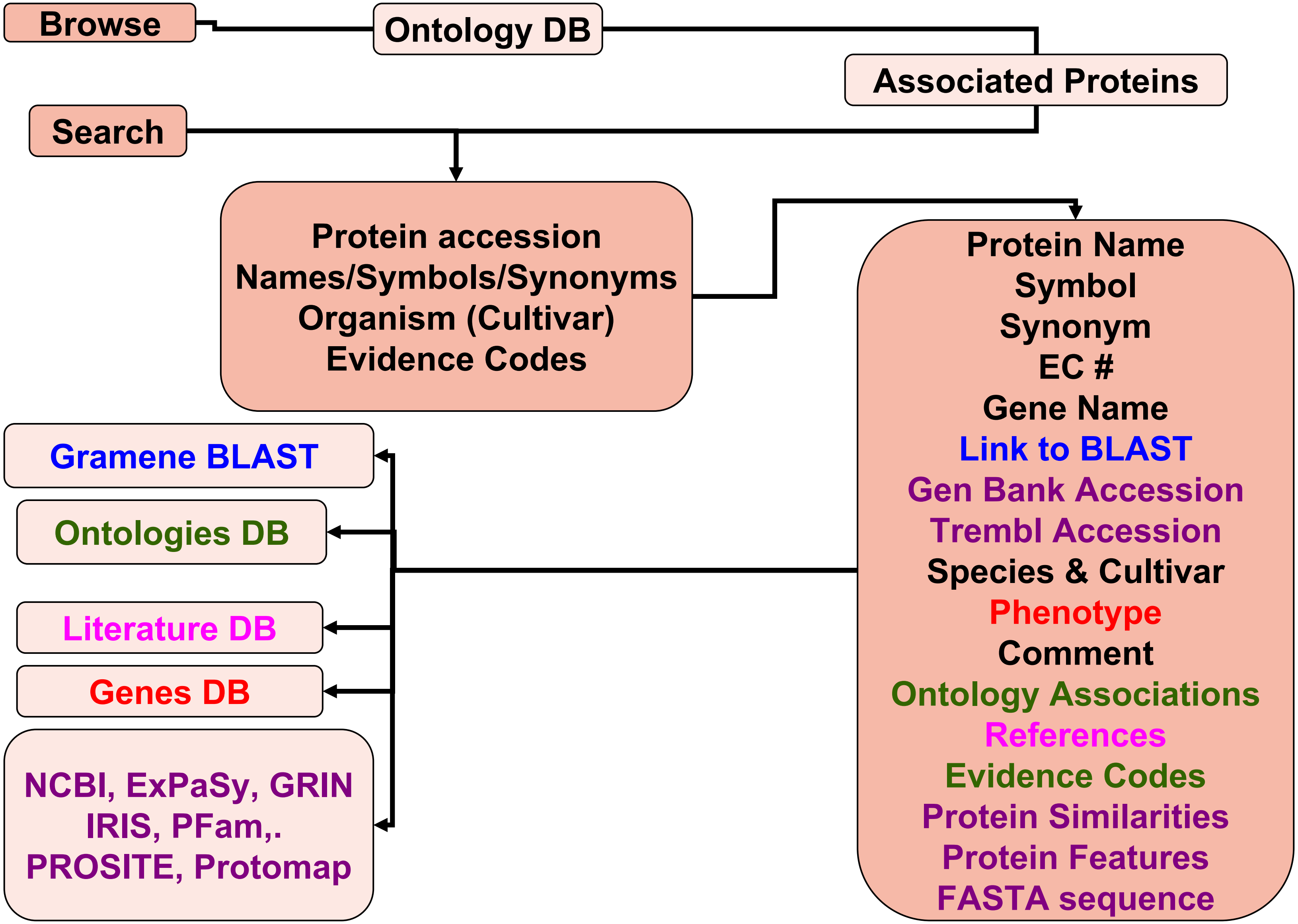
Welcome to the Protein Database Tutorial

This tutorial will describe how to navigate the section of Gramene that provides collective information on proteins from grasses (family *Poaceae/Gramineae)*. The protein entries are annotated by associating the following ontology concepts if known:

- Gene Ontology (GO)
    - Molecular function of the gene product.
    - Biological process in which the gene product is involved.
    - Cellular component where the gene product is localized.
- Plant Ontology
    - Plant structure where the gene is expressed (PO)
    - Plant growth stage at which the gene is expressed (GRO)

**** Only rice (Oryza) protein entries are manually curated.***

Detailed Information from Proteins Database



Browse

Ontology DB

Associated Proteins

Search

Protein accession
Names/Symbols/Synonyms
Organism (Cultivar)
Evidence Codes

Protein Name
Symbol
Synonym
EC #
Gene Name
Link to BLAST
Gen Bank Accession
Trembl Accession
Species & Cultivar
Phenotype
Comment
Ontology Associations
References
Evidence Codes
Protein Similarities
Protein Features
FASTA sequence

Gramene BLAST

Ontologies DB

Literature DB

Genes DB

NCBI, ExPaSy, GRIN
IRIS, PFam,.
PROSITE, Protomap

Tutorial Help

The hand icon indicates a link that allows you to go to the same page in your web browser.

If you are viewing this tutorial with Adobe Acrobat Reader, click the "bookmarks" on the left hand side of the Reader for easier navigation.

Action Options are noted in this type of font.

Notes or comments use this style font.

Gramene Home Page

GRAMENE *A Resource for Comparative Grass Genomics* *v20 (March 2006)*

Quick Search

Have Questions...?

- Gramene now has **tutorials** for every module.
- Ask questions through Feedback or Email

Gramene Tips:

The SSR Marker Search is listed under the Marker database.

All Tips

Quick Start

nced genomes for Rice, Maize & Arabidopsis; Look for rice/maize
h with GrameneMart; Search for sequence alignment with
tology.

m or ProSite or Browse by Gene Ontology using GO Slim.

l maps for Rice, Maize, Wheat, Barley, Oats, Sorghum, and
Compa Viewer (CMap) to compare maps of different

MOLECULAR MARK RIT); or
search by marker ty

TRAITS: Search the such as
Rice Genes, Rice QTL, Maize QTL. Don't forget to explore traits in Ontologies.

LITERATURE: Search the literature for your friends and topics of interest.

SUBMISSION: Submit a Rice Gene or Ontology Term to Gramene.

Gramene is a curated, open-source, web-accessible data resource for comparative genome analysis in the grasses. Our goal is to facilitate the study of cross-species homology relationships using information derived from public projects involved in genomic and EST sequencing, protein structure and function analysis, genetic and physical mapping, interpretation of biochemical pathways, gene and QTL localization and descriptions of phenotypic characters and mutations.

Featured News

- Gramene Release 20 for March 2006. See the release notes.
- Gramene News Archive
- RTWG Gramene workshop materials are available.
- Rice News Worldwide from IRRI

Visit with us at

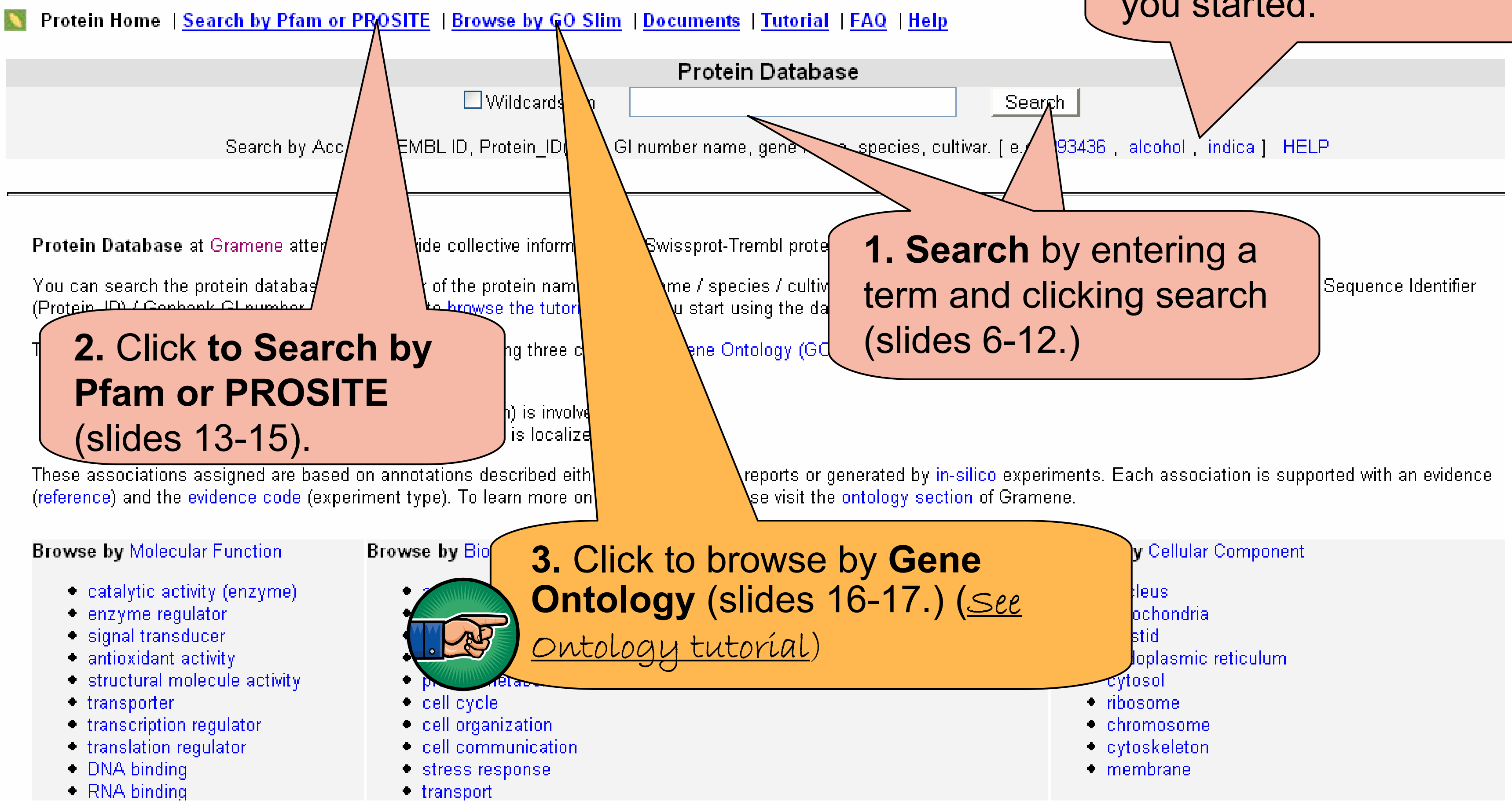
- 48th Maize Genetics Conference, March 9-12, 2006, Asilomar, Pacific Grove, CA

