



The i5k Workspace@NAL - Enabling Genomic Data Access, Visualization, and Curation of Arthropod Genomes



Monica Poelchau¹, Christopher Childers¹, Gary Moore¹, Vijaya Tsavatapalli¹, Jay D. Evans², Chien-Yueh Lee³, Han Lin³, Jun-Wei Lin⁴ and Kevin Hackett⁵

¹USDA/Agricultural Research Service/National Agricultural Library, Beltsville, MD, ²USDAARS Bee Research Lab, Beltsville, MD, ³Graduate Institute of Biomedical Electronics and Bioinformatics, Taipei, Taiwan, ⁴Graduate Institute of Electrical Engineering, Taipei, Taiwan, ⁵USDA-ARS, Beltsville, MD

What is the i5k initiative?

- The **5,000 arthropod genomes initiative (i5k)** coordinates the sequencing of 5,000 insect or related arthropod genomes¹.
- International effort to seek funding from academia, governments, industry, and private sources; prioritize insect genomes for sequencing; develop best practices for genome sequencing and curation.

What is the i5k Workspace@NAL?












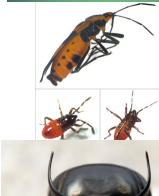

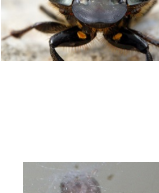

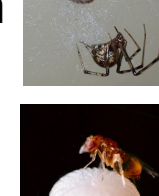

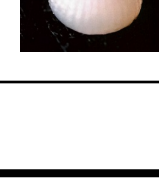


- A workspace for **genomic data access, dissemination, and curation** for any ‘orphaned’ arthropod genome project, hosted by the USDA’s National Agricultural Library (NAL)².
- Geared towards both **data producers** and **data consumers**:
 - Data producers**: we are actively soliciting new genome projects, in particular from groups with limited resources for genome hosting and curation.
 - Data consumers**: we aim to provide up-to-date genomic content from a variety of arthropod genome projects.
- URL: <https://i5k.nal.usda.gov>

How do I submit my data?

- Contact us (i5k@ars.usda.gov) for individual consultation about your genome project. In general, we require:
 - A ‘**frozen**’ genome assembly, preferably already submitted to NCBI.
 - Metadata about your assembly (contact us for the most current submission form).
- We host any other data mapped to the genome assembly, e.g.
 - Computational gene predictions
 - Transcriptomes
 - RNA-Seq data

Current content

39 arthropod species, many from the i5k pilot project³:

Latin name	Common name	Latin name	Common name
<i>Agilus planipennis</i>	 Emerald Ashborer Beetle	<i>Frankliniella occidentalis</i>	 Wester flower thrips
<i>Anoplophora glabripennis</i>	 Asian long-horned beetle	<i>Gerris buenoi</i>	 Water Strider
<i>Athalia rosae</i>	 Turnip sawfly	<i>Halyomorpha halys</i>	 Brown marmorated stink bug
<i>Blattella germanica</i>	 German Cockroach	<i>Homalodisca vitripennis</i>	 Glassy-winged sharpshooter
<i>Cataglyphis aquilonaris</i>	 Silvestri's Northern Forceptail	<i>Hyalella azteca</i>	 Amphipod
<i>Centruroides exilicauda</i>	 Bark scorpion	<i>Ladona fulva</i>	 Scarce Chaser
<i>Ceratitis capitata</i>	 Mediterranean fruit fly	<i>Latrodectus hesperus</i>	 Western black widow spider
<i>Cimex lectularius</i>	 Bed bug	<i>Leptinotarsa decemlineata</i>	 Colorado potato beetle
<i>Copidosoma floridanum</i>	 Parasitic Wasp	<i>Limnephilus lunatus</i>	 Caddis fly
<i>Diaphorina citri</i>	 Asian Citrus Psyllid	<i>Loxosceles reclusa</i>	 Brown recluse spider
<i>Drosophila biarmipes</i>	NA	<i>Manduca sexta</i>	Tobacco hornworm
<i>Drosophila bipunctinata</i>	NA	<i>Mayetiola destructor</i>	Hessian fly
<i>Drosophila elegans</i>	NA	<i>Oncopeltus fasciatus</i>	Milkweed Bug
<i>Drosophila eugracilis</i>	NA	<i>Onthophagus taurus</i>	Bull-headed Dung beetle
<i>Drosophila ficusphila</i>	NA	<i>Orussus abietinus</i>	Parasitic wood wasp
<i>Drosophila kikkawai</i>	NA	<i>Pachypsylla venusta</i>	Hackberry petiole gall psyllid
<i>Drosophila rhopaloea</i>	NA	<i>Parasteatoda tepidariorum</i>	Common house spider
<i>Drosophila takahashii</i>	NA	<i>Tigriopus californicus</i>	Harpacticoid copepod
<i>Ephemera danica</i>	Mayfly	<i>Trichogramma pretiosum</i>	Parasitic wasp
<i>Eurytemora affinis</i>	Common Copepod		

