Extending MediaWiki for community annotation

Daniel Renfro

daniel.paul.renfro@gmail.com
Texas A&M University





Outline

- Using Mediawiki vs. Wikipedia
- TableEdit & Mediawiki tables
 - Background
 - Function
 - Use cases
- Categories
 - As ontologies
 - GONUTS
- Educational Resource
 - Community Assessment of Community Annotation with Ontologies (CACAO)



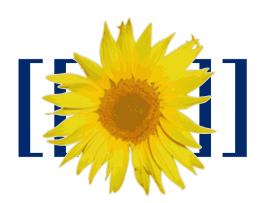


Using the MediaWiki software

- + Known track record
- + Easy to implement
- + Extensible
- + No/less compromising

- + Established community
- + Easy to maintain
- + Easy to customization

- Smaller audience (experts)
- Specific knowledge about one topic
- No "Google Factor"
- Cost of operation
- Higher barrier to entry







Freeform wikis & tabular data

```
ProbeSetName
                       Signal DetectionPValue Detection
                                                              DetectionString NumPairs
                                                                                              NumUsedPairs
       AFFX-BioB-5_at 25416.904297
                                       4.4e-05 0
       AFFX-BioB-M_at 42440.730469
                                       4.4e-05 0
                                                                      20
       AFFX-BioB-3_at 35719.859375
                                       4.4e-05 0
                                                                      20
       AFFX-BioC-5_at 70874.945313
                                       4.4e-05 0
       AFFX-BioC-3_at 47262.851563
                                       4.4e-05 0
       AFFX-BioDn-5_at 120802.867188
                                       4.4e-05 0
       AFFX-BioDn-3_at 201474.921875
                                       4.4e-05 0
       AFFX-CreX-5_at 429771.03125
                                       5.2e-05 0
       AFFX-CreX-3_at 429188.46875
                                       4.4e-05 0
       AFFX-DapX-5_at 1220.289673
                                       4.4e-05 0
10
       AFFX-DapX-M_at 941.896729
                                       0.011384
                                                                              20
       AFFX-DapX-3_at 729.904236
                                       0.000297
       AFFX-LysX-5_at 102.202423
                                       0.368427
13
       AFFX-LysX-M_at 172.067215
                                       0.131361
14
       AFFX-LysX-3_at 224.006241
                                       0.013804
15
       AFFX-PheX-5_at 355.921295
                                       0.00034 0
       AFFX-PheX-M_at 186.113037
                                       0.089478
                                                                              20
17
       AFFX-PheX-3_at 75.586524
                                       0.574044
                                                                              20
       AFFX-ThrX-5_at 295.346344
                                       0.010311
```

- Typical data for a Model Organism Database:
 - Structural data
 - Microarrays
 - Genomic coordinates / maps
 - Features (genes, ORFs, SNPs, etc.)





Wiki tables

- Hard to load data into
- Hard to mine data out-of
- Esoteric markup
 - Not good for the uninitiated
- Semantic tagging
 - More specialized markup

```
{| class="wikitable"
|-
! header 1
! header 2
! header 3
|-
| row 1, cell 1
| row 1, cell 2
| row 1, cell 3
|-
| row 2, cell 1
| row 2, cell 2
| row 2, cell 3
|}
```

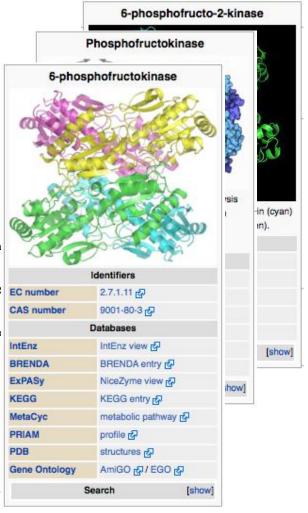
Header 1	Header 2	Header 3			
row 1,cell 1	row 1, cell 2	row 1, cell 3			
row 2,cell 1	row 2, cell 2	row 2, cell 3			
row 3,cell 1	row 3, cell 2	row 3, cell 3			





