# **Functional Annotation**



### May 23, 2007

Rama Maiti

# **Functional Annotation Overview**

- What is annotation
- Steps we take to annotate eukaryotic genes
- Software tools we use for functional annotation
- Steps we take to manually annotate or verify an automated annotation

# What are the questions?

- How did the gene get its structure and name?
- Does it really have a function assigned to it?
- Where did this information come from?
- Is it accurate? Can you rely on it?

# What is Functional Annotation?

- "To annotate" is "to make or furnish critical or explanatory notes or comments"
- For genomics the 'notes' are about
  - Names of the gene products
  - Functions of genes within an organism
- Elements of the functional annotation process
  - Validation of the gene structure
  - Literature search, if any is available
  - Homology / domain searches
  - Assignment of function
  - Maintenance of data availability

# **The Annotation Pipeline**



protein searches (nap) EST, cDNA alignments (gap2) custom database searches (nap/gap2) gene prediction algorithms Blastp searches HMM searches SignalP/TargetP/Interpro

Human intervention: -critical evaluation of automated assignments

# Manual vs. Automated Annotation

- Automated Annotation is complicated by high volumes of data derived from different methods at different centers
- High quality annotation requires manual review and intervention.

# **Steps in Functional Annotation**



# **Steps in Functional Annotation**

- Analyze the gene structure (Annotation Station or preferred gene viewer)
- Name the gene product (Manatee)
  - requires analysis of the gene product
  - gene product name is primarily homology based on different bioinformatics tools
- Assign Gene Ontology terms
  - Process
  - Function
  - Component

# Homology Searching (Tools that are available to characterize a sequence)

- WU BLAST <u>http://blast.wustl.edu/</u> with links to many servers
- NCBI BLAST <u>http://www.ncbi.nlm.nih.gov/blast/</u>
- **Pfam profiles** (profiles, or HMMs) <u>http://pfam.wustl.edu/</u>
- TIGRFAMS (profiles, or HMMs) <u>http://tigrblast.tigr.org/web-hmm/</u>
- **Prosite** (profiles & families) <u>http://ca.expasy.org/tools/scanprosite/</u>
- **Interpro** (families) <u>http://www.ebi.ac.uk/InterProScan/</u>
- TmHMM (transmembrane domain) <u>http://www.cbs.dtu.dk/services/TMHMM/</u>
- Swiss-Prot <u>http://au.expasy.org/sprot/</u>
- SignalP (signal peptide cleavage sites) <u>http://www.cbs.dtu.dk/services/SignalP/</u>
- TargetP (subcellular location) <u>http://www.cbs.dtu.dk/services/TargetP/</u>
- **PSI-BLAST** (NCBI) link at <u>http://www.ncbi.nlm.nih.gov/BLAST/</u>
- Protein families and clustering
  - **TIGR Paralogous Families** (not yet available outside of TIGR)
  - TribeMCL <u>http://www.ebi.ac.uk/research/cgg/tribe/</u>
- **TIGR Rice Workshop**

# Manatee

- Manatee is a web-based gene ٠ evaluation and genome annotation tool.
- Manatee displays the current ٠ annotation for prokaryotic and eukaryotic genomes.
- Manatee is an open source software ٠ available at

http://sourceforge.net/projects/manatee/

Oryza sativa spp japonica cv. Nipponbare Gene Curation Page Home | Logged into [osa1] as ecaler C GENE CURATION INFORMATION ß database: osa1 11667.t00202 () end5/end3: 1151896 / 1142631 Model: 11667.m00212 feat name / locus: Pub Locus:LOC Os01g03020 gene length: 2838 View Blastp Searches New Gene protein length: 946 seg id: 11667 molecular wt: 107175.54 Select Function T gene list pager pI: 6.67 ▶ Reload Page << 11667.m00211 11667.m00213 >> 3000 4000 5000 6000 7000 0 1000 2000 00 9000 8000 Gene Synonyms Alternative Splicing None None (Scale shown in nucleotides, CURATION STATUS submit | reset gene annotation curated gene structure curated pseudogene 5' partial 3' partial GENE IDENTIFICATION submit | reset | history | alias gene name aliases gene name product name product name aliases leucyl-tBNA synthetase make high quality functional assignments using genome gene symbol aliases analyses tools. These tools consist of, but are not limited to GO classifications, blast search data, protein families. ec number aliases

**TIGR Rice Workshop** 

### Verify evidence from automated annotation

- -BLAST matches
- -HMM
- -Prosite, Interpro classifications
- -Motifs
- -Signal Sequence
- -Target Sequence
- -Transmembrane domain
- -Protein families



### **Functional annotation**

#### Examine the gene structure

- does it make sense with respect to the alignments? do you need to re-curate the gene
- structure?

#### Name the gene product

Determine whether it is published, Fully characterized? Give it the Swiss-Prot name. Sequenced but not characterized? Look at the evidence.

#### Add comments to comment field

explain reasoning for othersadd personal communication informationmake comments about function or process

In many occasions after analyzing our data and make a decision about a gene function, we may need to go back and re-examine the gene structure.

