Tools and Techniques for Innovative Annotation:
Contributing to a Kinetoplastida Knowledge Base

Sandhya Subramanian[1], Isabelle Phan[1], Christian Olsen[1], Carol Farris[1], Gowthaman Ramasamy[1], Matthew B. Rogers[2,6], Flora Logan[2,5], Haiming Wang[3], Brian P. Brunk[4], Emily J. Glenn[1]

AUTHOR AFFILIATIONS: TriTrypDB Consortium: [1]Seattle BioMed, Seattle, WA; [2]Wellcome Trust Sanger Institute, Cambridgeshire, UK; [3]Center for Tropical and Emerging Global Diseases, University of Georgia, Athens, GA; [4]University of Pennsylvania, Philadelphia, PA; [5]University of Cambridge, Cambridge, UK; [6]University of York, York, UK.

Teams of scientists and bioinformaticians at four institutions are using web-based collaboration tools to facilitate the annotation of the genomes of Kinetoplastida. This collaborative effort provides foundational information for TriTrypDB, an integrated genomic and functional genomic database for pathogens of the family Trypanosomatidae, including organisms in both *Leishmania* and *Trypanosoma* genera. These pathogens are associated with Chagas' disease, leishmaniasis and African sleeping sickness – all parasitic infectious diseases for which no vaccine or drug to prevent infection is available.

Genomic annotations in TriTrypDB are prepared *via* rapid information and knowledge exchange between teams of literature annotators and data curators. Teams follow an open information management infrastructure using web-based applications: **Zotero** (a bibliographic and research information management system), **JIRA** (traditionally used for software development bug tracking), and **Google groups**. Dissemination of Kinetoplastida genome knowledge is accelerated through the innovative use of these web-based tools as an open information sharing platform.
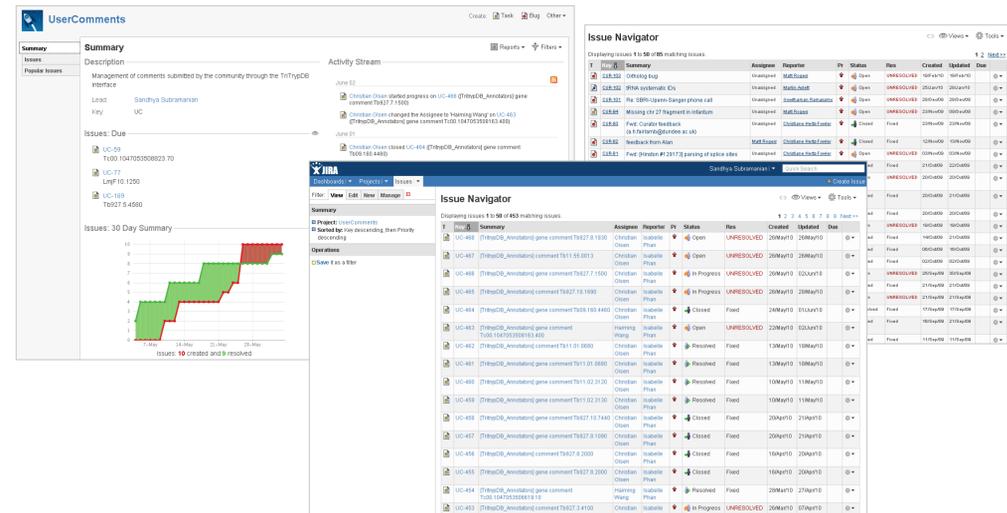
Kinetoplastids are a group of flagellate protozoa, including a number of parasites responsible for diseases like leishmaniasis, African sleeping sickness, Chagas' disease.