Mediawiki Tables

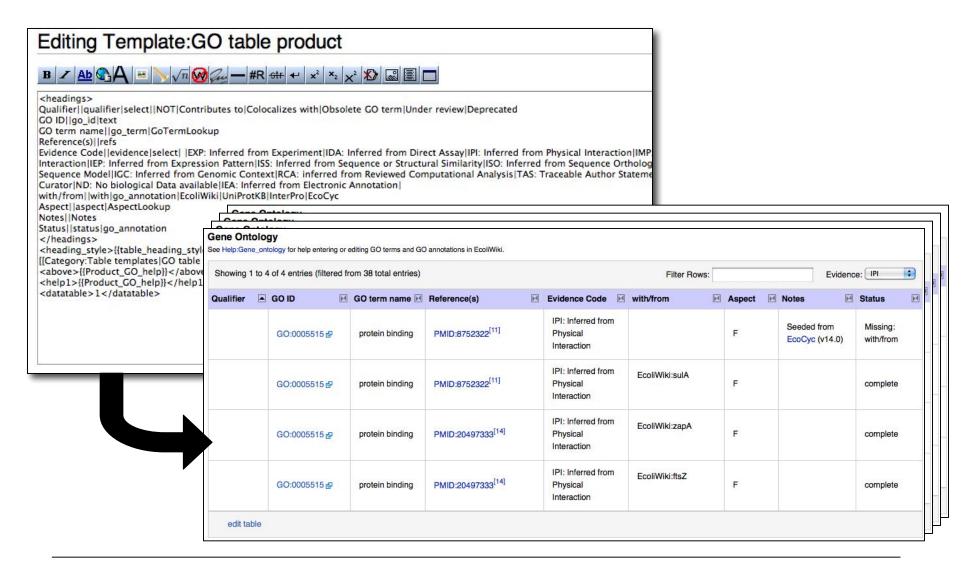
```
{{FixBunching|beg}}
{{enzyme
  Name = 6-phosphofructokinase
  EC number = 2.7.1.11
  CAS number = 9001-80-3
  IUBMB EC number = 2/7/1/11
  GO code = 0003872
  image = Phosphofructokinase 6PFK wpmp.png
  width =
  caption =
}}
{{FixBunching|mid}}
{{Infobox protein family
  Symbol = PFK
  Name = Phosphofructokinase
  image = Phosphofructokinase (active vs inactive form).png
  width =
 caption = Bacterial Phosphofructokinase: 3rd glycolysis enzyme (smaller than in
Eukaryotes). In yellow=sugars; in red=ATP-ADP; starred=regulatory sites.<ref
name="pmid6115424">{{PDB|4pfk}}; {{cite journal | author = Evans PR, Farrants GW,
| title = Phosphofructokinase: structure and control | journal = Philos. Trans. R
Lond., B, Biol. Sci. | volume = 293 | issue = 1063 | pages = 53-62 | year = 1981
June | pmid = 6115424 | doi = 10.1098/rstb.1981.0059 | laysummary =
http://www.pdb.org/pdb/static.do?p=education discussion/molecule of the month/pdb
 laysource = PDB Molecule of the Month }}</ref>
 Pfam = PF00365
  Pfam clan = CL0240
  InterPro = IPR000023
  SMART =
  PROSITE = PDOC00336
  SCOP = 5pfk
  TCDB =
  OPM family =
  OPM protein =
 PDB = \{ PDB2 | 1kzh \} \}, \{ PDB2 | 1mto \} \}, \{ PDB2 | 1pfk \} \}, \{ PDB2 | 1zxx \} \}, \{ PDB2 | 2f48 \} \}
{{PDB2|2pfk}}, {{PDB2|3pfk}}, {{PDB2|4pfk}}, {{PDB2|6pfk}}
{{FixBunching|end}}
```







