

Data Sustainability and Funding: AgBase

Fiona McCarthy, Shane Burgess, Bindu Nanduri.

PAG XXV, Data Sustainability and Funding Workshop
18 January 2017

1. About AgBase

- Resource to support functional modeling in agricultural organisms.
- Provide functional annotations
 - GO, MI annotation
 - manual & computational
- Provide tools to support functional modeling
- Provide support for community.



The screenshot shows the AgBase website interface. At the top left is the AgBase logo with the version number [Version: 2.00]. To the right is the Mississippi State University logo. Below the logo is a navigation bar with links: HOME, SEARCHES, TOOLS, ANIMALS, PLANTS, MICROBES, PARASITES, HELP, CONTACT. The main content area features a search box with the text "GO AgBase Search" and "Enter Multiple Queries (new line separated)". There are "Search" and "Reset" buttons. To the right of the search box are several database logos: Chickspress, Bird Base, iAnimal, Chicken Gene Nomenclature, and HPIDB. A sidebar on the left contains links for "About Annotation", "Community Requests & Submissions", "Educational Resources", "Downloads & Statistics", "Array Annotation", "Journal Database", and "Microbial GBrowsers".

<http://www.agbase.msstate.edu>



1. About AgBase: data sources

- External data consumed by AgBase:
 - Gene Ontology (annotations, ontologies, slim sets)
 - Plant Ontology
 - NCBI, Ensembl, UniProtKB sequence, accessions, mapping data, etc
 - PubMed
 - HGNC HCOP data



1. About AgBase: data generated

- Gene Ontology annotations (EBI-GOA)
 - 1,753,795 GO annotations for 341,795 gene products from 504 species (Nov 2016)
 - focused manual curation for 11 species
- Molecular Interactions
 - HPIDB via IntAct
 - 3,664 manually curated interactions
- Standardized gene nomenclature for chicken (via CGNC)
- Chicken expression data (via Chickspress)



1. About AgBase: use cases

- 2006:
 - livestock researchers with list of differentially expressed genes from microarray data
 - limited/no bioinformatics experience
 - limited set of species being studied
 - limited set of genes
 - functional enrichment (GO, pathways)
- 2016:
 - transcriptome data sets (larger, more complex)
 - very broad range of species
 - need to provide genome-wide functional annotation prior to functional enrichment & modeling
 - functional enrichment, networks, comparative

2. AgBase Activities

- Manual curation of GO and MI
 - 2 GO biocurators, 1 MI curator (targeted curation)
 - also provide gene nomenclature for chicken
- Computational annotation
 - 1 database manager/developer
 - developer new functional annotation workflows
- Support for functional annotation & modeling of non-traditional model species
 - ad hoc, as requested
 - time-intensive!



2. AgBase Activities

Are you working on any projects aimed at reducing costs?

1. Training users

- to reduce costs associated with community support
- documenting workflows for functional annotation & modeling using open software & public access computing resources

2. Computation

- moving computational workflows to public resources (e.g., CyVerse)



3. Biocuration

- eGIFT: text-mining to identify papers for biocuration (triage)
- Use of existing biocuration interfaces (EBI-GOA, IntAct)



3. Funding!

- Current funding:
 - USDA Animal Genomic Tools
 - NIH funding (now completed)
- Future funding:
 - USDA for support of livestock genomes
 - NSF for development of computational workflows
 - NIH (for chicken)?

3. Funding!

- Is your current funding adequate?



3. Funding!

- Is your current funding adequate?
 - on average, have had 2 biocurators/year for 6 major livestock species
 - main problem is one of continuity: lose trained biocurators, then [if we get more funding] we have to train new staff
- What could you do with more funding?
 - more biocuration...
 - ...or better still, functional biocuration to support FAANG Project
 - desperate need for training & workflows so that functional analysis of gene expression data sets is routine and informative

3. Funding: future sources

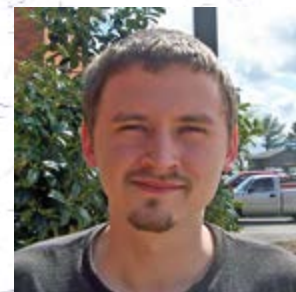
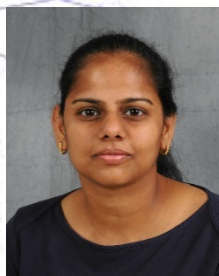
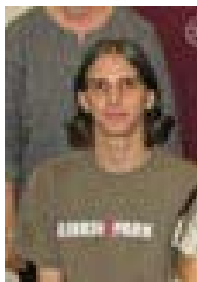
- Continued application to “traditional” sources:
 - USDA for FAANG animals
 - NSF for development of workflows
 - NIH for chicken (biomedical model)
- Collaborative model
 - incorporating partial funding in with specific grant (annotation & analysis support)
 - training and support
- Sustainability
 - distribution of data to public resources (GO, IntAct, PO, etc)
 - computational pipelines: CyVerse
 - documentation and public release of workflows for functional modeling (+ outreach & training)
 - knowledge ??



AgBase



AgBase



MISSISSIPPI STATE
UNIVERSITY™



THE UNIVERSITY
OF ARIZONA.