0						
GENE CURATION INFORMA	ATION					6
11667.t00202() Model: 11667.m0021 Pub Locus:LOC_0501g0 > View Blastp Search seq id: 11667 Select Function > Reload Page	2 13020 Nes	end5/end3: 1151896/1142631 gene length: 2838 protein length: 946 molecular wt: 107175.54 pl: 6.67	feat_na	database ume / locus <u>New C</u> gene list	e: osa1 s: cosa1 s: c	213 >>
		0 1000 2000 3000	4000 5000 6	200 7000	8000	9000
Gene Synonyms Alternative	Splicing		۱۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰	••••	• •	-f
INONE	e	(Scale sh	own in nucleotide:	s.)		
CURATION STATUS					subm	it   rese
gene structure curated		gene annotation curated		∏ p:	seudogene	
1 5' partiai		3 partia				
GENE IDENTIFICATION			submit	reset	history	alias
gene name				gene	name aliase	<u>IS</u>
product name				produ	ict name ali	ases
seesy and synatetase						
gene symbol				gene	symbol alia	ses
ec number				ec nu	mber aliase	<u>s</u>
comment:						
AUTOUPDATE: Updated st	tructure (	of 2699.m00121. ;				
pub_comment:						
pub_comment:				_		
pub_comment:				-		
pub_comment:				_		
pub_comment:				_		
pub_comment:						
pub_comment:						
auto comment:			submit	reset	go sug   se	arch
pub_comment: * auto comment GENE ONTOLOGY delete go id			submit	reset   assigned by	go sug   se assign date	arch
pub_comment: * auto comment GENE_ONTOLOGY delete go id GO:0000166_add	edit (F) m	ucleotide binding	submit	reset   assigned by autoGO	go sug seign date 11/08/04	arch
pub_comment: > auto comment GENE_ONTOLOGY delete go id □ GO:0000166_add □ GO:0003824_add	edit (F) m edit (F) c	ucleotide binding stalytic activity unleakese municatide municatide and	submit	reset   assigned by autoGO autoGO	go sug   se assign date 11/08/04 11/08/04	arch   evidend
pub_comment: > auto comment GENE_ONTOLOGY delete go id GO:000166_add _ GO:0003824_add _ GO:0006139_add _ _	edit (F) m edit (F) c edit (m tai	ucleotide binding stalytic activity ucleobase, nucleoside, nucleotide and n	submit ucleic acid	reset   assigned by autoGO autoGO sutoGO	go sug   se assign date 11/08/04 11/08/04 11/08/04	arch
pub_comment: Pauto comment GENE_ONTOLOGY delete go id GO:000166_add GO:0001824_add GO:000139_add GO:000612_add	edit (F) m edit (F) c edit (P) m edit (P) m edit (P) m	ucleotide binding atalytic activity ucleobase, nucleoside, nucleotide and n offism ordein biosynthesis	submit ucleic acid	reset   assigned by autoGO autoGO autoGO autoGO	go sug   se assign date 11,08,04 11,08,04 11,08,04 11,08,04	arch   eviden
pub_comment:           > auto comment           GENE ONTOLOGY           delete         go id           GO:3000166         add _	edit (F) m edit (F) c metait (P) m edit (P) m edit (P) p edit (P) a	ucleotide binding atalytic activity ucleobase, nucleoside, nucleotide and n olism rotein biosynthesis mino acid and derivative metabolism	submit ucleic scid	reset 1 assigned by autoGO autoGO autoGO autoGO autoGO	go sug   se assign date 11,08,04 11,08,04 11,08,04 11,08,04 11,08,04	arch
pub_comment:           > auto comment           GENE_ONTOLOGY           delete         go id           GO:0000166         add           GO:00003824         add           GO:00006139         add           GO:0006519         add	edit (F) m edit (F) c edit (PP) m edit (P) pa	ucleotide binding stalytic activity ucleobase, nucleoside, nucleotide and n bolism rotein biosynthesis mino acid and derivative metabolism	submit ucleic acid	reset   assigned by autoGO autoGO autoGO autoGO autoGO	go sug   se assign date 11,08,04 11,08,04 11,08,04 11,08,04	arch
pub_comment: > auto comment GENE_ONTOLOGY delete go id GO:0000166_add GO:0000824_add GO:00008129_add GO:0006519_add Function	edit (F) m edit (F) c edit (P) m edit (P) m edit (P) m edit (P) m process	ucleotide binding atalytic activity ucleobase, nucleoside, nucleoside and n bolism mino acid and derivative metabolism component	submit ucleic acid others	reset   assigned by autoGO autoGO autoGO autoGO autoGO	go sug 1 se assign date 11/08/04 11/08/04 11/08/04	arch   eviden
pub_comment: > auto comment GENE_ONTOLOGY delete go id	edit (F) m edit (F) c edit (P) m edit (P) m edit (P) a process	ucleotide binding atalytic activity ucleobase, nucleoside, nucleotide and n bolism mino acid and derivative metabolism <b>component</b>	submit ucleic acid others	reset   by autoGO autoGO autoGO autoGO	go sug   se assign date 11/08/04 11/08/04 11/08/04	arch   eviden
pub comment: auto comment GENE ONTOLOGY delete go id GO:0000166 add GO:0000166 add GO:0000163 add GO:00006139 add GO:0000619 add GO:0000619 add GO:0000619 add GO:0000619 add GO:0000619 add GO:0000619 add GO:0000619 add	edit (F) m edit (F) ci edit (P) m edit (P) sa edit (P) a process	ucieotide binding atalytic activity ucieobase, nucleoside, nucleostide and n bolism mino acid and derivative metabolism component component reference	submit ucleic acid others	reset   assigned by autoGO autoGO autoGO autoGO	go sug   se assign date 11/08/04 11/08/04 11/08/04	arch   evidence
pub comment:	edit (F) m edit (F) c edit (P) m metal edit (P) p edit (P) a process	ucieotide binding, atalytic activity ucieobase, nucleoside, nucleoside and n bolism mino acid and derivative metabolism component reference	submit ueleic acid others vit	reset   assigned by autoGO autoGO autoGO autoGO	go sug   se assign date 11/08/04 11/08/04 11/08/04	arch     evidence qua
pub_comment: > auto comment: GENE_ONTOLOGY delete go id GO:0000166_add GO:0000824_add GO:00006139_add GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006519_AC GO:00006510_AC GO:00006510_AC GO:00006510_AC GO:00006510_AC GO:00006510_AC GO:00006510_AC GO:00006510_AC GO:00006510_AC GO:00000000_AC GO:00000000_AC GO:0000000_AC GO:00000_AC GO:000000_AC GO:00000_AC GO:00000_AC GO:00000_AC GO:0000_AC GO:0000_AC GO:0000_AC GO:0000_AC GO:0000_AC GO:0000_AC GO:0000_AC GO:0000_AC GO:000_AC GO:0000_AC GO:0000_AC GO:0000_AC GO:0000_AC GO:000_AC GO:000_AC GO:0000_AC GO:00	edit (F) m edit (F) c edit (P) m metai edit (P) p edit (P) a process	ucleotide binding atalytic activity ucleobase, nucleotide and n bolism mino acid and derivative metabolism <b>component</b> <b>reference</b>	subruit ueleic acid ueleic acid uuters uviti	reset   assigned by autoGO autoGO autoGO autoGO	go sug   se assign date 11/08/04 11/08/04 11/08/04	arch     evidenc qual
pub_comment:	edit (F) m edit (F) c edit (P) m metail edit (P) m edit (P) an process	ucleotide binding atalytic activity ucleobase, nucleoside, nucleotide and n bolism mino acid and derivative metabolism <b>component</b> <b>component</b>	submit ucleic acid ucleic acid	reset   assigned by autoGO autoGO autoGO autoGO	go sug   se assign date 11/08/04 11/08/04 11/08/04	arch     evidenc
pub_comment:	edit (F) m edit (F) c edit (P) m edit (P) m edit (P) m edit (P) a process	ucleotide binding atalytic activity ucleobase, nucleoside, nucleotide and n bolism mino acid and derivative metabolism <b>component</b> <b>reference</b>	submit ucleic acid others vitil	reset   assigned by autoGO autoGO autoGO autoGO	go sug   se assign date 11/08/04 11/08/04 11/08/04 11/08/04	arch     evidene qual
pub_comment: > auto comment: GENE_ONTOLOGY delete go id GO:3000166 add GO:3000166 add GO:30006139 add GO:3000612 add GO:3000612 add GO:30006519 add GO:3000651	edit (F) m edit (F) c edit (P) m edit (P) m edit (P) m edit (P) a process	ucieotide binding atalytic activity ucieobase, nucieoside, nucieotide and n bolism mino acid and derivative metabolism component component reference	submit ucleic acid others vitil	reset 1 sssigned by autoGO autoGO autoGO autoGO	go sug seign date 11/08/04 11/08/04 11/08/04 11/08/04	arch
pub_comment:         > auto comment:         GENE_ONTOLOGY         delete       go id         GO:3000166       add ]         GO:3000166       add ]         GO:3000166       add ]         GO:3000166       add ]         GO:30006519       add ]         GO:30006519       add ]         function       ISS ]         ISS ]       ISS ]         ISS ]       ISS ]	edit (F) m edit (F) c edit (P) pr edit (P) pr edit (P) a process	ucicotide binding atalytic activity ucicobase, nucicotide and n bolism rotein biosynthesis mino acti and derivative metabolism component component v reference v v v	submit ucleic acid ucleic acid	reset   assigned by autoGO autoGO autoGO autoGO	go sug 1 se assign 11/08/04 11/08/04 11/08/04 11/08/04	arch 1
pub_comment:         > auto comment:         GENE_ONTOLOGY         delete       go id         GO:3000166       add ]         GO:3000166       add ]         GO:3000166       add ]         GO:3000166       add ]         GO:300016519       add ]         GO:30006519       add ]         GO:30006519       add ]         ISS I	edit (F) m edit (F) c edit (P) m edit (P) m edit (P) m process	ucicotide binding atalytic activity ucicobase, nucicoside, nucicotide and n bolism nino acid and derivative metabolism <b>component</b> <b>reference</b> <b>v</b> <b>v</b> <b>v</b> <b>v</b> <b>v</b> <b>v</b> <b>v</b> <b>v</b> <b>v</b> <b>v</b>	submit a cleic acid a cleic acid acid acid acid acid acid acid acid	reset   assigned by autoGO autoGO autoGO autoGO	go sug seign date 11/08/04 11/08/04 11/08/04 11/08/04	arch     evidence qual
pub_comment: > auto comment: GENE_ONTOLOGY delete go id GO:0000824 add GO:000824 add GO:0006519 add GO:	edit (F) m edit (F) c edit (P) p edit (P) p edit (P) a process	ucleotide binding atalytic activity ucleobase, nucleoside, nucleoside and n bolism rotein biosynthesis mino acid and derivative metabolism component yeference y y y y y y y y y y y y y y y y y y y	submit ueleic acid others vitil viti	reset   assigned by autoGO autoGO autoGO autoGO	go sug   se assign date 11,08,04 11,08,04 11,08,04 11,08,04 11,08,04	eviden. qual
pub_comment: > auto comment GENE_ONTOLOGY delete go id GO:0000824 add GO:000824 add GO:000812 add GO:0006519 add GO:00006519 add GO:0006519 add GO:0006519 add GO:0	edit (F) m edit (F) c edit metal edit (P) a edit (P) a process	ucleotide binding atalytic activity ucleobese, nucleoside, nucleoside and n bolism mino acid end derivetive metabolism <b>component</b> <b>reference</b> <b>v</b> <b>v</b> <b>reference</b>	submit ucleic acid i ucleic ac	reset   assigned by autoGO autoGO autoGO autoGO	go sug   se assign date 11/08/04 11/08/04 11/08/04 11/08/04	arch   eviden qua

# Use all possible resources...

end5/end3: 15409 / 140 gene length: 4558 protein length: 660 mol. w1.: 50055.59 pf: 5.20

New Gene

----

S sug tree Ø





S N	CBI	••••••	••••	••••••	Protei	n
PubMed	Nucleotide	Protein	Genome Struct	ure PapSet	Taxonomy C	MIM Boc
Search	Protein 🗆	for I			Go Cle	ar
		Limits	Preview/Index	History	Clipboard	Details
Display	default 🗆	Save Text	Add to Clipboard			
□1: AAC50	1613. cytosolia	: NADP(+)[g	gi:1465733] <sup>Nucleotic</sup>	le, Related Seq	uences, PubMed, BLin	Taxonomy, k, LinkOut
LOCUS DEFINITION ACCESSION PID VERSION DESOURCE KEYWORDS SOURCE ORGANISH	AACS0613 cytosolic N AACS0613 g1465733 AACS0613.1 locus HSU43' humon. Humon.	WDP(+)-depen GI:1465733 344 accession 844 accession 844 accession	565 aa dent malic enzyme. n <u>U43944.1</u> rdata; Craniata; Ye	lineor PRI	27-JUL-1996 eleostowi;	
REFERENCE AUTHORS TITLE JOURIVAL MEDLINE	Nonidentity enzyme to to J. Protein / 96397682	offering Prin s 1 to 565) Huang S.M. ar of the cDNA hat from the them. 15 (3),	nd Chang,G.G. sequence of human narmal human cell , 273–279 (1996)	breast concer	cell malic	
REFERENCE AUTHORS TITLE JOURNAL	2 (residue Chou, HY., Direct Submitted ( Submitted ( Medical Cerr China	s 1 to 565) Huang,SM. Ission 27-DEC-1995) ber, 18 Shih	and Chang.GG. HY. Chau, Bloche -Yuan St., Taipei,	mistry, Natio Taiwan 100, Re	nal Defense epublic of	
COMMENT FEATURES SOURCE	Wethod: con Lo 1. /a /d	:eptual tran :otion/Quali 565 rganism="Homo train="Mango o_xref="taxo	slation. fiers o sapiens" lian (Chinese)" n:9606"			