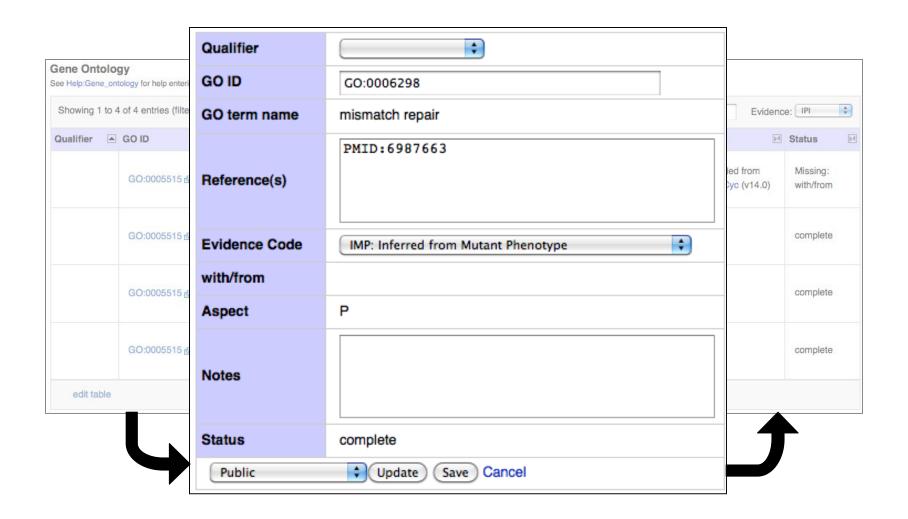
TableEdit







TableEdit







Example: Table Edit graph generation

Domains/Motifs/Modification Sites

See Help:Product_domains_motifs for help entering or editing information in this section of

Showing 1 to 5 of 5 entries Search all columns: ■ Residues Description Notes References M Type PF01624 PMID:19920124^[19] Domain 11..123 MutS domain I 图 PF05188 PMID:19920124^[19] Domain 131..256 MutS domain II @ PF05192 PMID:19920124^[19] Domain 264..564 MutS domain III @ PF05190 PMID:19920124^[19] Domain 428..519 MutS family domain IV @ PF00488 PMID:19920124^[19] 567..799 Domain MutS domain V @ edit table

```
PF01624 MutS domain I

PF05188 MutS domain II

PF05192 MutS domain III

PF05190 MutS family domain IV

PF00488 MutS domain V
```

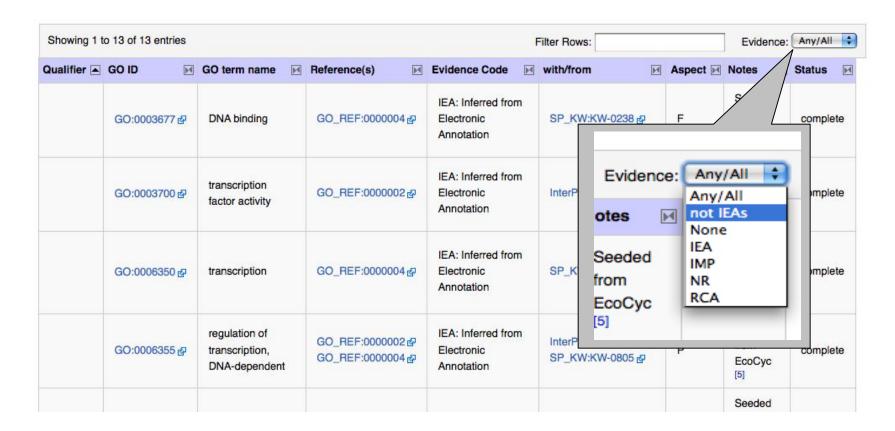
```
if ( defined( 'MW_SUPPORTS_PARSERFIRSTCALLINIT' ) ) {
    $wgHooks['ParserFirstCallInit'][] = 'efMotifMapExtension';
} else {
    $wgExtensionFunctions[] = "efMotifMapExtension";
}

function efMotifMapExtension() {
    global $wgParser;
    $wgParser->setHook( "motif_map", "efRenderMotif_map" );
    return true;
}
```