Gramene Calendar

Last modified: Thu Mar 2 15:54:55 2006

Click here to open Protein Home Page

Protein Home Page



Protein Home | Search by Pfam or PROSITE | Browse by GO Slim | Documents | Tutorial | FAQ | Help

Protein Database

☐ Wildcards Search

Search by Acc
EMBL ID, Protein_ID
GI number name, gene
species, cultivar. [e.
93436 , alcohol , indica] HELP

Protein Database at Gramene atte
ide collective inform
Swissprot-Trembl prote

You can search the protein databas
of the protein nam
me / species / cultiv
Sequence Identifier (Protein_ID) / Genbank GI number
to browse the tutori
u start using the da

T
ng three c
ene Ontology (GO

) is involve
is localize

These associations assigned are based on annotations described eith
reports or generated by in-silico experiments. Each association is supported with an evidence (reference) and the evidence code (experiment type). To learn more on
se visit the ontology section of Gramene.

Browse by Molecular Function

- catalytic activity (enzyme)
- enzyme regulator
- signal transducer
- antioxidant activity
- structural molecule activity
- transporter
- transcription regulator
- translation regulator
- DNA binding
- RNA binding

Browse by Bio

- cell cycle
- cell organization
- cell communication
- stress response
- transport

y Cellular Component

- cleus
- ochondria
- stid
- doplasmic reticulum
- cytosol
- ribosome
- chromosome
- cytoskeleton
- membrane

Possible searches. Click some examples to get you started.

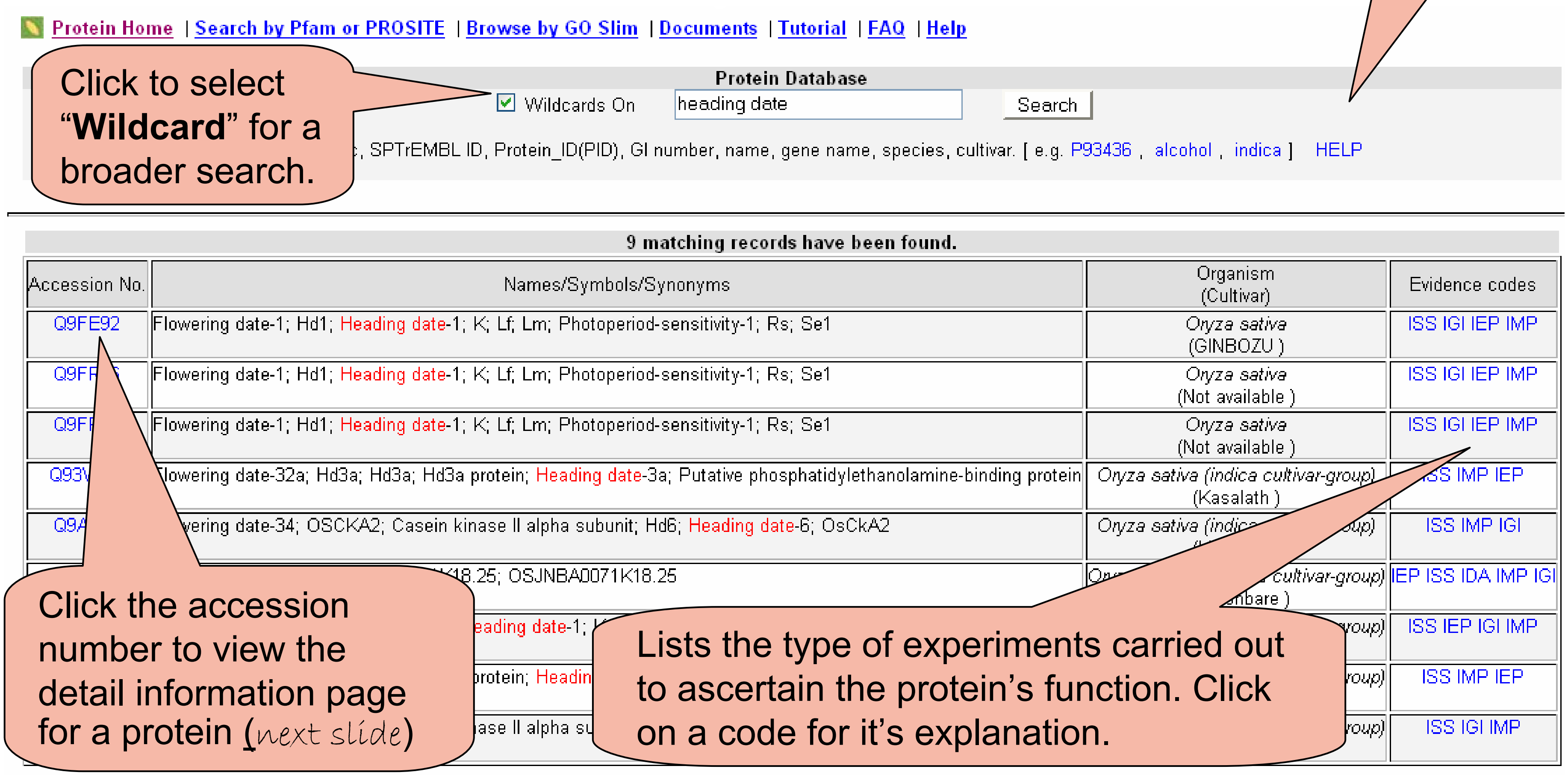
1. Search by entering a term and clicking search (slides 6-12.)

2. Click **to Search by Pfam or PROSITE** (slides 13-15).

3. Click to browse by **Gene Ontology** (slides 16-17.) (See Ontology tutorial)

1. Term Search: Results using “heading date”

Click for **Help Docs**.



Protein Home | **Search by Pfam or PROSITE** | **Browse by GO Slim** | **Documents** | **Tutorial** | **FAQ** | **Help**

Protein Database

☑ Wildcards On | heading date | Search

, SPTrEMBL ID, Protein_ID(PID), GI number, name, gene name, species, cultivar. [e.g. P93436 , alcohol , indica] HELP

9 matching records have been found.