szielyszes zincerziszter szielyszes zincerziszter rincerziszter sofiel by AC. sofiel by AC. For all matching prevológo7 / PROCOS7: Liporgerase PROCE246 / PROCAS6: Plant Liporgerase Family 500: Paralogous domain PPOL477 Family 500: Paralogous domain Process 3 V D gene, 1663.6 ron -119.20 Type [] Fan Dutes [] 199 / +10-08 17:07:25.0 (preded) E-pect 0 Coords 171-043 1-678 / 678 Score 1651.5 Centains (2) PROBASE Tyronine protein kinase PROBASE Sector Protein kinase gene\_syn: none Noise cutoff 23.20 role\_id: mone Total expect 4.0e-36 core 133.2 Coords 54-159 HMM Coords 1-1297129 Score 133.2 E-pect 4.8e - 36 IVER TO C - GO 3 Sinctural Inks [] Coords Recol Non 0.000 0.000 0 5 8 Mew Disches Mew Disches Searches %5000 04.0 85.3 03.4 83.2 03.2 82.8 01.9 Dem Pub Med Edit View Go Book lert 🗆 Send 18 Text 🗆 PLAInature Query- 11711.m00041 (857 letters) the plant journal PNA **Biological Chemistry** 0 geneti

Eve Set View Ge Boshamata Jook Window Mghi

 Vit711 B00004 /11711 B00004
 Vit711 B00004 /11711 /11711 B00004 /11711 /11111 /1111 /1111 /1111 /1111 /1111 /1111 /1111 /11111 /1111 /1111 /1

20 1 50 sect 20 1

Real Set Set Set

EMBL-EBI

# **Example:**-

A protein sequence from *Trypanosoma brucei*. Our task will be to annotate this protein sequence as fully as possible, given the tools at hand.

protein sequence:

>unknown\_T. brucei protein\_sequence MLRRLGVRHFRRTPLLFVGGDGSIFERY TEIDNSNERRINALKGCGMFEDEWIATE KVHGANFGIYSIEGEKMIRYAKRSGIMP PNEHFFGYHILIPELQRYITSIREMLCEK QKKKLHVVLINGELFGGKYDHPSVPKT RKTVMVAGKPRTISAVQTDSFPQYSPDL HFYAFDIKYKETEDGDYTTLVYDEAIEL FQRVPGLLYARAVIRGPMSKVAAFDVE RFVTTIPPLVGMGNYPLTGNWAEGLVV KHSRLGMAGFDPKGPTVLKFKCTAFQE ISTDRAQGPRVDEMRNVRRDSINRAGVQ LPDLESIVQDPIQLEASKLLLNHVCENRL KNVLSKIGTEPFEKEEMTPDQLATLLAK DVLKDFLKDTEPSIVNIPVLIRKDLTRYV IFESRRLVCSQWKDILKRQSPDFSE\*

# **Verify the gene structure**





**TIGR Rice Workshop** 

# **NCBI BLAST**

NCBI BLAST tools at: <u>http://www.ncbi.nlm.nih.gov/blast/</u>.

Program	Database	Query	
BLASTN	Nucleotide	Nucleotide	
BLASTP	Protein	Protein	
BLASTX	Protein	Nucleotide $\rightarrow$ Protein	
TBLASTN	Nucleotide →Protein	Protein	
TBLASTX	Nucleotide →Protein	Nucleotide $\rightarrow$ Protein	

TIGR Rice Workshop

S NCBI		protein-prot	ein <b>BLAST</b>
Nucleotide	Protein	Translations	Retrieve results for an
			RID
<b></b>			
<u>Search</u>			
Set subsequence From	m: To:	]	
Choose database nr	•		
Do CD-Search			
<u></u>		_	
Now: 🕒	BLAST! Or Reset query Res	et all	
Ontions	for edward blesting		
Options			
Limit by entrez query	or select f	rom:  All organisms	•
Compositional adjustments	Composition-based statistics	•	
Choose filter	Low complexity  Mask for	lookup table only 🗖 I	Mask lower case
Firment	10		
Expect			Verify Structural Annotation
Word Size	3 -		BLAST
Matrix	BLOSUM62 💽 Gap Costs Exis	stence: 11 Extension: 1	Domains
			Motifs
PSSM			Protein families
1			Functional Assignment

### **BLAST: What makes a good alignment?**

It depends on what you are trying to prove!

- minimum of 30% identity, better 40% & up
  - higher for short proteins
  - score is weighted for length
- full length match
  - at least 70% of both proteins

		Alignments		
Example : run NCBI BLAST		>gi 115504417 ref XP 001219001.1  G RNA editing ligase; RNA-editing complex protein; KREL2 ['		
		gi 83642483 emb CAJ16514.1  G RNA editing ligase; RNA-editi	ng complex protein; KREL2 [Trypa]	
		brucei] Le <sup>*</sup> h=416	_	
	natain against protain	Score = 860 bits (2222). Expect = 0.0. Method: Composition	-based stats.	
BLASIP-F	protein against protein	Identities = 416/416 (100%), Positives = 416/416 (100%), Gap	s = 0/416 (0%)	
		Query 1 MLRRLGVRHFRRTPLLFVGGDGSIFERYTEIDNSNERRINALKGCGMFE	DEWIATEKVHG 60	
Results <sup>.</sup>		Sbjet 1 MLRRLGVRHFRRTPLLFVGGDGSIFERYTEIDNSNERRINALKGCGMFE	DEWIATEKVHG 60	
recourto.		ery 61 ANFGIYSIEGEKMIRYAKRSGIMPPNEHFFGYHILIPELQRYITSIREM	LCEKQKKKLHV 120	
	BLASTP output a 100%	ANFGIYSIEGEKMIRYAKRSGIMPPNEHFFGYHILIPELQRYITSIREM	LCEKQKKKLHV LCEKQKKKLHV 120	
AUTHORS	$_{\rm H}$ match, is to a genom	erd N.J. Herris B.D. Hertz-Fowler C	AFDIKYKET 180	
A0111OK5	B project submissior	d.C.S., Atkin.R.J., Barron.A.J.,	YAFDIKYKET YAFDIKYKET 180	
	<sub>B</sub> which means that th	e Bringaud, F., Clark, L.N., Corton, C.H.,	PLVGMGNVPI. 240	
	c entry is no	tt,J., Fraser,A., Gruter,E., Hall,S.,	PLVGMGNYPL	
	Harper,À.D., Kay,M.P., Le	ech,V., Mayes,R., Price,C., Quail,M.A.,	PLVGMGNYPL 240	
	Rabbinowitsch, E., Reitter	,C., Rutherford,K., Sasse,J., Sharp,S.,	IRNVRRDSIN 300 IRNVRRDSIN	
	Shownkeen, R., MacLeod, A.,	Taylor,S., Tweedie,A., Turner,C.M.,	IRNVRRDSIN 300	
	Tait, A., Gull, K., Barrell	, B. and Melville, S.E.	ITPDQLATLL 360	
IIILE	The DNA sequence of Chrom	isotion recembination and nelumerphism	ITPDQLATLL ITPDQLATLL 360	
JOURNAL	Nucleic Acide Des 31 (16	A864_4873 (2003)		
PIIBMED	12907729	), 1001-1073 (2003)	SPDFSE 416 SPDFSE	
REFERENCE	2		SPDFSE 416	
AUTHORS	Berriman.M., Hertz-Fowler	.C.V.A., Hall.N., Kerhornou.A.X.,		
	Bowman, S., Quail, M., Kay,	M.P., Bray-Allen,S., Lennard,N.J.,	nitochondrial precursor (RNA li( ≥i]	
	Clark,L.N., Harris,B.R.,	Melville,S., Gerrard,C., Rajandream,M.A.		
	and Barrell, B.G.		based stats.	
TITLE	Direct Submission		0/416 (0%)	
JOURNAL	Submitted (20-SEP-2002) T	he Wellcome Trust Sanger Institute,	EWIATEKVHG 60	
	Wellcome Trust Genome Cam	pus, Hinxton, Cambridge CB10 1SA, UK	WIATEKVHG 60	
REMARK	revised by [3]		FEROKKELHY 120	
REFERENCE	3 (residues 1 to 416)		CEKQKKKLHV	
AUTHORS	Hertz-Fowler, C. and Berri	man,M.	CEKQKKKLHV 120	
TITLE	Direct Submission		YAFDIKYKET 180 💌	

Example :
navigating
<b>BLAST output</b>

>gi|47117107Lsp|P82864|TB48 TRYBB RNA editing ligase TbMP48, mitochondrial r RNA ligase MP48 [Trypanosoma brucei] gi|110670?>(gb|AAG27063.1) Length=2.6 Score = 856 bits (2212), Expect = 0.0, Method: Composition-based stats.

Identities

The se	cond	hit	in	the
BLAST	outp	out,	a g	9%
match,	is t	o a	Sw	iss-
Prot ent	trv.			