Example: Table Edit Javascript, jQuery, and Data Tables

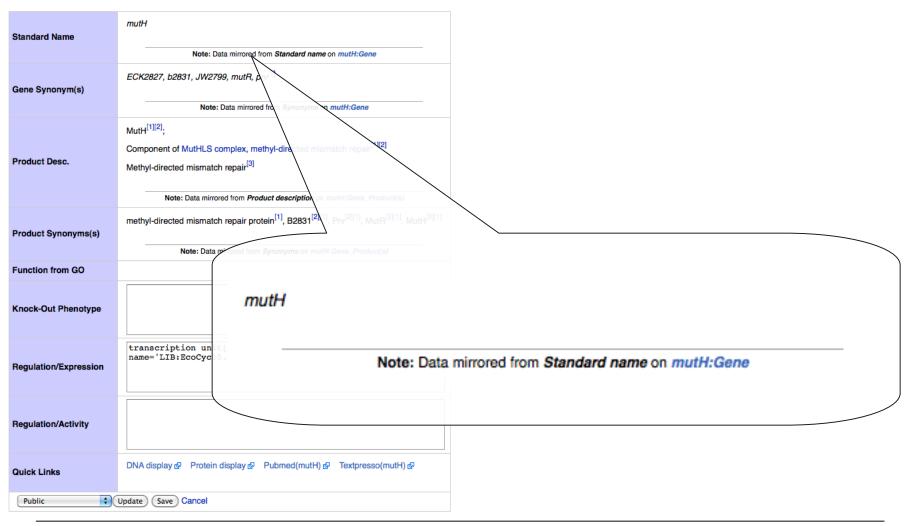


Many thanks to Allan Jardine at http://datatables.net





Example: Table Edit Mirroring data







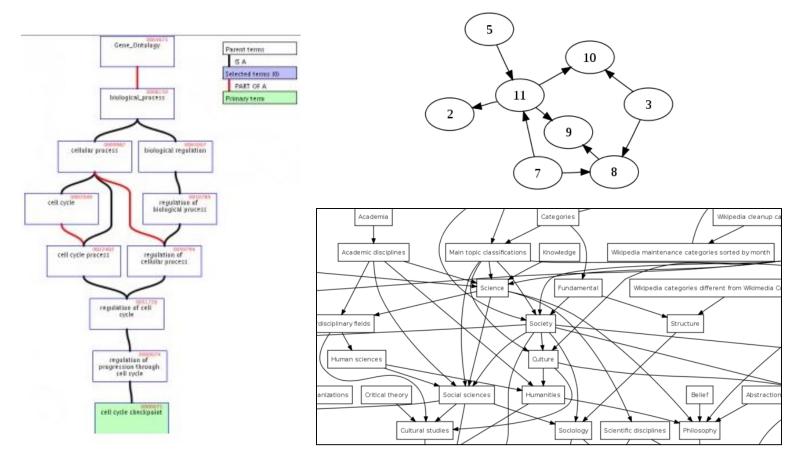
Outline

- Using Mediawiki vs. Wikipedia
- TableEdit & Mediawiki tables
 - Background
 - Function
 - Use cases
- Categories
 - As ontologies
 - GONUTS
- Educational Resource
 - Community Assessment of Community Annotation with Ontologies (CACAO)





Traversing the GO with categories



A typical GO term.

Two directed acyclic graphs (DAGs)