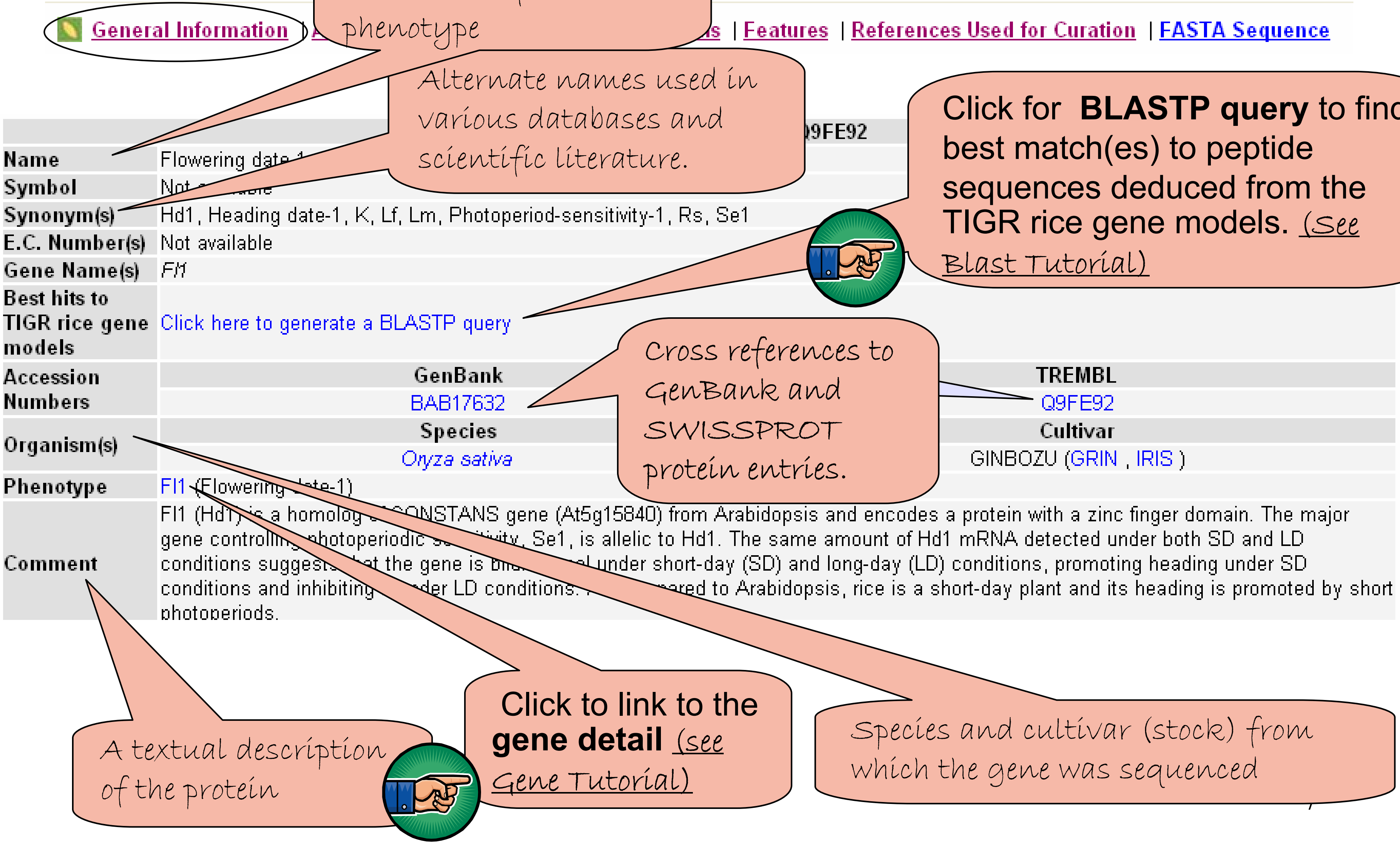
Accession No.	Names/Symbols/Synonyms	Organism (Cultivar)	Evidence codes
Q9FE92	Flowering date-1; Hd1; Heading date-1; K; Lf; Lm; Photoperiod-sensitivity-1; Rs; Se1	*Oryza sativa* (GINBOZU)	ISS IGI IEP IMP
Q9FF	Flowering date-1; Hd1; Heading date-1; K; Lf; Lm; Photoperiod-sensitivity-1; Rs; Se1	*Oryza sativa* (Not available)	ISS IGI IEP IMP
Q9FF	Flowering date-1; Hd1; Heading date-1; K; Lf; Lm; Photoperiod-sensitivity-1; Rs; Se1	*Oryza sativa* (Not available)	ISS IGI IEP IMP
Q93V	lowering date-32a; Hd3a; Hd3a; Hd3a protein; Heading date-3a; Putative phosphatidylethanolamine-binding protein	*Oryza sativa (indica cultivar-group)* (Kasalath)	SS IMP IEP
Q9A	vering date-34; OSCKA2; Casein kinase II alpha subunit; Hd6; Heading date-6; OsCkA2	*Oryza sativa (indic* *up)*	ISS IMP IGI
	K18.25; OSJNBA0071K18.25	*On* *cultivar-group)* nbare)	IEP ISS IDA IMP IGI
	eading date-1;	*roup)*	ISS IEP IGI IMP
	protein; Headin	*roup)*	ISS IMP IEP
	ase II alpha su	*roup)*	ISS IGI IMP

Click to select “**Wildcard**” for a broader search.

Click the accession number to view the detail information page for a protein (next slide)

Lists the type of experiments carried out to ascertain the protein’s function. Click on a code for it’s explanation.

Protein Detail Page: General Information



General Information | A… | …s | Features | References Used for Curation | FASTA Sequence

	…9FE92	
Name	Flowering date-1	
Symbol	Not available	
Synonym(s)	Hd1, Heading date-1, K, Lf, Lm, Photoperiod-sensitivity-1, Rs, Se1	
E.C. Number(s)	Not available	
Gene Name(s)	*Fl1*	
Best hits to TIGR rice gene models	Click here to generate a BLASTP query	
Accession Numbers	**GenBank**	**TREMBL**
	BAB17632	Q9FE92
Organism(s)	**Species**	**Cultivar**
	Oryza sativa	GINBOZU (GRIN , IRIS)
Phenotype	Fl1 (Flowering date-1)	
Comment	Fl1 (Hd1) is a homolog of CONSTANS gene (At5g15840) from Arabidopsis and encodes a protein with a zinc finger domain. The major gene controlling photoperiodic sensitivity, Se1, is allelic to Hd1. The same amount of Hd1 mRNA detected under both SD and LD conditions suggests that the gene is bifunctional under short-day (SD) and long-day (LD) conditions, promoting heading under SD conditions and inhibiting it under LD conditions. Compared to Arabidopsis, rice is a short-day plant and its heading is promoted by short photoperiods.	

Shows the name(s) of the protein molecule based on its function or phenotype

Alternate names used in various databases and scientific literature.

Click for **BLASTP query** to find best match(es) to peptide sequences deduced from the TIGR rice gene models. (See Blast Tutorial)

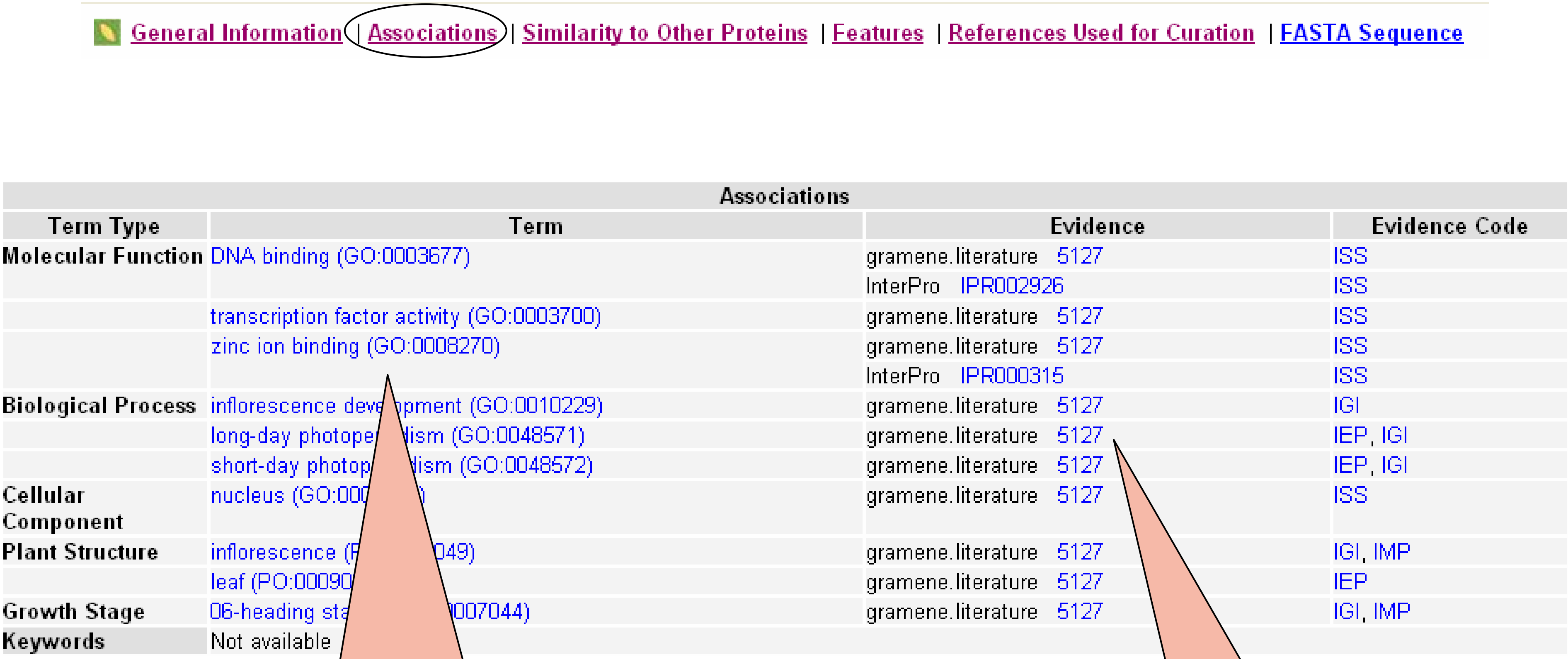
Cross references to GenBank and SWISSPROT protein entries.

