

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

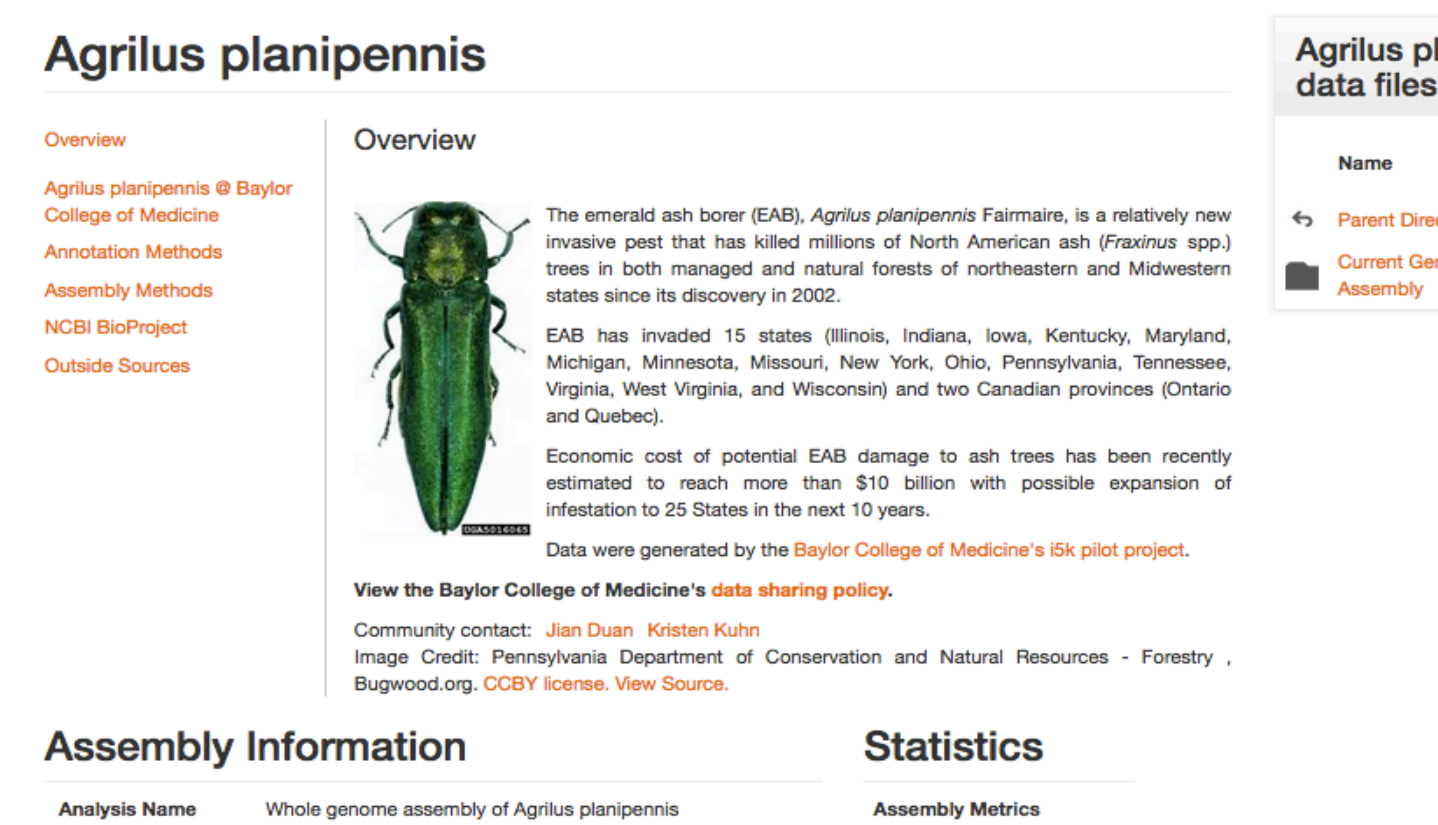
Monica Poelchau¹, Mei-Ju May Chen², Yu-Yu Lin², Gary Moore¹, Vijaya Tsavatapalli¹, Christopher Childers¹

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?