A wiki for Gene Ontology

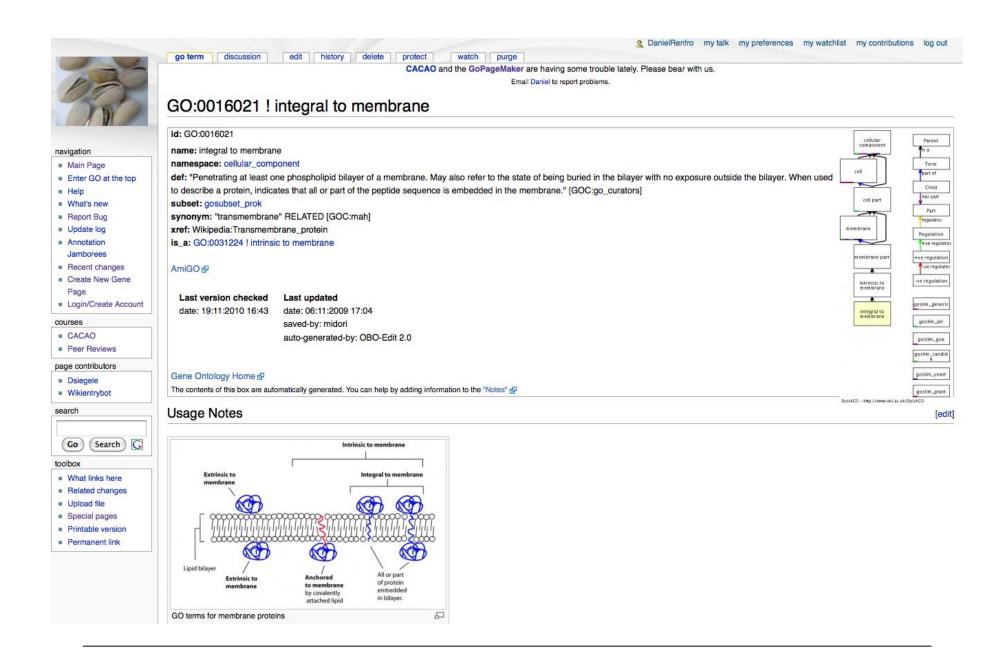
GONUTS

- Gene Ongology Normal Usage Tracking System
- Finding the right GO term (out of ~285,000 terms)
- Best practices, user notes, etc.
- Annotating any gene in UniProt
- Example annotations from MGI, FlyBase, WormBase, dictyBase...

(http://gowiki.tamu.edu)











From the Cellular Component Ontology Guidelines

References

See Help:References for how to manage references in GONUTS.

Child Terms

This term has the following 10 child terms.

- = [+] GO:0005887 integral to plasma membrane (39)
- [] GO:0009319 cytochrome o ubiquinol oxidase complex
- = [+] GO:0016472 sodium ion-transporting two-sector ATPase complex (2)
- [] GO:0017090 meprin A complex
- = [-] GO:0031301 integral to organelle membrane (10)
 - [] GO:0005639 integral to nuclear inner membrane
 - [] GO:0005779 integral to peroxisomal membrane
 - [] GO:0030173 integral to Golgi membrane
 - [-] GO:0030176 integral to endoplasmic reticulum membrane (5)
 - [] GO:0032937 SREBP-SCAP-Insig complex
 - [] GO:0042765 GPI-anchor transamidase complex
 - [+] GO:0042824 MHC class I peptide loading complex (1)
 - [] GO:0071458 integral to cytosolic side of endoplasmic reticulum membrane
 - [] GO:0071556 integral to lumenal side of endoplasmic reticulum membrane
 - [] GO:0030285 integral to synaptic vesicle membrane
 - [+] GO:0031166 integral to vacuolar membrane (2)
 - [] GO:0031303 integral to endosome membrane
 - [] GO:0031309 integral to nuclear outer membrane
 - [+] GO:0031351 integral to plastid membrane (2)
 - [+] GO:0032592 integral to mitochondrial membrane (2)
- Pages in category "GO:0016021! integral to membrane"

The following 200 pages are in this category, out of 21,495 total.

Show articles starting with: --All results-- Co

(previous 200) (next 200)

•

- APIME:Q8LUG3
- В
- BACSU:FTSK
- BISBI:A0A028
- C
- CHICK:A0FCL5
- CHICK:A0FEQ2

C cont.

- CHICK:A6N241
- = CHICK:A6N242
- _____
- CHICK:A6N8N7
- CHICK:A6QR74
- CHICK:A6YIJ1
- CHICK:A6YIJ2
- CHICK:A6YJX2
- CHICK:A6YJX3
- = CHICK-A7E3K3

= [] GO:0031361 - integral to thylakoid membrane

- [+] GO:0034702 ion channel complex (4)
- [] GO:0045203 integral to cell outer membrane
- [] GO:0046696 lipopolysaccharide receptor complex
- = [+] GO:0046930 pore complex (3)