A textual description of the protein

Click to link to the **gene detail** (see Gene Tutorial)

Species and cultivar (stock) from which the gene was sequenced

Protein Detail Page: Associations



General Information | Associations | Similarity to Other Proteins | Features | References Used for Curation | FASTA Sequence

Associations

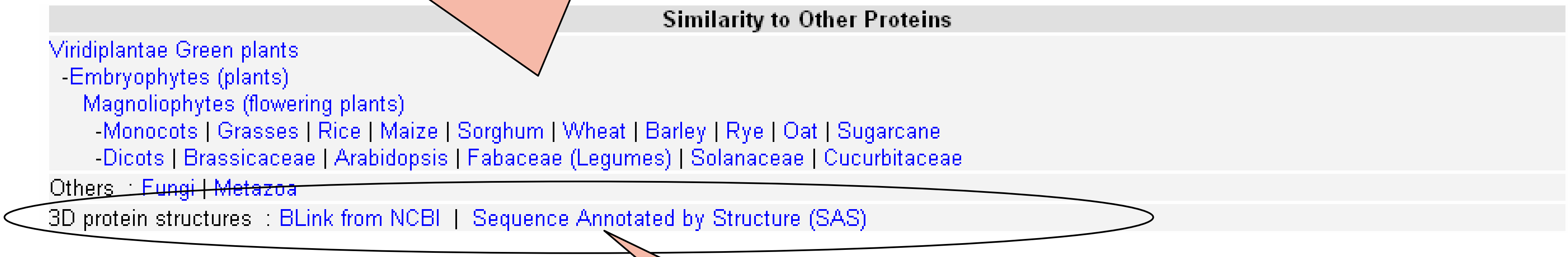
Term Type	Term	Evidence	Evidence Code
Molecular Function	DNA binding (GO:0003677)	gramene.literature 5127	ISS
		InterPro IPR002926	ISS
	transcription factor activity (GO:0003700)	gramene.literature 5127	ISS
	zinc ion binding (GO:0008270)	gramene.literature 5127	ISS
		InterPro IPR000315	ISS
Biological Process	inflorescence deve pment (GO:0010229)	gramene.literature 5127	IGI
	long-day photope ism (GO:0048571)	gramene.literature 5127	IEP, IGI
	short-day photop ism (GO:0048572)	gramene.literature 5127	IEP, IGI
Cellular Component	nucleus (GO:000)	gramene.literature 5127	ISS
Plant Structure	inflorescence (P 049)	gramene.literature 5127	IGI, IMP
	leaf (PO:00090	gramene.literature 5127	IEP
Growth Stage	06-heading sta 007044)	gramene.literature 5127	IGI, IMP
Keywords	Not available		

Click terms to view **Ontology Information** (See Ontology tutorial).

Click to view the **Reference(s)** used to assign the ontology term.

Protein Detail Page: Similarities

Click these options to link to NCBI's **BLink** to display the graphical output of pre-computed BLASTP results against the non-redundant (nr) protein database from NCBI.



Similarity to Other Proteins

Viridiplantae Green plants
-Embryophytes (plants)
Magnoliophytes (flowering plants)
-Monocots | Grasses | Rice | Maize | Sorghum | Wheat | Barley | Rye | Oat | Sugarcane
-Dicots | Brassicaceae | Arabidopsis | Fabaceae (Legumes) | Solanaceae | Cucurbitaceae
Others : Fungi | Metazoa
3D protein structures : BLink from NCBI | Sequence Annotated by Structure (SAS)

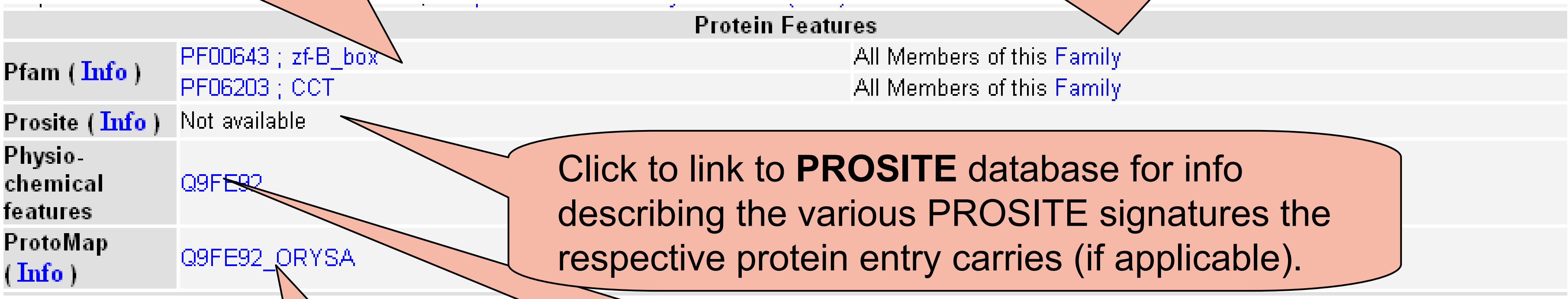
Click these options to find if there are any known proteins that share the **structural homology** with the given protein

Protein Detail Page: Features

General Information | Associations | Similarity to Other Proteins | Features | References Used for Curation | FASTA Sequence

Click to learn more about the **Protein Family** to which the protein belongs (if applicable).

Click to view a list of other protein entries that belong to this **Protein family** or share the Pfam/PROSITE **domain**.



Protein Features		
Pfam (Info)	PF00643 ; zf-B_box	All Members of this Family
	PF06203 ; CCT	All Members of this Family
Prosite (Info)	Not available	
Physio-chemical features	Q9FE92	
ProtoMap (Info)	Q9FE92_ORYSA	

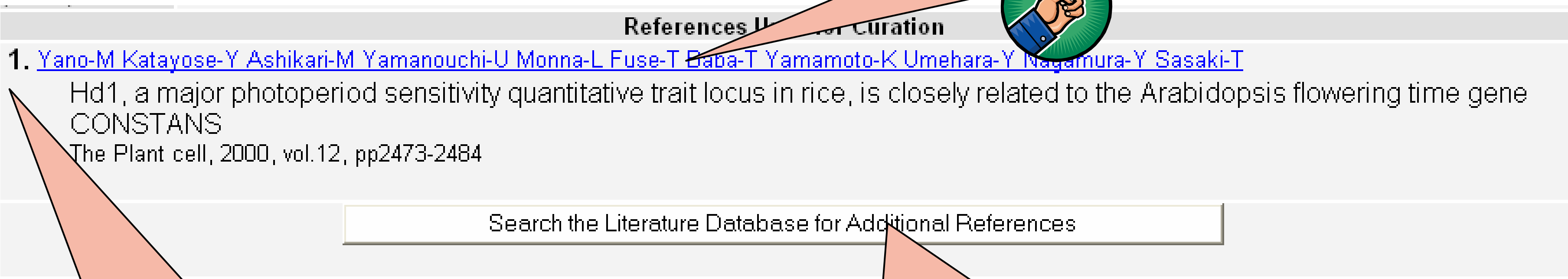
Click to link to **PROSITE** database for info describing the various PROSITE signatures the respective protein entry carries (if applicable).

Click to link to protein entry in **ProtoMap** database at Cornell University.

Click to link the **ExPAsy** server for analyses of physiochemical features analyzed by **ProtParam** tool.

Protein Detail Page: References

Click to see **Literature Detail Page** for that reference (See Literature tutorial).


References Used for Curation

1. Yano-M Katayose-Y Ashikari-M Yamanouchi-U Monna-L Fuse-T Baba-T Yamamoto-K Umehara-Y Nagamura-Y Sasaki-T
Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS
The Plant cell, 2000, vol.12, pp2473-2484

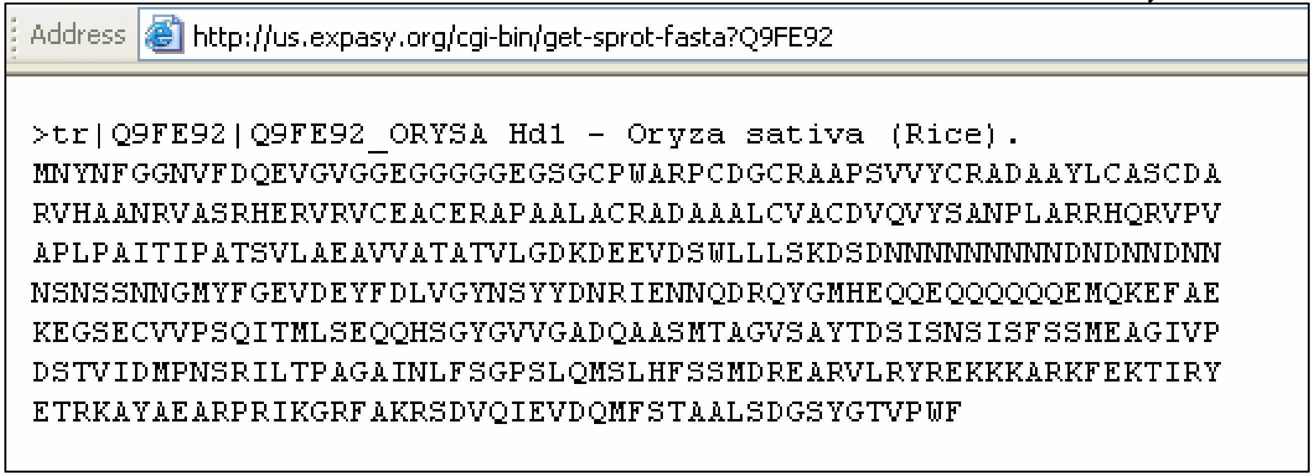
Lists of all the references used for annotation of the selected protein.

Click to search for **more references** on this protein that are in the literature database but were not cited in the curation (See Literature tutorial).

Protein Detail Page: FASTA Sequence

General Information | Associations | Similarity to Other Proteins | Features | References Used for Curation | FASTA Sequence

Click here to download the sequence in FASTA format for further analysis or your research purposes.