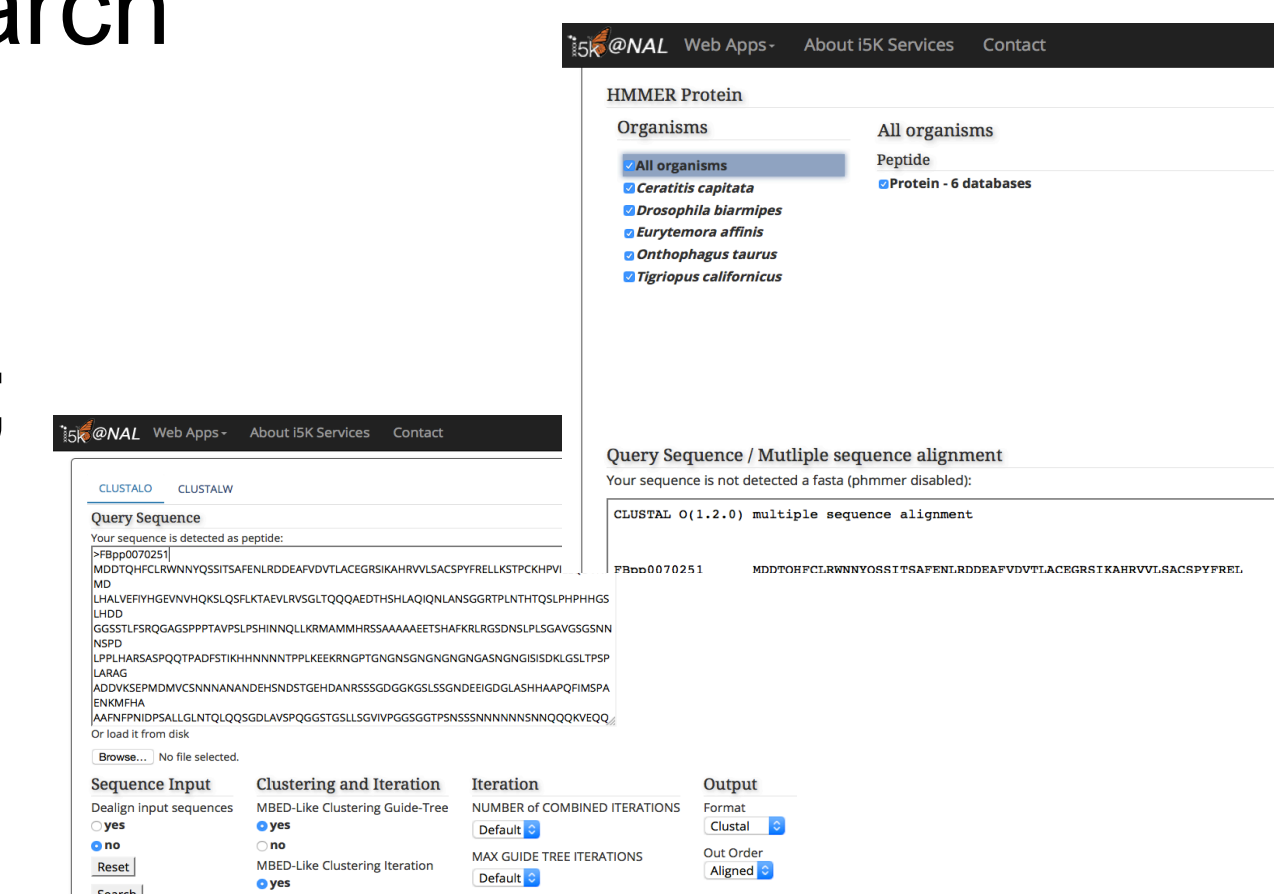
- A workspace for **genomic data access, dissemination, and curation** for **arthropods**, hosted by the USDA's National Agricultural Library (NAL)².
- We currently host genome project data for **53 arthropod species**.
- We provide a central **organism page** for each project, **gene pages** for projects 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**.