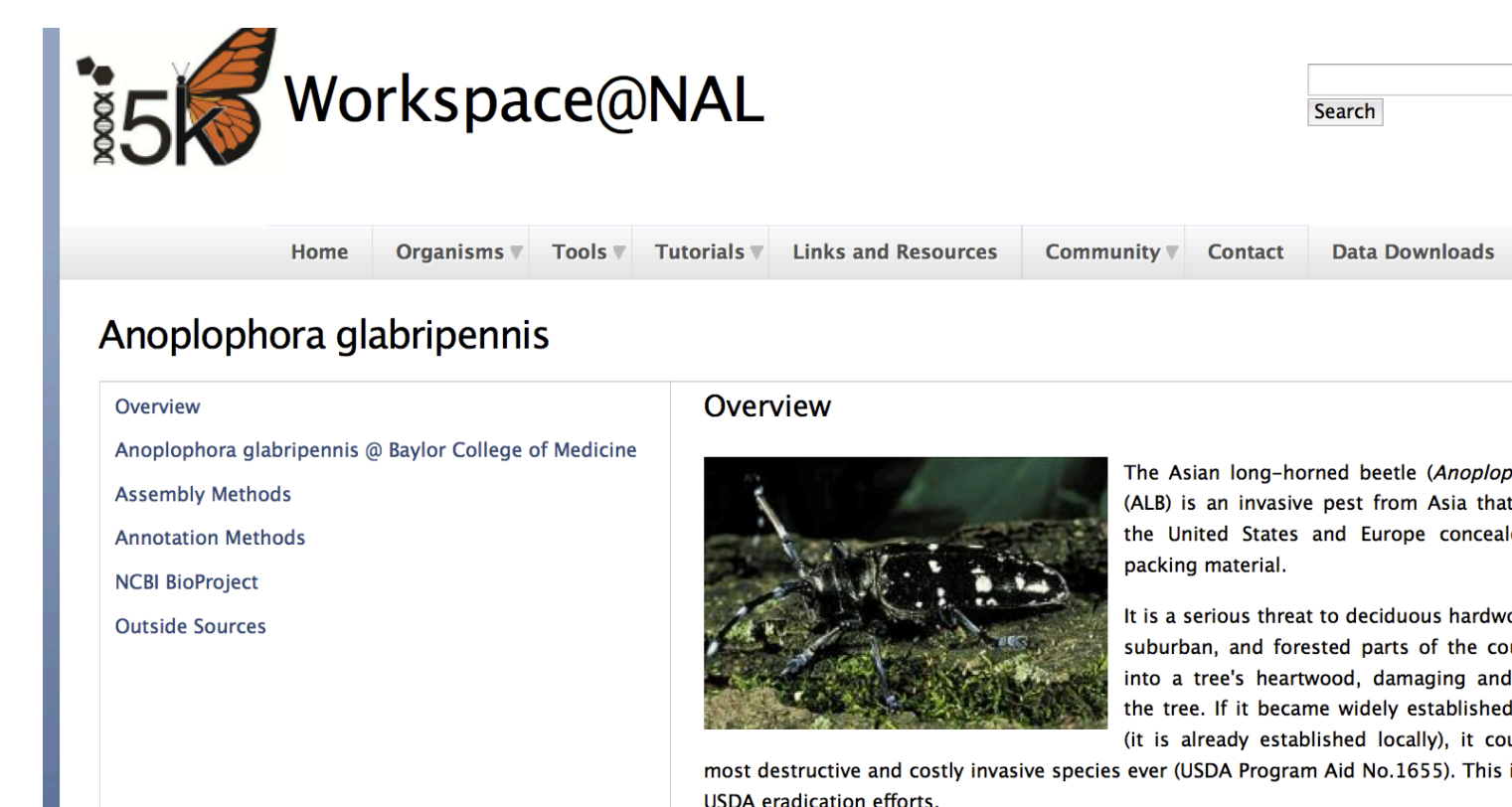
URL: <https://i5k.nal.usda.gov>

Contact: i5k@ars.usda.gov

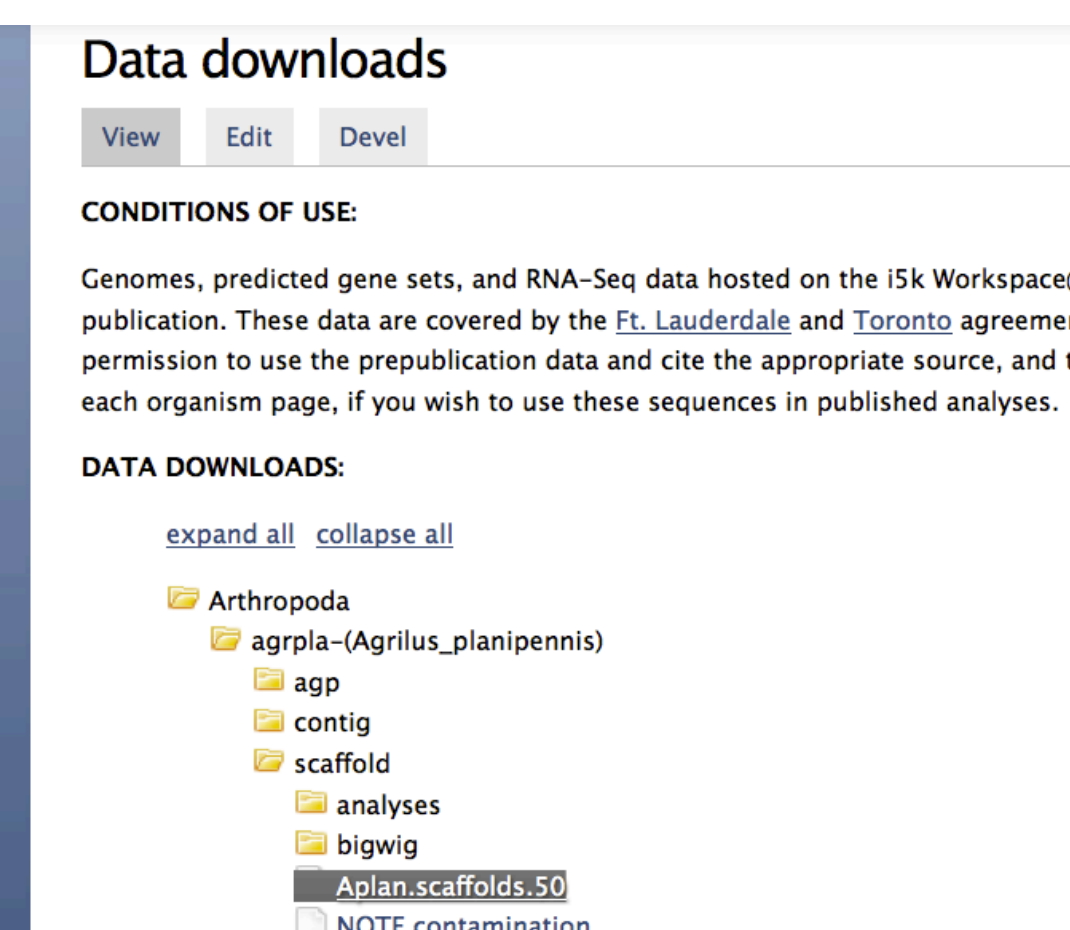