The alignment reveals three	
positions with variations:	

I103V (very similar, both hydrophobic) conservative

```
D182G (negative, hydrophilic
   to tiny polar) non-
   conservative
```

```
V364A (nonpolar, aliphatic,
    hydrophobic to tiny,
    nonpolar, aliphatic)
    conservative
```

3 =	413/41	.6 (99%), Positives = 414/416 (99%), Gaps = 0/416 (	0%)
Quer	y 1	MLRRLGVRHFRRTPLLFVGGDGSIFERYTEIDNSNERRINALKGCGMFEDEWIATEKVHG MLRRLGVRHFRRTPLLFVGGDGSIFERYTEIDNSNERRINALKGCGMFEDEWIATEKVHG	60
Sbjo	t 1	MLRRLGVRHFRRTPLLFVGGDGSIFERYTEIDNSNERRINALKGCGMFEDEWIATEKVHG	60
Quer	y 61	ANFGIYSIEGEKMIRYAKRSGIMPPNEHFFGYHILIPELORY <mark>ITSIREMLCEKOKKKLHV</mark> ANFGIYSIEGEKMIRYAKRSGIMPPNEHFFGYHILIPELORY+TSIREMLCEKOKKKLHV	120
Sbjo	t 61	ANFGIYSIEGEKMIRYAKRSGIMPPNEHFFGYHILIPELQRYVISIREMLCEKQKKKLHV	120
Quer	y 121	VLINGELFGGKYDHPSVPKTRKTVMVAGKPRTISAVQTDSFPQYSPDLHFYAFDIKYKET	180
Sbjo	t 121	VLINGELFGGKYDHPSVPKTRKTVMVAGKPRTISAVQTDSFPQYSPDLHFYAFDIKYKET	180
Quer	y 181	ED GDYTTLVYDEAIELFQRVPGLLYARAVIRGPMSKVAAFDVERFVTTIPPLVGMGNYPL E GDYTTLVYDEAIELFQRVPGLLYARAVIRGPMSKVAAFDVERFVTTIPPLVGMGNYPL	240
Sbjo	t 181	EG GDYTTLVYDEAIELFQRVPGLLYARAVIRGPMSKVAAFDVERFVTTIPPLVGMGNYPL	240
Quer	y 241	TGNWAEGLVVKHSRLGMAGFDPKGPTVLKFKCTAFQEISTDRAQGPRVDEMRNVRRDSIN TGNWAEGLVVKHSRLGMAGFDPKGPTVLKFKCTAFOFISTDRAOGPRVDEMRNVRRDSIN	300
Sbjo	t 241	TGNWAEGLVVKHSRLGMAGFDPKGPTVLKFKCTAFQEISTDRAQGPRVDEMRNVRRDSIN	300
Quer	y 301	RAGVQLPDLESIVQDPIQLEASKLLLNHVCENRLKNVLSKIGTEPFEKEEMTPDQLATLL	360
Sbjc	t 301	RAGVQLPDLESIVQDPIQLEASKLLLNHVCENRLKNVLSKIGTEPFEKEEMTPDQLATLL RAGVQLPDLESIVQDPIQLEASKLLLNHVCENRLKNVLSKIGTEPFEKEEMTPDQLATLL	360
Quer	y 361	AKDVLKDFLKDTEPSIVNIPVLIRKDLTRYVIFESRRLVCSQWKDILKRQSPDFSE 416 AKD LKDFLKDTEPSIVNIPVLIRKDLTRYVIFESRRLVCSOWKDILKROSPDFSE	;
Sbjo	t 361	AKDALKDFLKDTEPSIVNIPVLIRKDLTRYVIFESRRLVCSQWKDILKRQSPDFSE 416	i

See Glossary entry for SNP

# **Swiss-Prot**

Our sequence is 99% identical to the sequence of this Swiss-Prot entry.

Another name for this protein in the literature is 'REL2.'

ENTRY INFORMATION	
ENTRY NAME	TB48 TRYBB
ACCESSION NUMBER	P82864
Integrated into Swiss-Prot on	2004-05-10
Sequence was last modified on	2001-03-01 (Sequence version 1)
Annotations were last modified on	2006-10-31 (Entry version 24)
NAME AND ORIGIN OF TH	E PROTEIN
PROTEIN NAME	RNA-editing ligase TbMP48, mitochondrial precursor
Synonyms	EC 6.5.1.3 RNA ligase
GENE NAME	MP48
SOURCE ORGANISM	Trypanosoma brucei brucei
TAXONOMY ID	5702 [NCBI, NEWT]
LINEAGE	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma
REFERENCES	
[1]	<ul> <li>Panigrahi AK; Gygi SP; Ernst NL; Igo RP Jr, Palazzo SS; Schnaufer A et al. <u>View all</u>.</li> <li>Association of two novel proteins TbMP52 and TbMP48 with the Trypanosoma brucei RNA editing complex.</li> <li>2001, Mol. Cell. Biol., 21, 380-389.</li> <li>Position: NUCLEOTIDE SEQUENCE [GENOMIC DNA], PROTEIN SEQUENCE OF 18-37; 58-72; 118-139; 143-151; 200-207; 217-224; 255-263; 302-323; 336-340; 371-384 AND 410-416, FUNCTION, AND SUBCELLULAR LOCATION</li> <li>PubMed: <u>11134327</u>; Medline: <u>20576857</u>.</li> </ul>
COMMENTS	
FUNCTION	Part of the RNA editing complex essential for cell variability. RNA editing in kinetoplastid mitochondria inserts and deletes uridylates at multiple sites in pre-mRNAs as directed by guide RNAs.
CATALYTIC ACTIVITY	ATP + (ribonucleotide)(n) + (ribonucleotide)(m) = AMP + diphosphate + (ribonucleotide)(n+m).
SUBCELLULAR LOCATION	Mitochondrion.

# **Swiss-Prot**

#### Click on the NCBI hyperlink to look at this publication.

**TIGR Rice Workshop** 

ENTRY INFORMATION			
ENTRY NAME	TB48 TRYBB		
ACCESSION NUMBER	P82864		
Integrated into Swiss-Prot on	2004-05-10		
Sequence was last modified on	2001-03-01 (Sequence version 1)		
Annotations were last modified on	2006-10-31 (Entry version 24)		
NAME AND ORIGIN OF TH	E PROTEIN		
PROTEIN NAME	RNA-editing ligase TbMP48, mitochondrial precursor		
Synonyms	EC 6.5.1.3 RNA ligase		
GENE NAME	MP48		
SOURCE ORGANISM	Trypanosoma brucei brucei		
TAXONOMY ID	5702 [ <u>NCBI</u> , <u>NEWT</u> ]		
LINEAGE	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma		
REFERENCES			
	<ul> <li>Panigrahi AK; Gygi SP; Ernst NL; Igo RP Jr; Palazzo SS; Schnaufer A et al. <u>View all</u>.</li> <li>Association of two novel proteins TbMP52 and TbMP48 with the Trypanosoma brucei RNA editing complex.</li> <li>2001, Mol. Cell. Biol., 21, 380-389.</li> <li>Position: NUCLEOTIDE SEQUENCE [GENOMIC DNA], PROTEIN SEQUENCE OF 18-37; 58-72; 118-139; 143-151; 200-207; 217-224; 255-263; 302-323; 336-340; 371-384 AND 410-416, FUNCTION, AND SUBCELLULAR LOCATION</li> </ul>		
	PubMed: <u>11134327;</u> Medline: <u>20576857</u> .		
COMMENTS	PubMed: <u>11134327;</u> Medline: <u>20576857</u> .		
COMMENTS FUNCTION	PubMed: <u>11134327;</u> Medline: <u>20576857</u> . Part of the RNA editing complex essential for cell variability. RNA editing in kinetoplastid mitochondria inserts and deletes uridylates at multiple sites in pre-mRNAs as directed by guide RNAs.		
COMMENTS FUNCTION CATALYTIC ACTIVITY	PubMed: <u>11134327</u> ; Medline: <u>20576857</u> . Part of the RNA editing complex essential for cell variability. RNA editing in kinetoplastid mitochondria inserts and deletes uridylates at multiple sites in pre-mRNAs as directed by guide RNAs. ATP + (ribonucleotide)(n) + (ribonucleotide)(m) = AMP + diphosphate + (ribonucleotide)(n+m).		



# Pubmed

- Read abstract
- If promising, read paper to be sure protein is characterized
- If characterized, it is good <u>evidence</u> for naming our sequence

Association of two novel proteins, TbMP52 and TbMP48, with the Trypanosoma brucei RNA editing complex.

<u>Panigrahi AK, Gygi SP, Ernst NL, Igo RP Jr, Palazzo</u> <u>SS, Schnaufer A, Weston DS, Carmean N, Salavati</u> <u>R, Aebersold R, Stuart KD</u>.

Seattle Biomedical Research Institute, Seattle, Washington 98109, USA.

RNA editing in kinetoplastid mitochondria inserts and deletes uridylates at multiple sites in pre-mRNAs as directed by guide RNAs. This occurs by a series of steps that are catalyzed by endoribonuclease, 3'-terminal uridylyl transferase, 3'-exouridylylase, and RNA ligase activities. A multiprotein complex that contains these activities and catalyzes deletion editing in vitro was enriched from Trypanosoma brucei mitochondria by sequential ion-exchange and gel filtration chromatography, followed by glycerol gradient sedimentation. The complex size is approximately 1,600 kDa, and the purified fraction contains 20 major polypeptides. A monoclonal antibody that was generated against the enriched complex reacts with an approximately 49-kDa protein and specifically immunoprecipitates in vitro deletion RNA editing activity. The protein recognized by the antibody was identified by mass spectrometry, and the corresponding gene, designated TbMP52, was cloned. Recombinant TbMP52 reacts with the monoclonal antibody. Another novel protein, TbMP48, which is similar to TbMP52, and its gene were also identified in the enriched complex. These results suggest that TbMP52 and TbMP48 are components of the RNA editing complex.