Updates for 2016

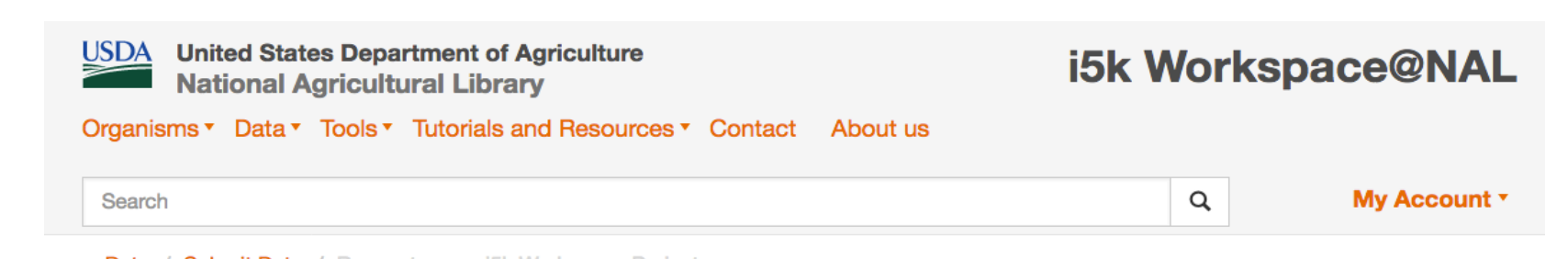
New web applications

- 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.



Request a new i5k Workspace Project

Thank you for your interest in submitting your genome project to the i5k Workspace! Please answer the following questions to help us decide if the resources at the i5k Workspace are a good fit for your project.

Genus *

- Sign up for an account: <https://i5k.nal.usda.gov/register/project-dataset/account>

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.