Genome annotation is the process of attaching biological information to sequences by identifying elements on the genome and attaching biological information to them.
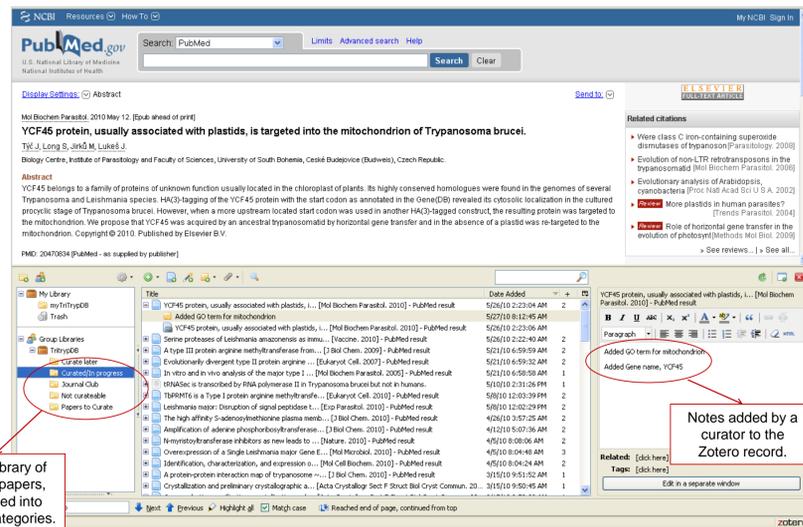

Shared library of relevant papers, subdivided into working categories.

Added GO term for mitochondrion.

Added Gene name, YCF45

Notes added by a curator to the Zotero record.

A browser-based **Zotero library** is used as a high-level staging area to identify publications, tag them, and track their association with existing curated database entries. Zotero thus permits annotators to see quickly what has already been extracted from the paper as annotation-worthy.

JIRA is used to coordinate TriTrypDB users' comments, track annotation issues, share pertinent unpublished information on genomics discovery, and trigger uploads of annotations to TriTrypDB. JIRA is also used for general project management, including time and effort allocation.
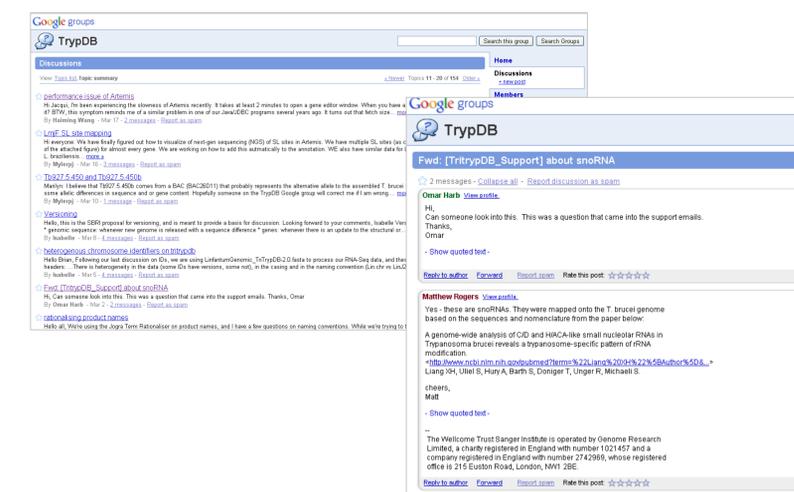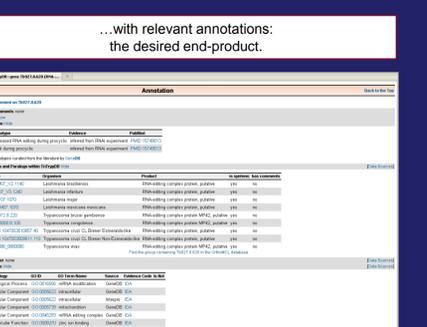
Google Groups provides a secure space for discussing the project, with group members. In this collaborative space, decision histories become part of a project knowledge base that can be accessed more readily than e-mail threads.

Aslett et al. **TriTrypDB: a functional genomic resource for the Trypanosomatidae.** Nucleic Acids Research 2010 38 (Database issue):D457-D462; doi:10.1093/nar/gkp851

Synteny map view of an entry in TriTrypDB

…with relevant annotations: the desired end-product.
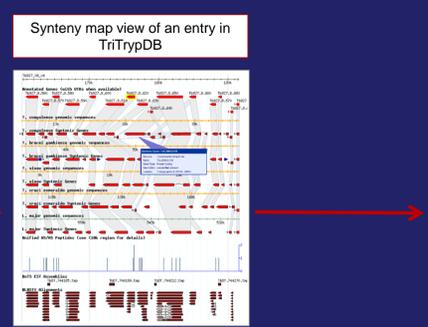
Tb927.8.620. T. brucei TREU927 protein coding gene on Tb927_08_v4 from 175,388 to 176,569 (Chromosome: 8)