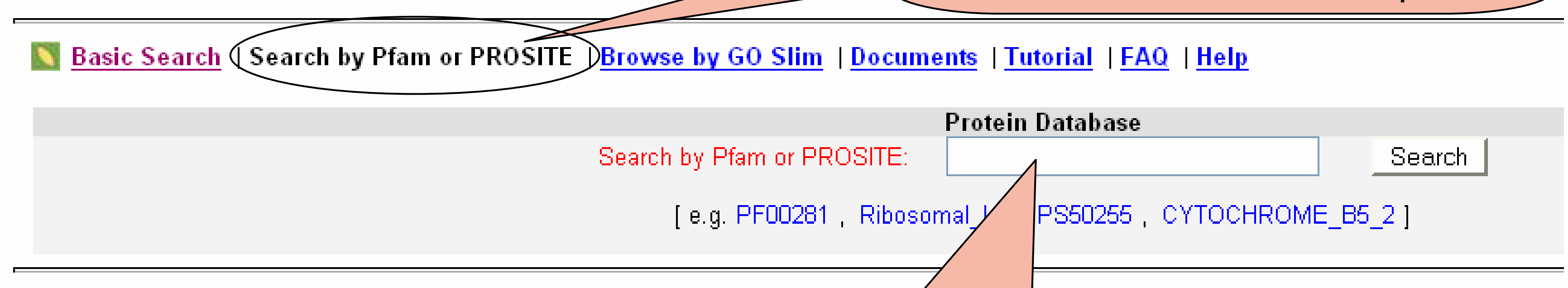
http://us.expasy.org/cgi-bin/get-sprot-fasta?Q9FE92

```
>tr|Q9FE92|Q9FE92_ORYSA Hd1 - Oryza sativa (Rice).
MNYNFGGNVFDQEVGVGGEGGGGGEGSGCPWARPCDGCRAAPSVVYCRADAAYLCASCDA
RVHAANRVASRHERVRVCEACERAPAALACRADAAALCVACDVQVYSANPLARRHQRVPV
APLPAITIPATSVLAEAVVATATVLGDKDEEVDSWLLLSKDSDNNNNNNNNNDNDNNDNN
NSNSSNNGMYFGEVDEYFDLVGYNSYYDNRIENNQDRQYGMHEQQEQQQQQEMQKEFAE
KEGSECVVPSQITMLSEQQHSGYGVVGADQAASMTAGVSAYTDSISNSISFSSMEAGIVP
DSTVIDMPNSRILTPAGAINLFSGPSLQMSLHFSSMDREARVLRYREKKKARKFEKTIRY
ETRKAYAEARPRIKGRFAKRSDVQIEVDQMFSTAALSDGSYGTVPWF
```

For bulk download of all the sequences please visit our ftp site ftp://ftp.gramene.org/pub/gramene/CURRENT_RELEASE/data/protein

Search by Pfam or PROSITE

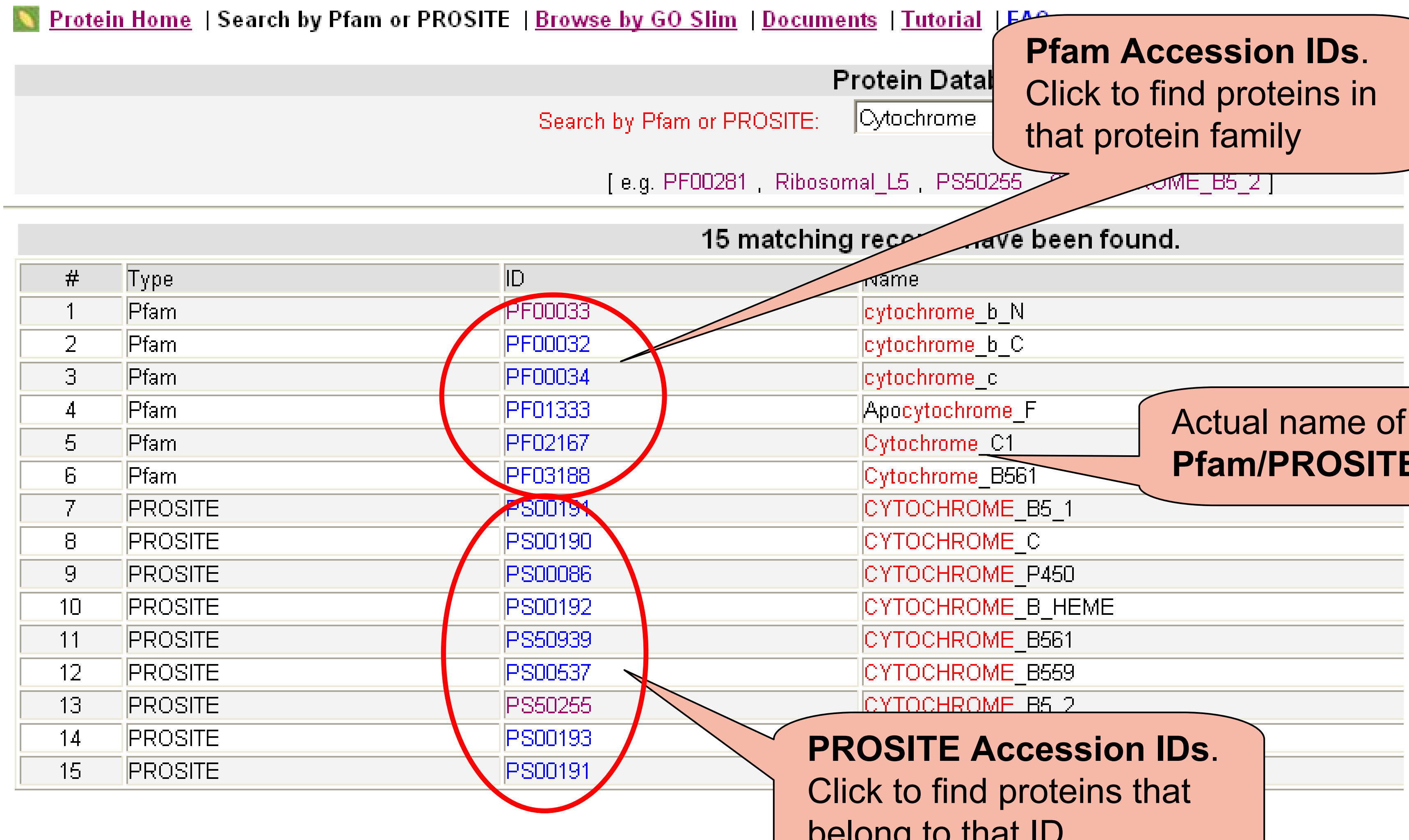
From the protein search section, click on the “**Search by Pfam or PROSITE**” for this search option.



Search Pfam / PROSITE by either a name or an id as given in examples.

****Note: Some proteins do not have associations to Pfam and PROSITE features. This will restrict your search.***

Results of Pfam or PROSITE Search



Protein Datal

Search by Pfam or PROSITE: Cytochrome

[e.g. PF00281 , Ribosomal_L5 , PS50255 , C OME_B5_2]

15 matching rec ave been found.