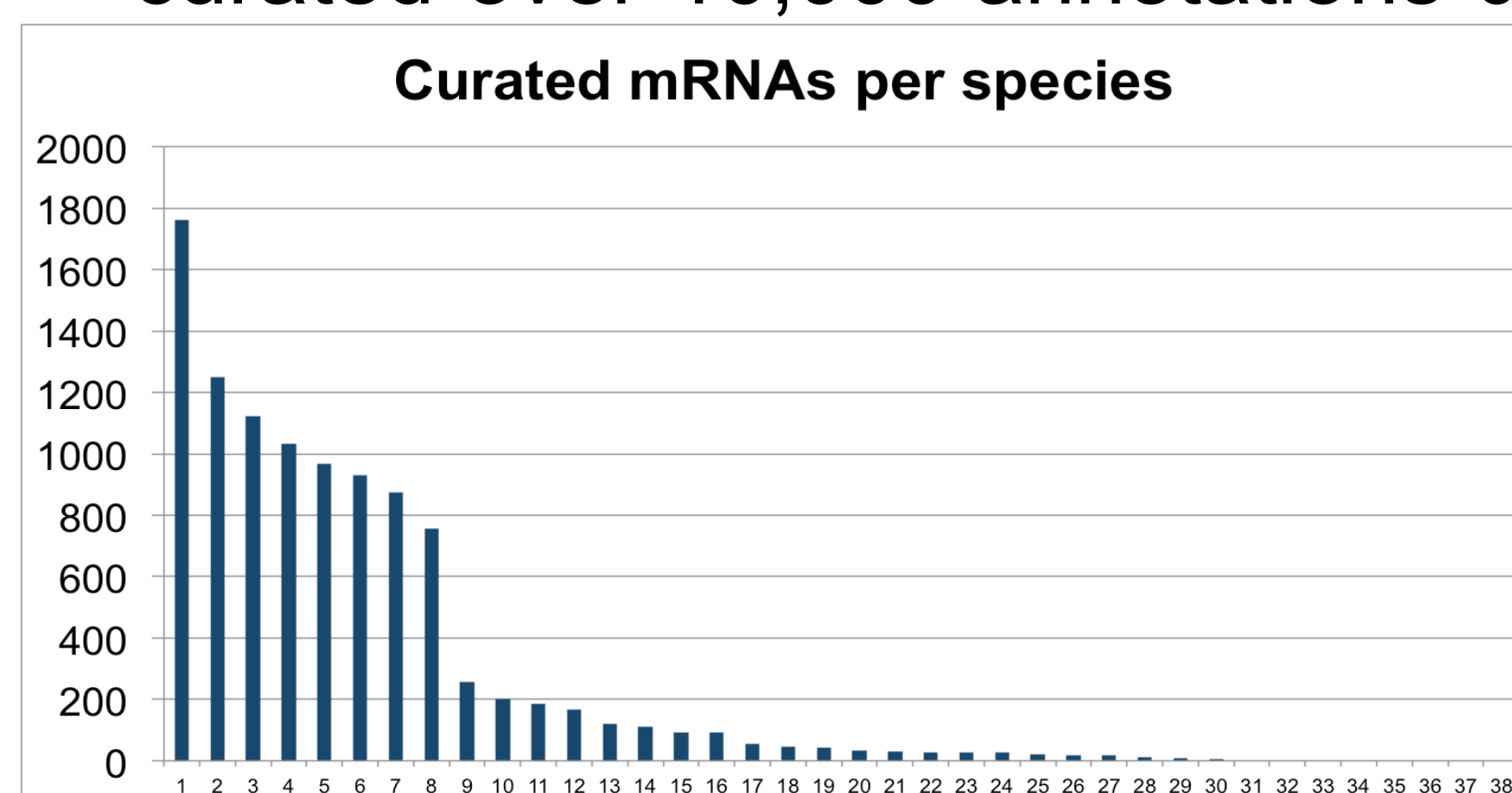
1. **Help annotate** an existing genome project.
 - Over 400 annotators curating over 30 genome projects
 - Register for an Apollo account: <https://i5k.nal.usda.gov/web-apollo-registration>
2. **Add data** to an existing genome project.
 - Browse our list of existing projects: <https://i5k.nal.usda.gov/species>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