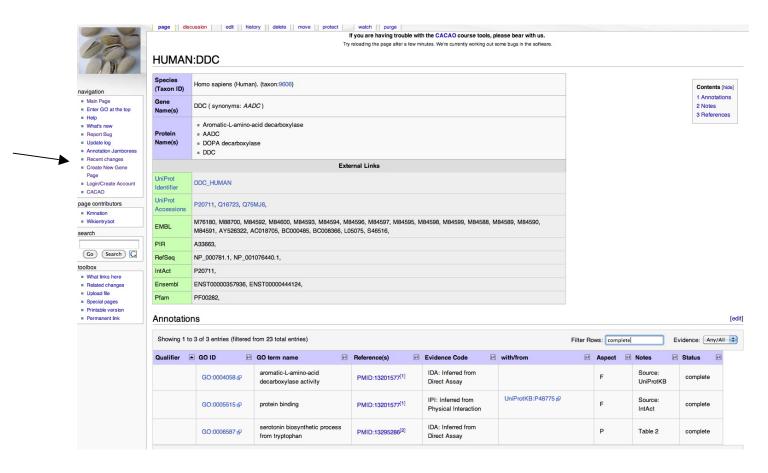
C cont.

- = CHICK:AT134
- = CHICK:AT1A1
- = CHICK:AT1A2
- = CHICK:AT1A3
- = CHICK:AT1B1
- = CHICK:AT1B3
- CHICK:AT2A1
- CHICK:AT2A2
- = CHICK-AT2A3





Annotation on GONUTS



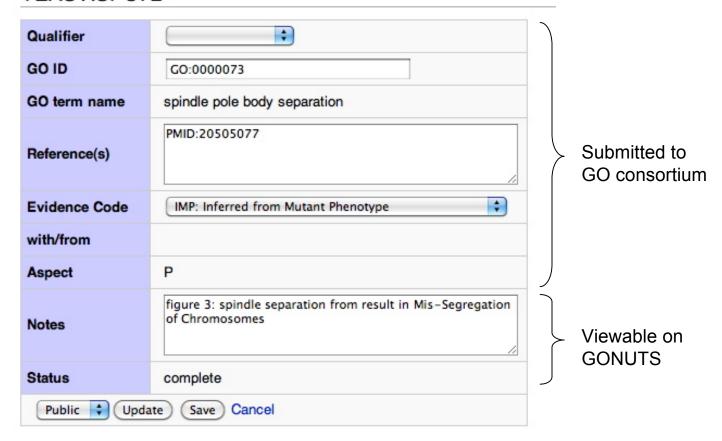
- Users can create gene pages for anything in UniProt.
 - New gene pages are populated with information, including previous GO annotations.





Annotation on GONUTS

YEAST:SPC72







Comparing annotations

Organisms

PFKL [edit]

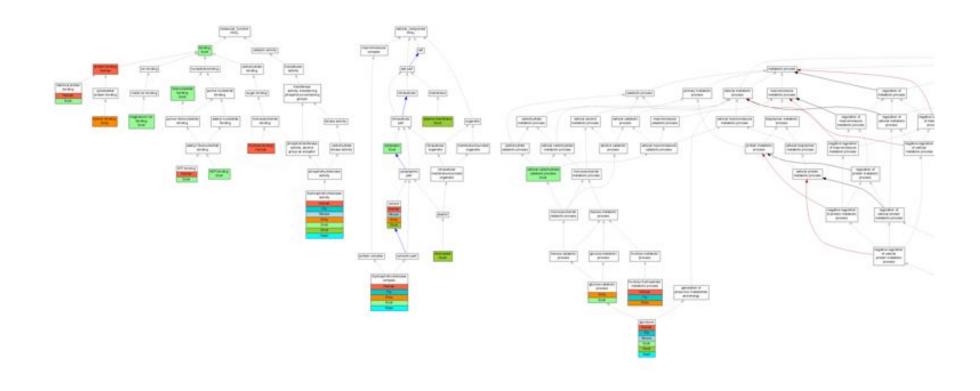
2		
=		
_		
J		
)		
)		

