Approaches to Community Annotation

Chris Stoeckert

May, 2005 BRC meeting

VBI

Community Annotation

- What do we mean by this?
  - Only a few staff annotators (0-5)
  - Unpaid volunteers
    - Deeper insights from individual experts
    - Less consistency as the sum is from many viewpoints
- Subsytems versus genomes
- Evidence trail
- Updating “official” annotation

Types of Annotation

- Gene Models
  - Sequencing centers
  - Staff annotators
  - Individual investigators
- Functional Annotation (Pathways, Processes, Roles)
  - Experimental
    - Literature
    - GO, E.C., others
  - Predicted
    - Sequence features (e.g., targeting sequence)
    - orthologs
    - Expression analysis

Annotation Mechanisms

- Tools
  - Apollo, Manatee, Artemis
    - Sequence Ontology
  - ASAP, Seed, CVE, others?
  - Text mining
    - Assign papers
    - Use as annotation aid
- Jamborees
- Editorial board
- Individual contributors

PlasmoDB Use Cases

- User comments
  - Comment on Gene Page item or region from Genome Browser
  - Can query text
  - Send to staff annotator (Andy Berry at Sanger)
- Relevance of papers
  - Ranked list of papers tied to a gene using text mining
  - Users can assign relevance of papers that can then be used for training
- Jamboree
  - Phase 2 annotation 2006
  - Community experts and staff annotators
  - Lots of planning!

Plasmodium falciparum / PFD1150c / Community annotation

hypothetical protein

ADD NOTE: Use the form below to add a new note for PFD1150c

Notes submitted by PlasmoDB users

Osamu Kaneko 03:23:13 Nov-18-2004 EST
ehime university school of medicine

PfRH4, reticulocyte binding protein homolog 4. One intron is missed in the annotation. Correct mRNA sequence can be found under accession number AF432854.

Submit a new note for PFD1150c

Anyone in the PlasmoDB user community is welcome to submit comments to PlasmoDB. However, before doing so, you need to register with us . This will allow you to securely set up a username and password of your choice, and will allow us to verify your e-mail address so that we can ensure that all of the comments in PlasmoDB are correctly attributed.
Forgotten your password?

Please note that if you have a username and password from an earlier version of PlasmoDB (e.g., prior to the official publication of the genome), they will not work on this form.

username:

password:

New note for *Plasmodium falciparum* gene PFD1150c:

Gene ID	Gene Names
MAL6P1.194	- Proposed Name for the gene MAL6P1.194: PfSOD2 (8 introns and 9 exons).
MAL6P1.146	The published PfPK4 sequence starts with the amino acid string sequence MKKRIRSSYK which maps at amino acid position ~1910 of PFF1370w and not ~1190 as mentioned in the PUB_COMMENT.
PFA0110w	RESA; ring-infected erythrocyte surface antigen; Pf155;
PFA0125c	JESEBL, EBA-181
PFA0680c	Mauer's cleft 2-transmembrane domain protein. This gene has been extended to include: "MFHYIYKIYIFTIILCASNLFNNNVVEIGTYKLSYHNGGIQFR" upstream of the originally annotated protein coding start.
PFB0300c	Synonyms for merozoite surface protein 2: MSP2; MSP-2; MSA-2; MSA2; merozoite surface antigen 2
PFB0340c	serine rich antigen; SERA; p126; cysteine protease; serine rich protein; SERP;
PFC0025c	Plasmodium falciparum gene
PFC0490w	putative dolichyl-diphospho-oligosaccharide-protein (2.4.1.119) (based on orthologs in human, yeast, etc. See ortholog group 746046 for further details)
PFC0890w	Putative sec22
PFC0975c	This cyclophilin (PFCYP19) is also refered to as CPR3. See RefSeq entry NP_473329.
PFD0795w	putative histone acetyltransferase (EC 2.3.1.48) (based on orthologs in human, yeast, etc.; see ortholog group 766015 for further details)
PFD0965w	putative phosphatidylinositol kinase (EC 2.7.1.67) (based on orthologs in yeast, worm, etc.; see ortholog group 765426 for further details)
PFD1150c	PfRH4, reticulocyte binding protein homolog 4. One intron is missed in the annotation. Correct mRNA sequence can be found under accession number AF432854.
PFD1155w	PEBL, Psi EBA-165
chr5.glm_15	Is actually PFE0040c- PfEMP2 with 100% identity (annotation must be wrong!!)
PFE0120c	PFE0120c is also refered to as MSP8. See RefSeq NP_703370.
PFE0660c	This protein has now been functionally characterized as "Purine Nucleoside Phosphorylase" (See references below). The annotation should be changed appropriately.
PFE1150w	PFE1150w is also called pfmdr1, and Pgh-1
PFE1150w	The gene PFE1150w shows identity to RefSeq entry NP_703574. In this entry the gene is named MDR1.
PF07_0001	This is identical in sequence to RefSeq NP_703978. The gene product is identified as a RIFIN, but is given the name 1838.m00006.
PF07_0128	EBA-175, SABP. Two introns are missed in the annotation.

Discussion Summary

- Experimental rather than predicted (algorithmic) annotation
- Community coherence is an issue
  - Jamborees less than successful other than a publication
  - Smaller groups with follow up are better
- Pay with publication
  - Editorial board (Alliance for Cell Signaling and Nature publications)
- Evidence
  - Who, when
  - Text box
  - Evidence box