PMID: 11134327 [PubMed - indexed for MEDLINE]

# **Domains (HMMs)** TIGRFAMs search



There are no Pfam-A domains higher than gathering threshold

Help

[416 residues]

About

Total score:	923.1
Trusted cutoff:	100.0
Gathering cutoff:	100.0
Noise cutoff:	-165.0



This is a very positive hit to the RNA ligase RNL2 family domain (TIGR02307).

hmmpfam - search a HMMER 2.1.1 (Dec 1 Copyright (C) 1992 HMMER is freely di	single seq against HMM database 998) -1998 Washington University School of stributed under the GNU General Publi	Medicine Ic License	(GPL).	
HMM file: Sequence file:	ALL_LIB_bin.HMM hmmpfam-search-14395-11722551	.72.in		
Query: unknown_T.	brucei protein_sequence			
Scores for sequence Model Descript	e family classification (score includ ion 	les all doma Score l	ains): E-value	N
TIGRO2307 RNA_lig_	 RNL2: RNA ligase, Rnl2 family	923.1 7	.7e-274	1
Parsed for domains Model Domain	: seq-f seq-t hmm-f hmm-t score	e E-value		
TIGR02307 1/1	25 408 1 421 [] 923.1	7.7e-274		
Alignments of top-: TIGR02307: domain	scoring domains: 1 of 1, from 25 to 408: score 923.1, *->FkkYTsleNssyrrifaeKltglglrGGEWVA F++YT+++Ns++rri+a+K++g++++ EW+A+	E = 7.7e-2' EKiHGaNFSi: EK+HGaNF+i-	74 ivee +++e	
unknown_T. 25	FERYTEIDNSNERRINALKGCGMFEDEWIAT	TEKVHGANFGI	YSIE 69	
unknown_T. 70	dPNEAqDGAEkkVtfAKRtGiidPnEdGDYDFFGY) + Ek++++AKR+Gi++PnE+ FFGY) GEKMIRYAKRSGIMPPNEHFFGY)	nilieeytakv) nili+e+++++ HLLIPELQRYI'	kAis ++i+ ISIR 107	,
	dlLkekaGvikklesvivyGELaGkgyqkpvvPKsr	KtvtlanKkR:	iISG	
unknown_T. 108	EMLCEKQKKKLHVVLINGELFGGKYDHPSVPKT	KTVMVAGKPR	ΓΙS- 154	F
unknown_T. 155	vevQsdsFPQYsPDkdFyAFDIkyketGeeeddvtI +vQ+dsFPQYsPD++FyAFDIkyket e++d++tI -AVQTDSFPQYSPDLHFYAFDIKYKET-EDGDYTTI	.vyDevlEvfe .vyDe++E+f+ .VYDEAIELFQ	cvpk cvp+ RVPG 202	:
	lkyåkelvRGtldEllafDNDLDSVVqvenFvtdlH l+yå++++RG+++++afD ve+Fvt++F	aLVdlgnypLl +LV++gnypL	NAEA	
unknown_1. 203	LLIAKAVIRGPMSKVAAFDVERFVITIF	PLACENCE ALP-	240	,

# Verify HMM

Total score:923.1Trusted cutoff:100.0Gathering cutoff:100.0Noise cutoff:-165.0

Score is well above the trusted cutoff.

	► TIGR02307: RNA ligase, I	Rnl2 family (View Sanger Pfar	m)		gene_sym: 1	ione ec	#: 6.5.1.3	role_id: <b>152, 166</b>
.1	Isology: equivalog							
$\cap$	Total score: 923.1	Trusted cutoff: 100.00	Gathering	cutoff: <b>100.00</b>	Nois	e cutoff: - <b>165.0</b>	0	Total expect: 1.3e-274
		Trusted cutoff2: -165.00	Gatheri	ng cutoff2:	Noise	e cutoff2: <b>100.0</b>	10	
.0								
$\cap$	View Alignment	Coords	HMM Coords	Score	Expect	Curation	[A	.dd To GO Evidence]
	▶ align page	670354-671505	1-421 / 421	923.1	1.3e-274			
1e								
	▶ No HMM-GO Sugge:	stions To Display.						

X Belvu: /home/lhannick/belfile.tmp		×
File Edit Colour Sort Pic	sked:	
(5×422)	60708090100120130140150160170	1
338.m00295 1 384 SPIP82864ITB48_TRYBB 25 408 PIRIG81461IT02843 76 468 GBIAAQ64204.1134333049 3 332 SPIP32277IY10A_BPT4 2 334	<pre>EKMIRYAKRSGIMPPNEHFFGYHILIPELQRYITSIREMLCEKQKKKLHVVLINGELFGGKYDHPSVPKTRKTVMVAGKPRT.ISAVQTDSFPQYSPDLHFYAFDIKYKET-EI EKMIRYAKRSGIMPPNEHFFGYHILIPELQRYVTSIREMLCEKQKKKLHVVLINGELFGGKYDHPSVPKTRKTVMVAGKPRT.ISAVQTDSFPQYSPDLHFYAFDIKYKET-EI EKMIRYAKRSGIMPPNEHFFGYHILIPELQRYVTSIREMLCEKQKKKLHVVLINGELFGGKYDHPSVPKTRKTVMVAGKPRT.ISAVQTDSFPQYSPDLHFYAFDIKYKET-EI EKMIRYAKRSGIMPPNEHFFGYHILIPELQRYVTSIREMLCEKQKKKLHVVLINGELFGGKYDHPSVPKTRKTVMVAGKPRT.ISAVQTDSFPQYSPDLHFYAFDIKYKET-EI EKMIRYAKRSGIMPPNEHFFGYHLLIPELQRYVTSIREMLCEKQKKKLHVVLINGELFGGKYDHPSVPKTRKTVMVAGKPRT.ISAVQTDSFPQYSPDLHFYAFDIKYKET-EI EKMIRYAKRSGIMPPSENFFGYHLLIDDFTAQVRALCALLKRKYGVTGRMGRVVLHGELFGAKYKHPLVPKSTKWCTLPNKKRIPIGVEIQSEPFPQYSPELHYFAFDVKYSVGAN EFTVTPAKRTSTIGANVMGDYDFYGCTSVVEAHTAKMEAISNWLWARG-IINVGETIIVYGELAGKGVQKEVN</pre>	

Our sequence contains an RNA ligase, Rnl2 family domain, with a very strong match. Members of this TIGRfam family ligate RNA.



#### SignalP-HMM result:

# **TargetP**

The sequence contains a mitochondrial targeting peptide, mTP.



cutoff

**TargetP 1.1 Server - prediction results** 

0.000 0.000

0.000

**Technical University of Denmark** 

Number of query sequences: 1 Cleavage site predictions not included. Using NON-PLANT networks. Len mTPSP other LdC RC Name 416 0.728 0.070 0.209 з unknown\_Tb\_seq



### **Transmembrane domains**

There are no transmembrane domains.

#### **TMHMM result**

**<u>HELP</u>** with output formats

#	unknown	Length: 416			
#	unknown	Number of pr	redicted TMHs:	0	
#	unknown	Exp number d	of AAs in TMHs:	0.0	0491
#	unknown	Exp number,	first 60 AAs:	0.0	0077
#	unknown	Total prob d	of N-in:	0.0	0474
ur	nknown TM	инмм2.О	outside	1	416



# **Annotation of Example Protein**

**BLAST:** A protein match at Swiss-Prot is 99% identical, with 2 conservative and one nonconservative amino acid substitutions. "RNA-editing ligase TbMP48, mitochondrial precursor" is the Swiss-Prot name for this close protein match.

This mitochondrial precursor of an RNA ligase was identified as a <u>member of a multi-protein complex that catalyzes deletion editing in vitro</u>. It was isolated from an enriched sample of Trypanosoma brucei mitochondria by sequential ion-exchange and gel filtration chromatography, followed by glycerol gradient sedimentation. The protein was not functionally characterized, but was identified as a member of an RNA-editing complex. The complex was shown to have RNA-editing function. (PMID:11134327)

**Domain:** Our sequence contains an RNA ligase, Rnl2 family <u>TIGRFAMs domain</u>, with a very strong match. Members of this TIGRfam family ligate (seal breaks in) RNA.

Signal sequence: none

Targeting Sequence: It contains a mitochondrial targeting sequence.

Under the standards of this annotation project, "RNA-editing ligase TbMP48, mitochondrial precursor," is a suitable name.



### **Evidence from homology searching**

#### **Compare sequences of unknown function to those of known function.**

Shared sequence identity may imply shared function:-

- Full-length match with significant identity (>30%)
- Domains and motifs
- Binding sites
- Catalytic sites

#### But :

- there are occurrences where one amino acid substitution changes the function of an enzyme.
- synonymous or "silent" codon substitutions may result in functional differences.
- Mutations may result in modification or deletion of function.
- all functional assignments made by similarity should be considered tentative until confirmed by experiment.

### **Transitive annotation**

Beware! A is like B B is like C C is like D D is NOT like A!

Take a conservative approach. Err on the side of missing homology rather than stretching weak data.