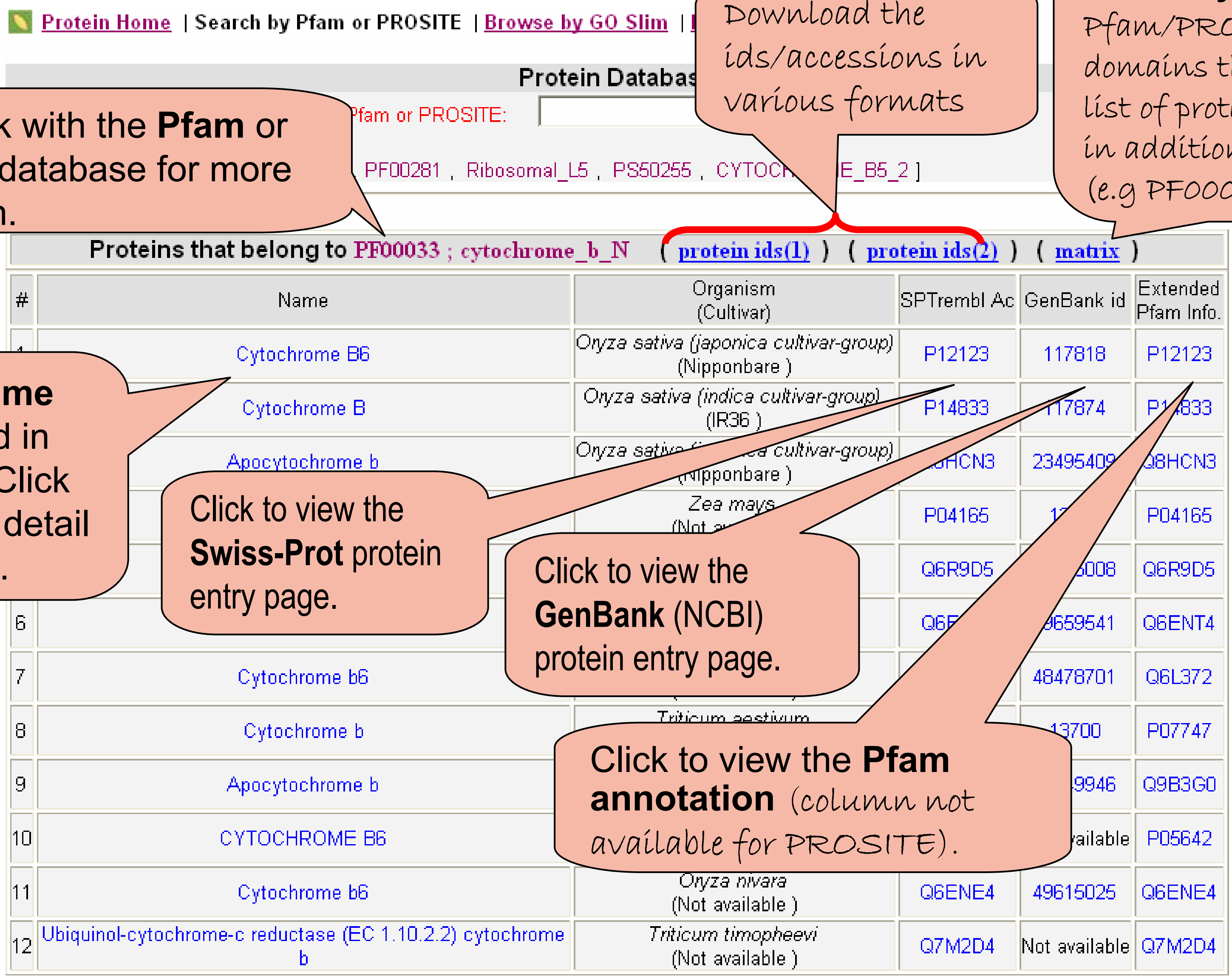
#	Type	ID	Name
1	Pfam	PF00033	cytochrome_b_N
2	Pfam	PF00032	cytochrome_b_C
3	Pfam	PF00034	cytochrome_c
4	Pfam	PF01333	Apocytochrome_F
5	Pfam	PF02167	Cytochrome_C1
6	Pfam	PF03188	Cytochrome_B561
7	PROSITE	PS00191	CYTOCHROME_B5_1
8	PROSITE	PS00190	CYTOCHROME_C
9	PROSITE	PS00086	CYTOCHROME_P450
10	PROSITE	PS00192	CYTOCHROME_B_HEME
11	PROSITE	PS50939	CYTOCHROME_B561
12	PROSITE	PS00537	CYTOCHROME_B559
13	PROSITE	PS50255	CYTOCHROME_B5_2
14	PROSITE	PS00193	
15	PROSITE	PS00191	

Pfam Accession IDs. Click to find proteins in that protein family

Actual name of the **Pfam/PROSITE Class.**

PROSITE Accession IDs. Click to find proteins that belong to that ID.

List of Proteins belonging to Pfam or PROSITE Entry



Proteins that belong to PF00033 ; cytochrome_b_N (protein ids(1)) (protein ids(2)) (matrix)

#	Name	Organism (Cultivar)	SPTrembl Ac	GenBank id	Extended Pfam Info.
1	Cytochrome B6	*Oryza sativa (japonica cultivar-group)* (Nipponbare)	P12123	117818	P12123
	Cytochrome B	*Oryza sativa (indica cultivar-group)* (IR36)	P14833	17874	P1 833
	Apocytochrome b	*Oryza sativa a cultivar-group)* Nipponbare)	HCN3	23495409	Q8HCN3
		Zea mays (Not a	P04165	1	P04165
			Q6R9D5	008	Q6R9D5
6			Q6E	9659541	Q6ENT4
7	Cytochrome b6			48478701	Q6L372
8	Cytochrome b	*Triticum aestivum*		13700	P07747
9	Apocytochrome b			9946	Q9B3G0
10	CYTOCHROME B6			ailable	P05642
11	Cytochrome b6	*Oryza nivara* (Not available)	Q6ENE4	49615025	Q6ENE4
12	Ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b	*Triticum timopheevi* (Not available)	Q7M2D4	Not available	Q7M2D4

Download the ids/accessions in various formats

Matrix displays **summary** of all the Pfam/PROSITE domains the given list of proteins share in addition to this (e.g PF00033).

Click to link with the **Pfam** or **PROSITE** database for more information.

Protein Name as recorded in Gramene. Click to view the detail information.

Click to view the **Swiss-Prot** protein entry page.

Click to view the **GenBank** (NCBI) protein entry page.

Click to view the **Pfam annotation** (column not available for PROSITE).

Browse by GO Slim

Browse by Molecular Function

- catalytic activity (enzyme)
- enzyme regulator
- signal transducer
- antioxidant activity
- structural molecule activity
- transporter
- transcription regulator
- translation regulator
- DNA binding
- RNA binding
- nucleotide binding

Browse by Biological Process

- amino acid metabolism
- carbohydrate metabolism
- lipid metabolism
- nucleic acid metabolism
- protein metabolism
- cell cycle
- cell organization
- cell communication
- stress response
- transport
- electron transprot
- photosynthesis
- energy pathway

Browse by Cellular Component

- nucleus
- mitochondria
- plastid
- endoplasmic reticulum
- cytosol
- ribosome
- chromosome
- cytoskeleton

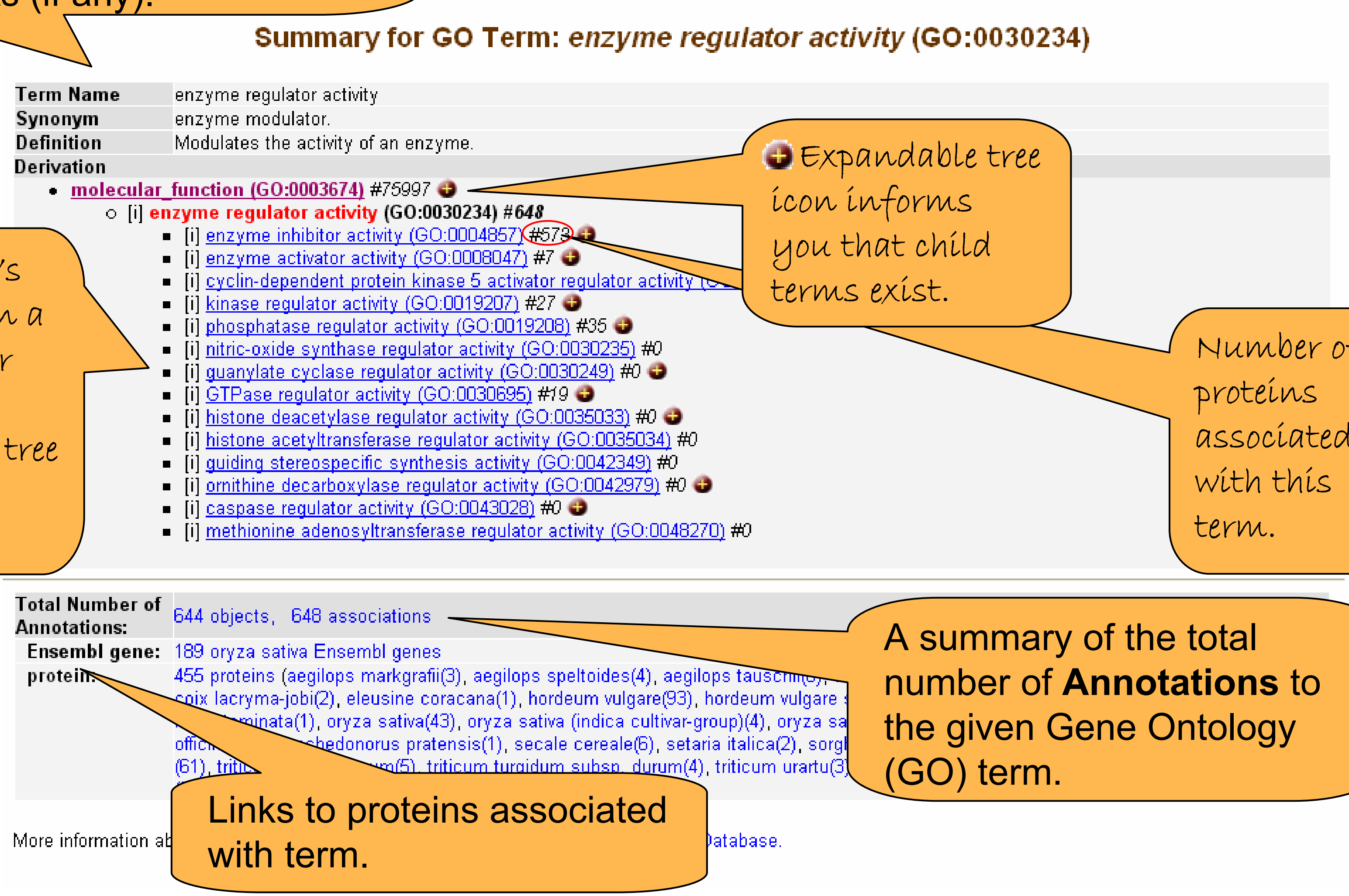
Click to search category in GO (See Ontology tutorial).

Click a term to quickly **browse the Ontology database** to find a list of proteins associated with these particular keywords (See Ontology tutorial).

Note: Many proteins do not have associations to Gene Ontology (GO) terms. This will restrict your search.

