全面呈现了从序列、结构、功能到通路、进化的多组学数据库体系。

相对第一次总结，对本课程课余学习时间明显增加，对课程学习框架有了进一步深刻的理解和掌握，继续保持！

5. 问题

Q1 当在 Uniprot 高级搜索一个物种的一个基因时，通常会出现蛋白质长短不一的条目，如何确定自己想要的是那一条呢



(organism_id:7227) AND (gene:nompC)

The unreviewed UniProtKB/TrEMBL database will be reduced in size in release 2026_02 (first half of 2026).

☑ Entries to be retained in UniProtKB:

- Entries from reference proteomes
- All reviewed (Swiss-Prot) entries
- Selected unreviewed (TrEMBL) entries with experimental or biologically important data

☒ Entries to be removed: Unreviewed (TrEMBL) entries that are not part of a reference proteome

Entries removed from unreviewed UniProtKB/TrEMBL will remain accessible in the UniParc sequence archive.
Please read our help page, view affected entries and proteomes, or contact us with any questions.

UniProtKB 6 results

Entry	Entry Name	Length	Organism	Gene Names (Synonyms)	Organism ID	Gene Names	Protein Names	Proteomes cross-reference
Q7KIQ2	Q7KIQ2_DROME	1,619 AA	Drosophila melanogaster (Fruit fly)		7227	**nompC**, CG11020	Mechanosensory transduction channel NOMPC	
A0A0S0WN65	A0A0S0WN65_DROME	1,755 AA	Drosophila melanogaster (Fruit fly)	CG17468, CT30855, Dmel\CG11020, DmNompC, jf24, l(2)25Dc, l(2)jf24, nomp, NOMPC, NompC, Nompc, nompc, ORE-13, TRPN, TRPN1	7227	**nompC**, CG17468, CT30855, Dmel\CG11020, DmNompC, jf24, l(2)25Dc, l(2)jf24, nomp, NOMPC, NompC, Nompc, nompc, ORE-13, TRPN, TRPN1, CG11020, Dmel_CG11020	No mechanoreceptor potential C, isoform G	UP000000803 Chromosome 2L
Q7KTN8	Q7KTN8_DROME	1,735 AA	Drosophila melanogaster (Fruit fly)	CG17468, CT30855, Dmel\CG11020, DmNompC, jf24, l(2)25Dc, l(2)jf24, nomp, NOMPC, NompC, Nompc, nompc, ORE-13, TRPN, TRPN1	7227	**nompC**, CG17468, CT30855, Dmel\CG11020, DmNompC, jf24, l(2)25Dc, l(2)jf24, nomp, NOMPC, NompC, Nompc, nompc, ORE-13, TRPN, TRPN1, CG11020, Dmel_CG11020	No mechanoreceptor potential C, isoform F	UP000000803 Chromosome 2L
A8DYV6	A8DYV6_DROME	1,726 AA	Drosophila melanogaster (Fruit fly)	CG17468, CT30855, Dmel\CG11020, DmNompC, jf24, l(2)25Dc, l(2)jf24, nomp, NOMPC, NompC, Nompc, nompc, ORE-13, TRPN, TRPN1	7227	**nompC**, CG17468, CT30855, Dmel\CG11020, DmNompC, jf24, l(2)25Dc, l(2)jf24, nomp, NOMPC, NompC, Nompc, nompc, ORE-13, TRPN, TRPN1, CG11020, Dmel_CG11020	No mechanoreceptor potential C, isoform D	UP000000803 Chromosome 2L
Q9VMR4	Q9VMR4_DROME	1,761 AA	Drosophila melanogaster (Fruit fly)	CG17468, CT30855, Dmel\CG11020, DmNompC, jf24, l(2)25Dc, l(2)jf24, nomp, NOMPC, NompC, Nompc, nompc, ORE-13, TRPN, TRPN1	7227	**nompC**, CG17468, CT30855, Dmel\CG11020, DmNompC, jf24, l(2)25Dc, l(2)jf24, nomp, NOMPC, NompC, Nompc, nompc, ORE-13, TRPN, TRPN1, CG11020, Dmel_CG11020	No mechanoreceptor potential C, isoform H	UP000000803 Chromosome 2L
E0A9E1	E0A9E1_DROME	1,732 AA	Drosophila melanogaster (Fruit fly)	CG17468, CT30855, Dmel\CG11020, DmNompC, jf24, l(2)25Dc, l(2)jf24, nomp, NOMPC, NompC, Nompc, ORE-13, TRPN, TRPN1	7227	**nompC**, CG17468, CT30855, Dmel\CG11020, DmNompC, jf24, l(2)25Dc, l(2)jf24, nomp, NOMPC, NompC, Nompc, ORE-13, TRPN, TRPN1, CG11020, Dmel_CG11020	No mechanoreceptor potential C isoform L[...]	UP000000803 Chromosome 2L