Category M	ID M	Term ⋈	Human 🖂	Mouse M	Rat M	Chicken M	Zfish 🖂	Fly 🖂	Worm ⋈	Dicty 🖂	Dicot 🖂	Yeast ⋈	Pombe M	Ecoli 🖂
Biological Process	GO:0040018	positive regulation of multicellular organism growth	х	х	x	x	х	х	IMP	х	x	х	х	х
Biological Process	GO:0006096	glycolysis	IDA	IDA	Х	X	Х	IC	X	X	IDA ISS	IMP IDA	X	IDA
Biological Process	GO:0006002	fructose 6- phosphate metabolic process	IDA IMP	x	x	х	х	IMP	x	IDA	х	x	x	х
Biological Process	GO:0031115	negative regulation of microtubule polymerization	х	x	x	х	х	x	x	IDA	х	x	х	x
Biological Process	GO:0006007	glucose catabolic process	x	x	х	X	х	х	x	IGI	х	x	х	IMP
Biological Process	GO:0009792	embryonic development ending in birth or egg hatching	х	x	x	x	x	x	IMP	х	x	x	х	x
Biological Process	GO:0000003	reproduction	Х	X	Х	X	х	Х	IMP	X	X	X	Х	X
Biological Process	GO:0044275	cellular carbohydrate catabolic process	х	x	x	x	х	x	x	х	x	x	x	IMP
Biological Process	GO:0016052	carbohydrate catabolic process	x	x	x	х	X	x	x	х	X	x	x	IMP
Biological Process	GO:0046676	negative regulation of insulin secretion	х	IDA	x	X	х	х	х	х	x	х	х	х
Biological Process	GO:0009749	response to glucose stimulus	х	IDA	x	х	х	х	x	X	х	х	х	x
Molecular Function	GO:0003872	6- phosphofructokinase activity	IDA IMP	IDA	x	x	х	IMP	х	IDA IGI	IDA ISS	IMP	×	IMP
Molecular Function	GO:0015631	tubulin binding	Х	X	X	X	X	X	X	IPI	Х	X	X	X
Molecular Function	GO:0005524	ATP binding	IDA	X	Х	X	Х	Х	Х	X	X	X	X	IMP
		identical protein	7	19	P P									²⁰





Comparing annotations



Thanks to Mary Dolan @ MGI





Outline

- Using Mediawiki vs. Wikipedia
- TableEdit & Mediawiki tables
 - Background
 - Function
 - Use cases
- Categories
 - As ontologies
 - GONUTS
- Educational Resource
 - Community Assessment of Community Annotation with Ontologies (CACAO)





CACAO

coupling annotation to teaching credit





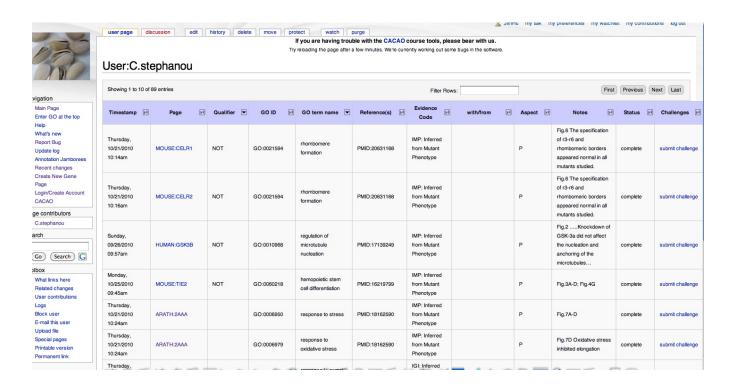
- Teams of students curate
- Faculty supervision
- Support from our team
- Intramural or Intercollegiate competition
- Distributed annotation jamborees
- Assessment via surveys and wiki data-mining