Results for GO Slim

The **GO term details** include name, synonyms, definition and comments (if any).



Summary for GO Term: *enzyme regulator activity* (GO:0030234)

Term Name	enzyme regulator activity
Synonym	enzyme modulator.
Definition	Modulates the activity of an enzyme.

Derivation

- molecular_function (GO:0003674) #75997 ⊕
  - [i] **enzyme regulator activity (GO:0030234) #648**
    - [i] enzyme inhibitor activity (GO:0004857) #573 ⊕
    - [i] enzyme activator activity (GO:0008047) #7 ⊕
    - [i] cyclin-dependent protein kinase 5 activator regulator activity
    - [i] kinase regulator activity (GO:0019207) #27 ⊕
    - [i] phosphatase regulator activity (GO:0019208) #35 ⊕
    - [i] nitric-oxide synthase regulator activity (GO:0030235) #0
    - [i] guanylate cyclase regulator activity (GO:0030249) #0 ⊕
    - [i] GTPase regulator activity (GO:0030695) #19 ⊕
    - [i] histone deacetylase regulator activity (GO:0035033) #0 ⊕
    - [i] histone acetyltransferase regulator activity (GO:0035034) #0
    - [i] guiding stereospecific synthesis activity (GO:0042349) #0
    - [i] ornithine decarboxylase regulator activity (GO:0042979) #0 ⊕
    - [i] caspase regulator activity (GO:0043028) #0 ⊕
    - [i] methionine adenosyltransferase regulator activity (GO:0048270) #0

Total Number of Annotations:	644 objects, 648 associations
Ensembl gene:	189 oryza sativa Ensembl genes
protein:	455 proteins (aegilops markgrafii(3), aegilops speltoides(4), aegilops tausch coix lacryma-jobi(2), eleusine coracana(1), hordeum vulgare(93), hordeum vulgare minata(1), oryza sativa(43), oryza sativa (indica cultivar-group)(4), oryza sa offic hedonorus pratensis(1), secale cereale(6), setaria italica(2), sorg (61), trit m(5), triticum turgidum subsp. durum(4), triticum urartu(3

More information ab atabase.

⊕ Expandable tree icon informs you that child terms exist.

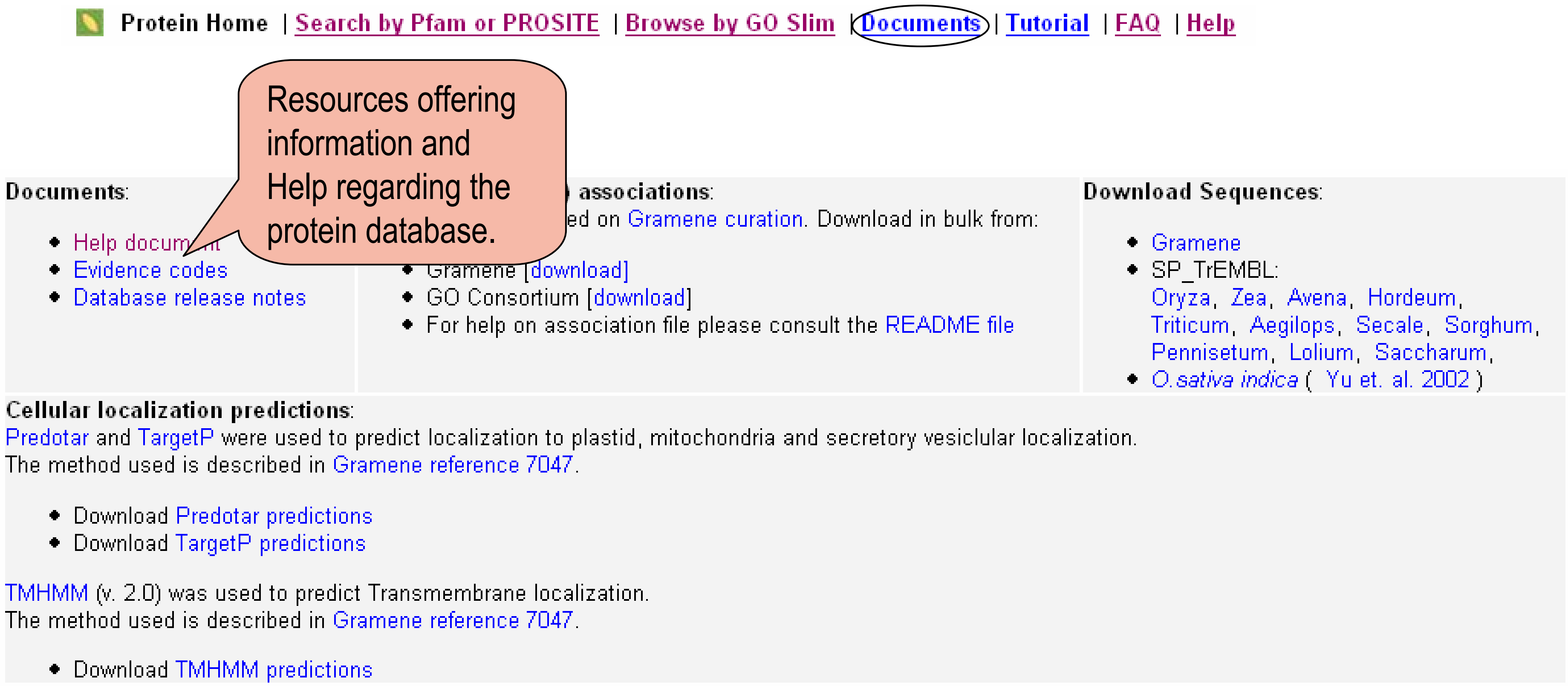
The term's lineage in a molecular function category tree and its children.

Number of proteins associated with this term.

A summary of the total number of **Annotations** to the given Gene Ontology (GO) term.

Links to proteins associated with term.

Documents



Documents:

- Help docum nt
- Evidence codes
- Database release notes

) associations:

ed on Gramene curation. Download in bulk from:

- Gramene [download]
- GO Consortium [download]
- For help on association file please consult the README file

Download Sequences:

- Gramene
- SP_TrEMBL:
  Oryza, Zea, Avena, Hordeum, Triticum, Aegilops, Secale, Sorghum, Pennisetum, Lolium, Saccharum,
- *O.sativa indica* (Yu et. al. 2002)

Cellular localization predictions:

Predotar and TargetP were used to predict localization to plastid, mitochondria and secretory vesiclular localization.
The method used is described in Gramene reference 7047.

- Download Predotar predictions
- Download TargetP predictions

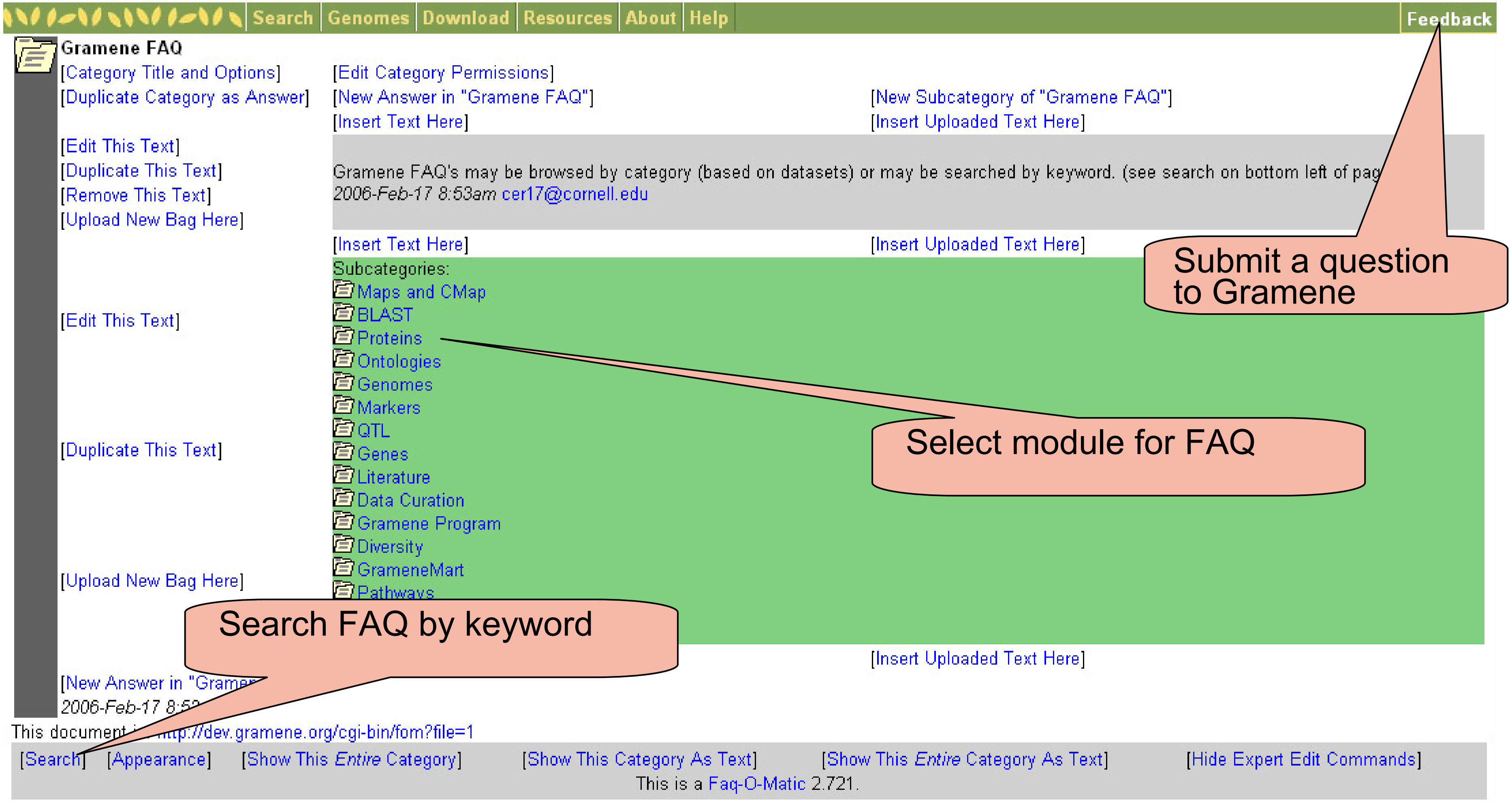
TMHMM (v. 2.0) was used to predict Transmembrane localization.
The method used is described in Gramene reference 7047.