# **Gene Ontology**

### GO is...

- a <u>method</u> used to structure biological knowledge using a dynamic controlled vocabulary across organisms.
- a <u>database</u> containing a shared vocabulary of descriptive terms for the description of the molecular function, biological process and cellular component of gene products.
- ◆ The Gene Ontology Consortium<sup>™</sup> is a <u>collaboration</u> among model genome organism databases.

# Topics

- Reasons GO has been developed
- Nuts and bolts of GO
- Tools
- Searching GO
- Assigning terms
- GO Slims

# **The Basics**

- GO is a controlled vocabulary
- GO has three aspects, or ontologies:
  - Molecular function
  - -Biological process
  - Cellular component
- The 3 aspects refer to genes and gene products

# The specificity of GO

There is a limit to how much information can be contained in the name of a protein. For example:"translation initiation factor 2 subunit"

GO terms assigned to this tell much more:

- GO:0005525 (MF) GTP binding
- GO:0006413 (BP) translational initiation
- GO:0005851 (CC) eukaryotic translation initiation factor 2B complex

The Gene Ontology is like a dictionary



### Each concept has:

- a name
- a definition
- an ID number

term: transcription initiation

id: GO:0006352

definition: Processes involved in the assembly of the RNA polymerase complex at the promoter region of a DNA template resulting in the subsequent synthesis of RNA from that promoter.

# **GO** terms

- A GO term, or ID, is attached to every function, process or component
- There are relationships between them
- Relationships are shown by a graph
  - Directed acyclic graph
  - Sometimes called a "tree"

# **GO Tools**

# **GO tools** are available at the GO Consortium: http://www.geneontology.org/GO.tools.shtml

<u>Developed and maintained by GO:</u> AmiGO - Searching through terms and annotations OBO-Edit - Editing and viewing the DAG

Many others developed independently, for: Annotation Gene expression/microarray data GO Slims

# AmiGO

# The GO Browser





Terms of use • Disclaimer • GO helpdesk Copyright © 1999-2006 the Gene Ontology Consortium



#### 🗆 all : all [184843] 🔶

Cellular Component

Molecular Function

#### GO:0008150 : biological\_process [139437] •

dictyBase

FlyBase

- GO:0065007 : biological regulation [18316]
- ⊞ 
   GO:0009987 : cellular process [81676]
- GO:0032502 : developmental process [16502]
- ⊕ GO:0043062 : extracellular structure organization and biogenesis [313]
- GO:0042592 : homeostatic process [1533]
- ⊞ 
   GO:0051179: localization [20043]
- ⊞ 
   GO:0040011 : locomotion [458]
- GO:0051235 : maintenance of localization [200]
- ⊞ 
   GO:0008152 : metabolic process [53763]
- ⊞ 
   GO:0051704 : multi-organism process [1640]

 Graphical View Permalink Download as XML Download as flat file

Binder Bi	AmiGO
Search Advanced Search BLAST search Browse	Help
Filter results Filter by ontology Ontology All Biological Process Cellular Component Molecular Function Filter Gene Product Counts Data source All CGD dictyBase FlyBase FlyBa	
<ul> <li>all : all [184843] •</li> <li>GO:0008150 : biological_process [139437]</li> <li>GO:0022610 : biological adhesion [1691]</li> <li>GO:0065007 : biological regulation [18316]</li> <li>GO:0009987 : cellular process [81676]</li> <li>GO:0032502 : developmental process [16502] •</li> <li>GO:0009838 : abscission [6]</li> <li>GO:0007571 : age-dependent general metabolic decline [9]</li> <li>GO:0007568 : aging [435]</li> <li>GO:0048856 : anatomical structure development [10162]</li> <li>GO:0009653 : anatomical structure formation [860]</li> <li>GO:0009653 : anatomical structure morphogenesis [5807]</li> <li>GO:0016265 : death [2387]</li> </ul>	Graphical View Permalink Download as XML Download as flat file

### **GO** information to include

Independent of interface, add: GO ID Evidence code Reference

The date is an important part of the annotation .

In Manatee:

Qualifier

	functio	n	process	compone	ent oth	iers	
		-	•			-	
	add go id	ev code	ref	erence		with	qualifier
		ISS 💌		<b>-</b>			
		ISS 🔽	] [	•		•	
		ISS 🖵	]	•			
		ISS 👤		•		•	
2		ISS 💌		•		•	

# **Filling in the GO information**

function	process	component	others		
function	PMID: TIGR_Tba1:annotatio	component unknown component cytoplasm extracellular integral to membrane membrane nucleus mitochondrion	others TmHMM signalP targetP CBI_gi: GR_Tba1: rterPro: JniProt/Swiss-Prot: Pfam: PIR: JniProt/TrEMBL: SGD: SGDID: TAIR:gene: -B: Prosite:	with	qualifier
		4 ) M	orotein_id: GenProtEC: MGI:		
		F	RGD: JniProt: GDB:		

# Assigning a GO term

- 1) Read the literature, not just the abstract
- 2) Search for GO terms
- 3) Record the data

# **GO** Annotations based on similarity

- Sequence or structure
  - Similarity to GO-annotated gene products
- Domains
- EC numbers
- Pathways
- Protein families

and many more...

http://www.geneontology.org/GO.indices.shtml

# **Annotating by similarity**

use the evidence code 'ISS'—inferred from sequence or structural similarity.

enter the database ID of the entity used to infer similarity in the 'With' field.

### **IEA: Inferred from Electronic Annotation**

IEA is used when no curator has checked the annotation to verify its accuracy.

#### Use when an annotation:

- is based on "hits" in sequence similarity searches, if they have not been reviewed by curators
- is transferred from database records, if not reviewed by curators
- that depend directly on computation or automated transfer of annotations from a database.
  - The actual method used (BLAST search, SwissProt keyword mapping, etc.) doesn't matter.
  - If the method is match-based, a valid database ID *must* be entered in the with column.

# **GO Slim**

- cut-down versions of the ontologies
- useful summary of GO annotation
- versions of GO Slims available
  - Eukaryotic GO slim
  - Plant GO slim
  - Yeast GO slim

# **Points to remember**

- GO enables querying across annotations
- The GO Consortium website has documentation and lists available tools
- AmiGO is available online and as downloadable resource
- GO Slims summarize your annotation
- GO annotations are worth the trouble—they enhance the value of research

# MANATEE-

- Navigation, inspection & <u>curation</u> of gene products
  - Gene/Gene products
  - GO Assignments
- Available at:
  - http://manatee.sourceforge.net

