The i5k Workspace@NAL – A place for arthropod genome communities to curate, visualize and share data

Monica Poelchau¹, Mei-Ju May Chen², Yu-Yu Lin², Gary Moore¹, Vijaya Tsavatapalli¹, Christopher Childers¹

¹USDA/Agricultural Research Service/National Agricultural Library, Beltsville, MD,
²Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan

What is the i5k Workspace@NAL?

- A workspace for genomic data access, dissemination, and curation for arthropods, hosted by the USDA’s National Agricultural Library (NAL)².
- We currently host genome project data for 53 arthropod species.
- We provide a central organism page for each project, gene pages for projects with an Official Gene Set, data downloads, BLAST³, and the JBrowse⁴ genome browser.
- We facilitate community gene curation via Apollo⁵.
- The i5k Workspace is built on a customized version of Tripal⁶.

Start an i5k Workspace project.

- Any orphaned arthropod genome project in need of manual curation or other genome portal resources can submit their data.
- Our main requirement is that the assembly needs to be accessioned by the INSDC (e.g. in GenBank). If your genome is already hosted in another database, let’s discuss.
- Contact i5k@ars.usda.gov to get started.

Join an i5k Workspace project.

1. Help annotate an existing genome project.
   - Over 400 annotators curating over 30 genome projects
   - Register for an Apollo account:
     https://i5k.nal.usda.gov/web-apollo-registration
2. Add data to an existing genome project.
   - Browse our list of existing projects:
     https://i5k.nal.usda.gov/species
   - Register here for a data submission account:
     https://i5k.nal.usda.gov/register/project-dataset/account

How is the i5k Workspace being used?

- We have over 400 registered annotators, who have manually curated over 10,000 annotations using Apollo.

Updates for 2016

New web applications

- ClustalW⁷, ClustalOmega⁸, and Hmmer⁹ web applications provide more sequence search options. Built in the Django framework. Features include: automatic input file format detection; query queuing system; and user accounts for result retrieval.

New data submission forms

- Our new web-based data submission forms make it easier to start i5k Workspace projects, and to add data to existing projects.
- Sign up for an account:
  https://i5k.nal.usda.gov/register/project-dataset/account

New organisms and datasets

- New species for 2016 include Dufourea novaeangliae, Habropoda laboriosa, Lasiosglossum abipes, Megachile rotundata, Melipona quadrifasciata, Micropilis demolitor, Diachasma alloaeum, Neodiprion lecontei, Amyelois transitella, and Tribolium castaneum.
- Official Gene Set pipeline developed at the NAL
  - Four are complete
  - Two are in progress
  - https://github.com/NAL-i5K/i5KNAL_OGS
- We regularly receive new RNA-Seq datasets from our contributors to facilitate manual curation.

For more information about i5k

- Visit us online: https://i5k.nal.usda.gov
- Visit us at Booth #320.

References

Acknowledgments and Funding

We would like to thank our data providers, the i5k coordinating committee, NAL leadership, and the NAL Information Systems Division team for their support and encouragement of this project. United States Department of Agriculture-Agricultural Research Service provided project support through the offices of the National Agricultural Library, Office of National Programs, and the Bee Research Laboratory.