New organisms and datasets

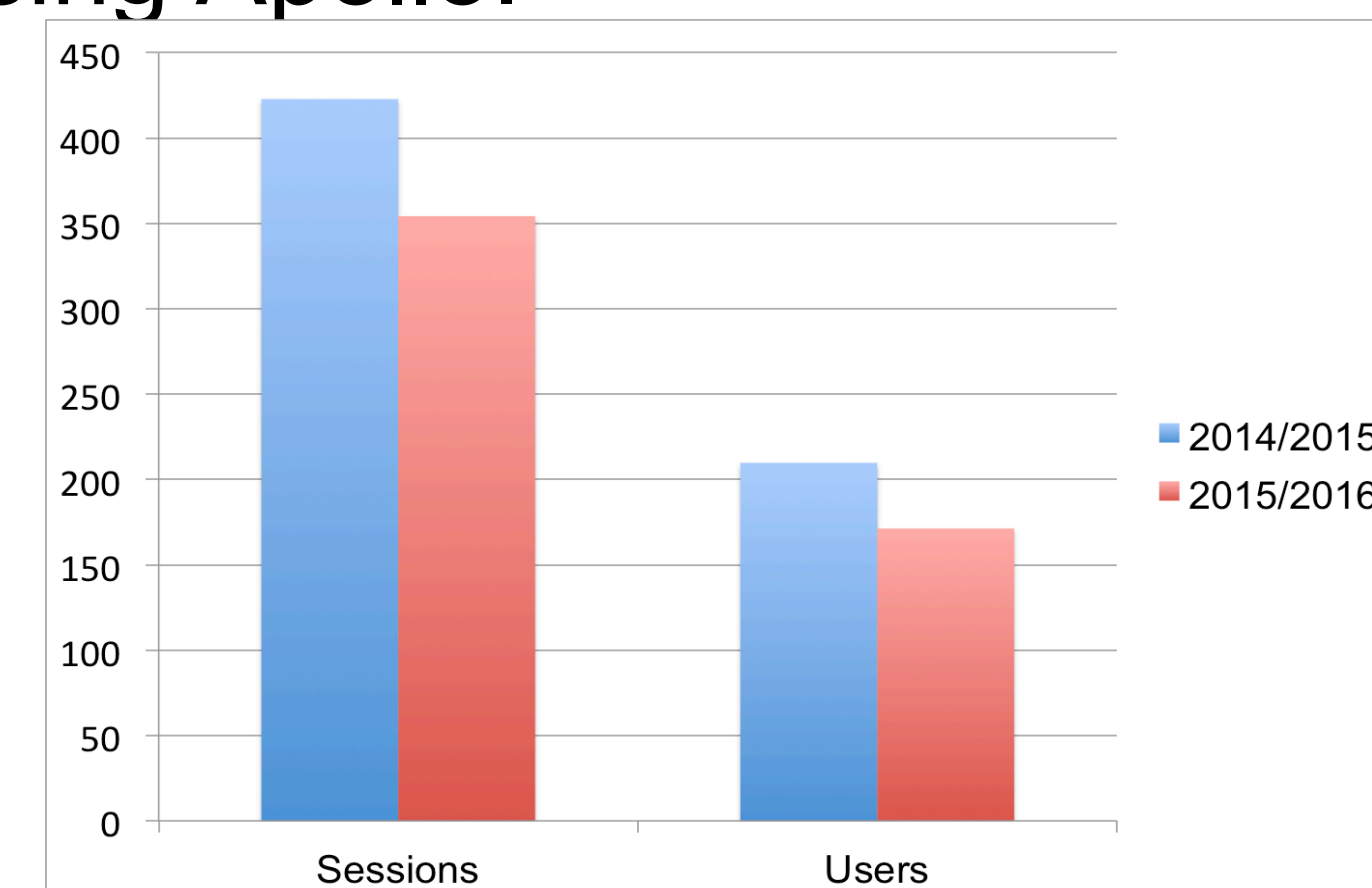
- New species for 2016 include *Dufourea novaeangliae*, *Habropoda laboriosa*, *Lasioglossum albipes*, *Megachile rotundata*, *Melipona quadrifasciata*, *Microplitis demolitor*, *Diachasma alloeum*, *Neodiprion lecontei*, *Amyelois transitella*, and *Tribolium castaneum*.
- Official Gene Set pipeline 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.

How is the i5k Workspace being used?

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



Number of curated mRNAs by species, demonstrating high variation in curation activity among species.



Average number of sessions or users per week in 2014/2015 vs 2015/2016.

For more information about i5k

- Visit us online: <https://i5k.github.io>
- Visit us at Booth #320.

References

1. i5K Consortium (2013) The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. *J. Hered.*, **104**, 595–600.
2. Poelchau, M.F., et al. (2014) The i5k Workspace@NAL – enabling genomic data access, visualization, and curation of arthropod genomes. *Nucl. Acids Res.* doi:10.1093/nar/gku983
3. Camacho, C., et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
4. Skinner, M.E., et al. (2009) JBrowse: A next-generation genome browser. *Genome Res.*, **19**, 1630–1638.
5. Lee, E., et al. (2013) Web Apollo: a web-based genomic annotation editing platform. *Genome Biol.*, **14**, R93.
6. Ficklin, S.P., et al. (2011) Tripal: a construction Toolkit for Online Genome Databases. *Database*: bar044.
7. Larkin, M.A., et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:21: 2947–2948.
8. Sievers, F., et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**:539
9. Eddy, S.R. (2009) A new generation of homology search tools based on probabilistic inference. *Genome Informatics* **23**(1):205–11.

Visit us online
Web: <https://i5k.nal.usda.gov>
Email: i5k@ars.usda.gov

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.