60742.100037	ani ion				
00/42.0003/	101101180				
Model: 6074	(F14621.34) 2.m00252	STATUS	E NOT CURATED	database	jath1
Pub Locus /	#1g55020	end5/end3	3: 19409714052	6	
asmbi_id:	60742	gene lengt	h: 4558	feat_name.focus	I
		protein leng	gan: 060		
Select Fund	ation 🗆	mot. wt.	5 30044.39	New G	ene
Reload	Page.	- P*	9.2.9		
			1000 20	2000	4000
Gene Synonyms: Internet	Even UTR stor ture:	freeer	f~~ f	·····	
None	Contro or fit and contro.	_			
GENE IDENTIFICATION					SHÉ
Gene Name				Gene Symbol	EC Number
Dinovvaenase (1001)				Dicert	3. 13. 11. 12
wh commant-				Jeans,	
Relation to S0100623	71 concerted by d	NR: of 200202 of	104637 1 ATHI TO/	(Y )	
Transiences on selfdoore	, supported by to	Net gr_coscoc_g		" D	
				- IZ	
auto_comment					
GENE ONTOLOGY					S sus tree
delete GO ID				assigned by assign date	evidence
9				cronning 08/28/02	
- GOMONEYSZ and	edt (P) detense res	ponse			1AS: PMID:7506
GO:0009511 add	edit (P) response to	wounding		cronning 08/28/02	TAS: PMID:7506
EMIDENCE PROTURE			_		
0 100 200	300 400 4	500 600 700	0.000		8
		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~			
			signal	P (SP-H99) 020	
			- IPROOD	907 / PF00305: Lipoxygenase	
			IPR001 PS0021	024 / PF01477: Lipoxygenape 11 Lipoxygenapes iron-biodi	, region 2 ng region signat
			IPR000	907 / PS00081: Lipoxygenase	
			LPR000	907 / PR000871 Lipoxygenase 246 / PR00458: Plant linovo	genase
			Family	589: Paralogous domain PFO	1477
			- Family	509: Paralogous domain PFO	0305
ным					S V 6
CO0305: Linexupanasa		0606 510	. 0008	ACR: 0000 17	la id: none
Isology: PFAM equivalog	domain Tota'	score: 1653.5	Trusted cutoff: -119,2	Noise cutoff: -354.70	Total expect I
View Alignment	Coords	HMM Coords	Score E	xpect Curation (A	dd To GO Eviden
align page	171-843	1-678 / 678	1653.5	e =	
No HMM-GO Suggestion	a To Display.				
PF01477: PLAT/LH2 doma	in and a set of	gene,	sym: none	ec#: none r	ole_id: mone
Isology: domain	Total score: 133.2	Trusted cutoff	34.20 Norse	cutott: 23.20 Total e	opect 4.8e-36
View Alignment	Coords	HMM Coords	Score Eq	pect Curation (A	dd To GO Eviden
align page	54-159	1-129 / 129	133.2 4.86	-36 1	
No HMM-GO Suggestion	s To Display.				
No HMM-GO Suggestion	s To Display.				
No HMM-GO Suggestion	s To Display.				• 6
No HMM-GO Suggestion PROSITE	s To Display.	instan 9			S 🖬
No HMM-GO Suggestion PROSITE PS00001: Lipoxygenases	s To Display. iron-binding region si	ignature 2.			s 6
No HMM-GO Suggestion PROSITE PSOCOL: Upoxygenases Match sequence: LHPVFR	s To Display. iron-binding region si ILLEPH	ignature 2.			s G
No HMM-GO Suggestion PROSITE PS0001: Upoxygenases Match sequence: UHPVF8 Coords	s To Display, iron-binding region si LLEPH	ignature 2. Recall	Quartien		S 🖬
No HMM-GO Suggestion PROSITE PSUBBL: Upprygenases Match sequence: LHPVFP Coords 541/551	s To Display. iron-binding region si LLEPH Procision 1.00	ignature 2. Recall 0.97	Quation		s 🕯
No HMM-GO Suggestion PROSITE PSOD01: Liptorygenases Match sequence: LHPVFR Coerds 541/551 PS00711: Liptorygenases	s To Display. iron-binding region si LLEPH Procision 1.00 iron-binding region si	ignature 2. Recall 0.97 ignature 1.	Quration E	Add To GO Even	s @ ence]
No HHM-GO Suggestion PROSITE PS00011: Uptoxygenases Match sequence: LHPVFP Coords S41/551 PS00711: Uptoxygenases Match sequence: HOUS9	s To Display. iron-binding region si ILLEPH Procision 1.00 iron-binding region si mMOTHASIE	ignature 2. Recall 0.97 Ignature 1.	Quastion	jada To GO Evid	s E
No HMM-GO Suggestion PROSTE PSSOUT: Upprygenases Match sequence: UHPVFP Coords S41/551 PSS0711: Upprygenases Match sequence: HOLISP	s To Display, iron-binding region si LLEPH Procision 1.00 iron-binding region si min-binding region si	ignature 2. Rocall 0.57 ignature 1.	Quastion	JARK TO GO Even	s C
No HMM-GO Suggestion PRIOSITE PSIOSIT: Upoxygenases Match sequence: UHVYF Coords S41/551 PSIO/T1: Upoxygenases Match sequence: HOUSH Coords	s To Display. iron-binding region si LLEPH Procision 1.00 iron-binding region si nuMOTHASIE Precision	Ignature 2. Rocal 0.97 Ignature 1. Rocal	Quartien E Quartien		s E
No HMM-GO Suggestion PROSTE PS000011 Uporygenases Math sequence: LHPVP S41/551 PS002111 Uporygenases Math sequence: HOLISH Opents 514/528	s To Display. iron-binding region si CLEPH Precision 1.00 Procision 1.00	ignature 2. Recal 0.97 Ignature 1. Recal 0.94	Curation E Curation E	[465 To 50 Evid	encel
No HIMM- GO Suggestion PROSITE PSS0011: Upprygenases Match sequence: UHPVFP Opensis S41/551 PSS0711: Upprygenases Match sequence: HOLISP Opensis Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 Start520 St	s To Display. Inon-bindleg region si (LLEPH 100 Inon-bindleg region si metoThASSE Precision 1.00	Ignature 2. Recall 0.97 1. Recall 0.94	Curation E Curation	[Add To GO Eved	encel
Ne HMM-GO Suggestion Priositie Proceedia Lipeorygenases Match sequence: HIV/FR Salacts acquence: HOLISH Coords S14/520 S16/520 S16/64_P	s To Display. iron-binding region si LLEPH Procision 1:00 Iron-binding region si mo-binding region si Procision 1:00	ignature 2. Rocall 0.97 Ignature 1. Rocall 0.94	Ouration E Ouration E	Add To GO Evid	3 (2 ence) encel s (2
No HMM-GO Suggestion PROSTE PS00011 Upoxygenases Math sequence: LHPVP Set1/551 P300711: Upoxygenases Math sequence: HOLISH Coords 514/520 SIGMA_P SignalP-2.8 Results: [Gr SuggestP-2.8 Results: [Gr	s To Display. Iron-binding region si LLEPH Precision 1.00 Iron-binding region si memorrhading region si memorrhading region 1.00	Ignature Z. Rocal 0.97 Ignature 1, Rocal 0.94	Curation E Curation E	(Add To GO Evid	s d encel s d
No HMM-GO Suggestion PHOSITE PS000011 Uptorygenases Match sequence: UHPVFP Coords S41/551 PS00711: Uptorygenases Match sequence: H0USP Coords S14/520 SIGIML_P SignelP-2.0 Hestits: [Sr: SignelP-2.0 HMM	s To Display. Iron-binding region si LLEPH Procision 1:00 Procision 1:00 Procision	Ignature 2. Recall 0.97 Recall 0.94 contput. for: SP-HM	Curation E Curation E M/INI	[Add To GO Eved	s 6 ence] 5 6
Ne HMM-GO Suggestion PROSITE PS00011 Upoxypenases Match sequence: HIVVR Coords 541,651 PS00111 Upoxypenases Match sequences 514,650 SIGHAL_P SignalP-2,8 HMM Prediction	s To Display. Iron-binding region si CLEPH Procision 1:00 Iron-binding region si Iron-binding region si Iron-binding region si Iron-binding region si Iron-binding region si Iron-secretary Non-secretary	Recall 0.57 Ignature 1. Recall 0.54 conjust for SP-HM protein _ CO	Ouration E Ouration E M/RM] rated	(Add To GO Evid	ence) ence) s C
No HMM-GO Suggestion PROSITE PS00011 Uposygenases Match sequence: LHPVP S41/551 PS002111 Uposygenases Match sequence: HOLISH Coords S14/525 SIGNA_P SignalP-2.8 HMM Prediction	s To Display. Iron-binding region si LLEPH Iron binding region si Iron bin	Ignature Z. Rocal 0.97 Ignature 1, Rocal 0.94 codpart for 3P-HM	Ourstien E Ourstien E MARKI rated	[Add To DO Evid	s C encej encej s C
Ne HMM-GO Suggestion PRIOSITE Proceeding Lipeorygenases Match sequence: HOUSE Coords S11/551 PS9/211: Upeorygenases Match sequence: HOUSE Coords S14/520 SIGMA_P SignalP-2.8 Results: [Gr Signal Results: [Gr SignalP-2.8 Res	s To Display. iron-binding region s LLEPH Precision 1.00 Inon-binding region si netwoTMASE Precision 1.00 ubical Display] (Rear Non-secretary 0.000 0.000	Ignature 2. Recall 0.57 Recall 0.54 conjust for SP-HM iprotein  © Co	Oursteen E Ourstien E MARKI Irated	[Add To GO Eved [Add To GO Eved [Add To GO Eved	s C encel s C
Ne HMM-GO Suggestion PriOSITE Present: Unorygenases Match sequence: UNVVP Coords S41/551 Prese/11: Unorygenases Match sequence: HOUSP Coords S14/520 Signal-2,8 HMM Prediction Signal peptide probability Signal exchor probability Signal exchor probability Signal exchor probability	s To Display. Iron-binding region s LLEPH Iron-binding region s LLEPH Iron-binding region s Iron-binding regio	Recall 0.97 Ignature 1. Recall 0.94 output for 3P-HM protein _ Cu	Ouration E Ouration E M/1991 rated	(Add To GO Evid	s d ancaj s d
Ne HMM-GO Suggestion PROSITE PS00011: Upprygenases Match sequence: UHPVFB Coords S41/55 PS00711: Upprygenases Match sequence: H0LISB Coords S14/525 SIGMA_P SignalP-2.8 Results: [Gr SignalP-2.8 Hesults: [Gr SignalP-2.8 Hesults: [Gr Signal periods probability Signal anchor probability Signal anchor probability Signal anchor probability Signal anchor probability	s To Display. Iron-binding region s CLEPH Procision 1.00 Iron-binding region si rom/OTMASE Procision 1.00 Iron-secretary Non-secretary 0.000 0.000 0.000	kgrature 2. Recoil 0.97 0.97 0.97 0.97 0.97 0.94 0.94 0.94 0.94	Ouration E Curration E M/TBN rated	Add To GO Evid	s C encel s C
Ne HMM-GO Suggestion PROSITE Proceeding Upperspectation Ratch sequences HVVR Coords S41/551 PS00711: Upperspectation S14/520 SIGHM_P SignalP-2.0 Head HX: [Crr SignalP-2.0 Head HX: [Crr Signal Peptide probability Signal anchor probability Signal enchor probability Signal Exclore gains probability	s To Display. iron-bindleg region s LLEPH Procision 1.00 Iron-bindleg region si mo-bindleg region si mo-bindleg region si Iron-bindleg region si procision 1.00 Iron-bindleg region si procision 1.00 Iron-bindleg region si Iron-bindleg region si Iron-bindl	Ignature 2. Recall 0.97 Recall 0.94 codput for 3P-HM protein = 000	Ouration E Curation E MATRAI rated	1465 To GO Evis	8 0 80(4) 8 0 5 0