- Download TMHMM predictions

Resources offering information and Help regarding the protein database.

Tutorial & FAQ

Tutorials offered in PDF, Powerpoint, or HTML.



Gramene FAQ

[Category Title and Options] [Edit Category Permissions]

[Duplicate Category as Answer] [New Answer in "Gramene FAQ"] [New Subcategory of "Gramene FAQ"]

[Insert Text Here] [Insert Uploaded Text Here]

[Edit This Text]
[Duplicate This Text]
[Remove This Text]
[Upload New Bag Here]

Gramene FAQ's may be browsed by category (based on datasets) or may be searched by keyword. (see search on bottom left of pag

2006-Feb-17 8:53am cer17@cornell.edu

[Insert Text Here] [Insert Uploaded Text Here]

[Edit This Text]
[Duplicate This Text]
[Upload New Bag Here]

Subcategories:
- Maps and CMap
- BLAST
- Proteins
- Ontologies
- Genomes
- Markers
- QTL
- Genes
- Literature
- Data Curation
- Gramene Program
- Diversity
- GrameneMart
- Pathways

[Insert Uploaded Text Here]

[New Answer in "Gramen

2006-Feb-17 8:5

This document i http://dev.gramene.org/cgi-bin/fom?file=1

[Search] [Appearance] [Show This *Entire* Category] [Show This Category As Text] [Show This *Entire* Category As Text] [Hide Expert Edit Commands]

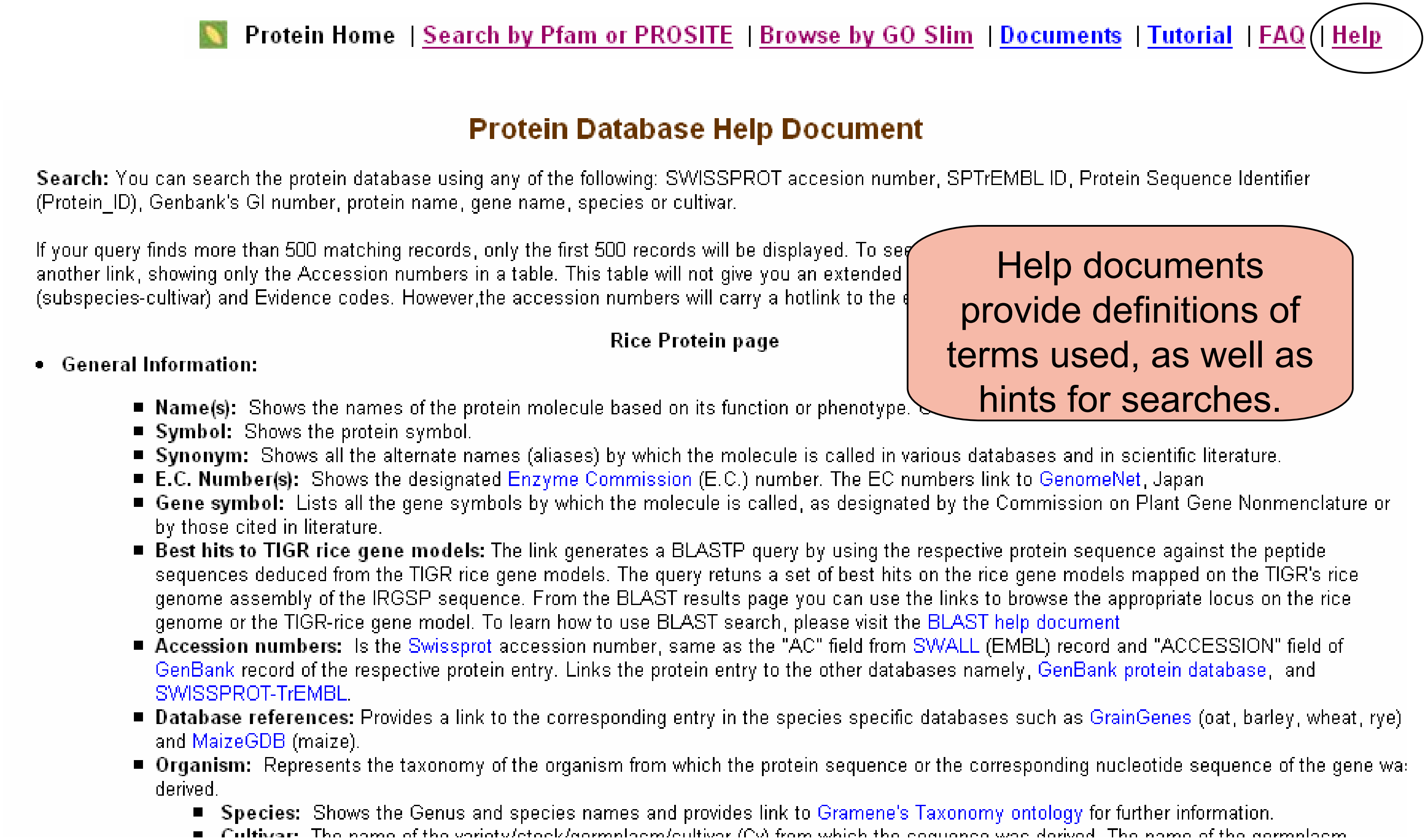
This is a Faq-O-Matic 2.721.

Submit a question to Gramene

Select module for FAQ

Search FAQ by keyword

Help



Protein Home | Search by Pfam or PROSITE | Browse by GO Slim | Documents | Tutorial | FAQ | Help

Protein Database Help Document

Search: You can search the protein database using any of the following: SWISSPROT accesion number, SPTrEMBL ID, Protein Sequence Identifier (Protein_ID), Genbank's GI number, protein name, gene name, species or cultivar.

If your query finds more than 500 matching records, only the first 500 records will be displayed. To see
another link, showing only the Accession numbers in a table. This table will not give you an extended
(subspecies-cultivar) and Evidence codes. However,the accession numbers will carry a hotlink to the

Rice Protein page

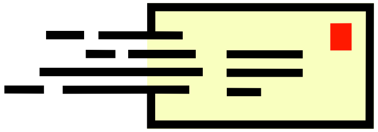
- **General Information:**
  - **Name(s):** Shows the names of the protein molecule based on its function or phenotype.
  - **Symbol:** Shows the protein symbol.
  - **Synonym:** Shows all the alternate names (aliases) by which the molecule is called in various databases and in scientific literature.
  - **E.C. Number(s):** Shows the designated Enzyme Commission (E.C.) number. The EC numbers link to GenomeNet, Japan
  - **Gene symbol:** Lists all the gene symbols by which the molecule is called, as designated by the Commission on Plant Gene Nonmenclature or by those cited in literature.
  - **Best hits to TIGR rice gene models:** The link generates a BLASTP query by using the respective protein sequence against the peptide sequences deduced from the TIGR rice gene models. The query retuns a set of best hits on the rice gene models mapped on the TIGR's rice genome assembly of the IRGSP sequence. From the BLAST results page you can use the links to browse the appropriate locus on the rice genome or the TIGR-rice gene model. To learn how to use BLAST search, please visit the BLAST help document
  - **Accession numbers:** Is the Swissprot accession number, same as the "AC" field from SWALL (EMBL) record and "ACCESSION" field of GenBank record of the respective protein entry. Links the protein entry to the other databases namely, GenBank protein database, and SWISSPROT-TrEMBL.
  - **Database references:** Provides a link to the corresponding entry in the species specific databases such as GrainGenes (oat, barley, wheat, rye) and MaizeGDB (maize).
  - **Organism:** Represents the taxonomy of the organism from which the protein sequence or the corresponding nucleotide sequence of the gene wa derived.
    - **Species:** Shows the Genus and species names and provides link to Gramene's Taxonomy ontology for further information.
    - **Cultivar:** The name of the variety/stock/germplasm/cultivar (Cv) from which the sequence was derived. The name of the germplasm

Help documents provide definitions of terms used, as well as hints for searches.

Contact Gramene

Use the feedback button, located at the top of every page, to provide feedback or to ask questions about Gramene.

or



Email Gramene at gramene@gramene.org