April 2010 (CACAO v0.1) 117 / 153
Fall semester ? / 757





Tracking the players/teams



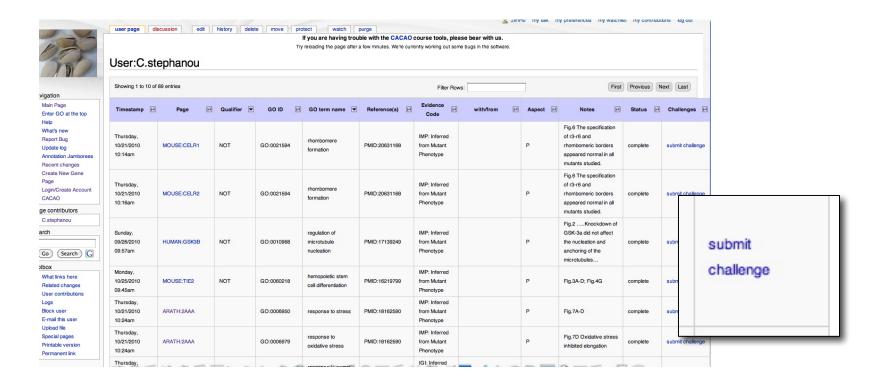
An extension tag added to a user page identifies all the annotations made by that user

<myAnnotations />





Tracking the players/teams



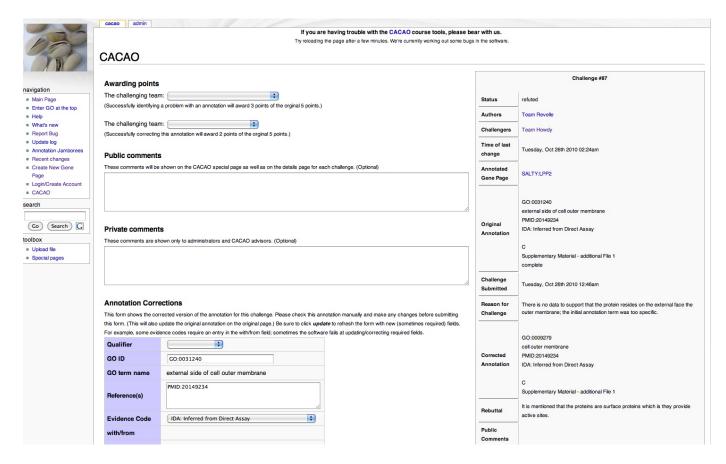
An extension tag added to a user page identifies all the annotations made by that user

<myAnnotations />





Judgement



Mentors with curator experience judge the challenges/rebuttals





Overall scoreboard

Scoreboard

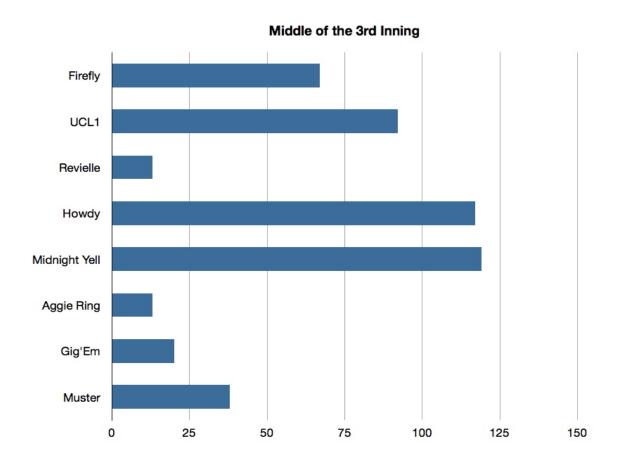
current standing	team name	number of unchallenged, complete annotations	number of problems correctly identified	number of problems corrected	total points	
1	Team Howdy	193	18	13	900	
2	Team Firefly	135	43	38	815	
3	Team UCL1	159	0	0	795	
4	Team Midnight Yell	116	2	2	532	
5	Team Muster	55	3	3	290	
6	Team Aggie Ring	49	0	0	243	
7	Team Gig Em	25	5	1	137	
8	Team Revelle	25	1	1	113	
total		757	72	58		

• A scoreboard page gathers information about all teams and challenges





Students want to win



The lead has been changing hands every inning





Acknowledgements

- EcoliWiki/GONUTS Team
 - Jim Hu, P.I.
 - Debby Siegele, co-P.I.
 - Brenley McIntosh
 - Adrienne Zweifel
 - Dave Clements
 - Nathan Liles
 - Amanda Supak
 - Chanchala Lairikyengbam
 - Joy Wang
- CACAO
 - Mentors at University College London
 - undergrads / grad students



