No HMM-GO Suggestion PROSITE PSOCOL: Upoxygenases Match sequence: LINVFR S41/551 PSOCIL: Upoxygenases Match sequence: HOLISP S14/520 SIGMA_P Signal P-2.8 Results: [Gr Signal Petidion Signal petide probability Max cleavage site probability Max cleavage site probability	s To Display. Iron-binding region s LLEPH Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1.00 Procision 1	Ignature 2. Recall 0.97 Ignature 1. Peccell 0.94	Curution E Curution E M/19/1 rated	[Add To GO Evid	s 0 encaj s 0 s 0
Ne HMM-GO Suggestion PRIOSITE Proceed: Lipersymmases Match sequence: HVHF Coords S11/551 PSW/11: Lipersymmases Match sequence: H0LISH Coords S14/520 SIGMA_P SigmaiP-2.8 Results: [Gr: SigmaiP-2.8 Resul	s To Display. iron-binding region s LEFH Precision 1.00 Precision 1.00 Precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.00 precision 1.	Ignature 2. Recall 0.97 Recall 0.94	Ourstien E Ourstien E MARKI Parated	[Add To GO Evel [Add To GO Evel [Add To GO Evel	encel encel s C
Ne HMM-GO Suggestion PROSITE PSS0011 Upoxygenases Match sequence: HVVF8 Sequence: HVVF8 Coords S41:651 PSS0211: Upoxygenases Match sequence: HUJS9 Coords S14:620 SignalP-2.0 HMM Pediction Signal peptide probability Signal anchor probability Mac cleavage site probability BLAST SXIM	s To Display. Iron-binding region s LLEPH Procision 1:00 Iron-binding region si Iron-binding region si Iron-binding region si Iron-binding region si Iron-binding region si Iron-secretory 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:000 0:0000 0:0000 0:000 0:000 0:000	Ignature 2. Recall 0.97 Ignature 1. Recall 0.94 output for SP-HM protein  Cu Vecu DES (search date: Wed	Ouration E Ouration E Mittel arated B_Beaches I = Searches Oct 9 11:17:49 2002)	[Add To GO Evid	s 0 ence) s 0 o s 0
No HMM - GO Suggestion PROSITE PS00011: Uposygenases Match sequence: LHPVP Sequence: LHPVP Sequence: HOLISH Coords S14/520 SIGMA_P Signal P-2.8 HMM Prediction Signal Probability Signal acchor probability Signal acchor probability Max cleavage site probability	s To Display. iron-binding region s LLEPH Procision 1.00 Procision	kgnature 2. Rocat 0.97 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.97 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.97 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat 0.94 Rocat	Curation E Curation E M/1911 In Searchest In Searchest In Searchest Oct 9 11:17:49 2002) Do	Add To DO Evid	ence) ence) s C o s C
Ne HMM-GO Suggestion PRIOSITE Proceeding Lipeosygenases Match sequences HOUSH S41/551 PSS0711: Upersygenases Match sequence: HOUSH Coords 514/520 SIGRM_P Signal P-2.0 Results; [Crr Signal P-2.0 HMM Prediction Signal peptide probability Signal acchor	s To Display. iron-binding region s LLEPH Procision 1.00 iron-binding region si webriding region si phical Display [Ran binding region iso binding region bi	Ignature 2. Recall 0.97 Recall 0.94 Contput for 3P-HM contput fo	Ouration E Ouration E MATERI F R. Bourchess Get 3 11:17:45 2002) Get 3 11:17:45 2002	[Add To GO Evid [Add To GO Evid [Add To GO Evid [Sold To	0 5 C 0 5 C 0 5 C
Ne HMM- GO Suggestion PROSITE PS00011: Uporygenases Match sequence: LHPVP S41/551 PS0011: Uporygenases Match sequence: HOLISP S107AL_P SignalP-2.8 HMM Prediction Signal peptide probability Mac cleavage site probability Mac cleavage site probability Accession SIF-0638267	s To Display. Iron-binding region s LLEPH Iron-binding region s I.00 Iron-binding region s Iro	Ignature 2. Recall 0.97 Ignature 1. Procell 0.94 protein  Compart for 3P-HM protein  Compart for 3P-HM pro	Curation E Curation E M/1941 mated R_Searches ID_Searches ID_Searches December 10001 December 1000000000000000000000000000000000000	Add To GO Evid	s () ence) soce) s () s () s () s () s ()
Ne HMM-GO Suggestion PRIOSITE Proceeding Lipeorygenases Match sequence: HOLISE Coords S41/551 Provide Samerice: HOLISE Coords S14/520 SiteMa_P SignalP-2.8 Results: [Grg SignalP-2.8 HMM Prediction Signal peptide probability Max cleavage site probability Max cleavage site probability Accession SIF-006527 GP-16306627 GP-1630627 GP-16306627 GP-	s To Display. iron-binding region s LEFH Precision 1.00 Precision 1.00 Unical Display (Plan: Non-secretary 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0	Ignature 2. Recall 0.97 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Rec	Curation E Curation E Curation E MORE A Searches D Searches Curation E Searches Curation E Searches Curation E Searches Curation E Searches Curation E Searches Curation E Searches Curation E Searches Curation E Searches Curation E Searches Curation E Searches Curation E Searches Curation Curation E Searches Curation Curation Curation E Searches Curation Curation Curation E Searches Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Curation Cur	[Add To GO Evid [Add To GO Evid [Add To GO Evid ] [Add To GO Evid [Add To GO E	8 CE 80CE 80CE 80 S C 80 S C 80 S C 80 S C 80 S C 80 S C
Ne HMM-GO Suggestion PriOSITE Presource: Liposygenases Match sequence: LiPVPB Coords S41/551 Presource: HOUSP Coords S14/520 Signal P-2.0 HMM Prediction Signal peptide probability Signal anchor probability Mac cleavage probability Mac cle	s To Display. Iron-binding region s LLEPH Inclusion 1.00 Iron-binding region s ILO Iron-binding region s ILO Iron-binding region s ILO Iron-secretary 0.000 Iron-binding region s ILO Iron-secretary 0.000 Iron-secretary I	Ignature 2.  Recall 0.57  Recall 0.54  Recall 0.55  Recal	Curation E Curation E Mittel Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address Address A	Add To GO Evid	s ( ence) socal P-value 0.0 0.0 0.0 0.0
Ne HMM-GO Suggestion PROSITE Proceed: Unexpenses Author sequence: HVMP Coords S11/551 Coords S14/520 StofAL_P SignalP-2.8 Results: [Gr: SignalP-2.8	s To Display. iron-binding region s ILLEPH Precision 1.00 Procision 1.00 Procisio	Ignature 2. Recall 0.97 Recall 0.94 Contput for SP-HM protein Contput for SP-HM (search date: Wed Length 656 651 651	Curation E Curation E Curation E Mitten B. Sourches Ip. Sourches	Add To GO Evid Add To GO Evid (Add To GO Evid (Add To GO Evid (Corpus averance) (Corpus averance) (Corpus averance) (Corpus averance) (Corpus averance)	ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence) ence)
Ne HMM-GO Suggestion PRIOSITE Proceeding Lipeorygenases Match sequence: HOLISH Coords S41/551 PSS0711: Uperygenases Match sequence: HOLISH Coords S14/520 SIGHM_P Signal P-2.0 Results: [Grg SignalP-2.0 HMM Prediction Signal peptide probability Signal acchor probability Signal acchor probability Signal Accession RLAST SKIM	s To Display. iron-bindleg region s LLEPH Procision 1.00 iron-bindleg region si mo-bindleg region si mo-bindleg region si no.0 phical Display [Ban bindleg region si bindleg reg region si bindleg region si bindl	Ignature 2. Recall 0.97 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Rec	Curation E Curation E MARINI R. Banches Des Sanches Cot 3 11:17:45 2002) Des Sanches Des Sponyenas Sponyenas Sponyenas Sponyenas	And To GO Evid [And To GO Evid [And To GO Evid [And To GO Evid (And To	9 5 6 (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5) (0.5)(
Ne HMH-GO Suggestion PRIOSITE PSICOLI Uporypenases Match sequence: LHPVFB S41/551 PSICOLI Uporypenases Match sequence: HOU39 S10744_P Signal P-2.0 Results: [Gr Signal Petide probability Mac cleavage site probability Mac cleavage site probability Accession SIGAST SIGM CP-125455 CP-1505622 CP-1505622 CP-1505622 CP-1505622 CP-15054533 CP-250-504 CP-1722545 EGAD-134725 EGAD-13472 EGAD-13472 EGAD EGAD EGAD-13472 EGAD EGAD EGAD EGAD EGAD EGAD EGAD EGAD	s To Display. Iron-binding region s LLEPH Iron-binding region s I.00 Iron-binding region s Iro	Ignature 2. Recall 0.97 Recall 0.94 Conferit for 3P-HM protein C Conference (search dates) (search dates) Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference Conference	Curation  Curation  Curation  Curation  E  Curation  R Searches  (p.Searches (	Add To GO Evid (Add To GO Evid (Add To GO Evid (Add To GO Evid (State and State and State and State (State and State and State and State (Contra avelana) (Contra avelana) (Contra avelana) (Contra avelana) (Contra avelana) (Contra avelana) (Contra avelana) (Contra avelana)	s 0 ence) s 2 P-value 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.
Ne HMM-GO Suggestion PROSITE PSSOB11 Upoxygenases Match sequencil IUPOX Starts Starts5 Coords Coord	s To Display. iron-binding region s LEFH iron-binding region s iron-binding region s iron-binding region si ren40714.038E Procision 1.00 whicai Display [Rear binding region si ren40714.038E iron-secretary 0.000 atty 0.00	Ignature 2. Recall 0.97 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Recall 0.94 Rec	Curation	Jast To GO Evid Jast T	