What resources and tools does the i5k Workspace provide?

Resources

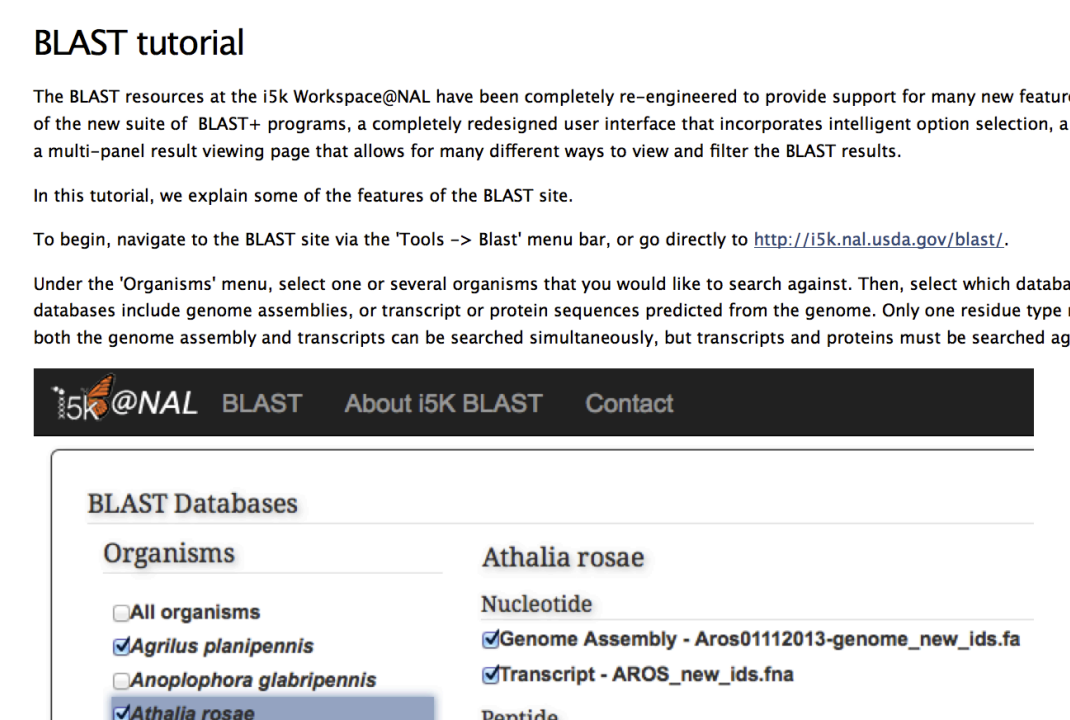
Organism home page



Data downloads

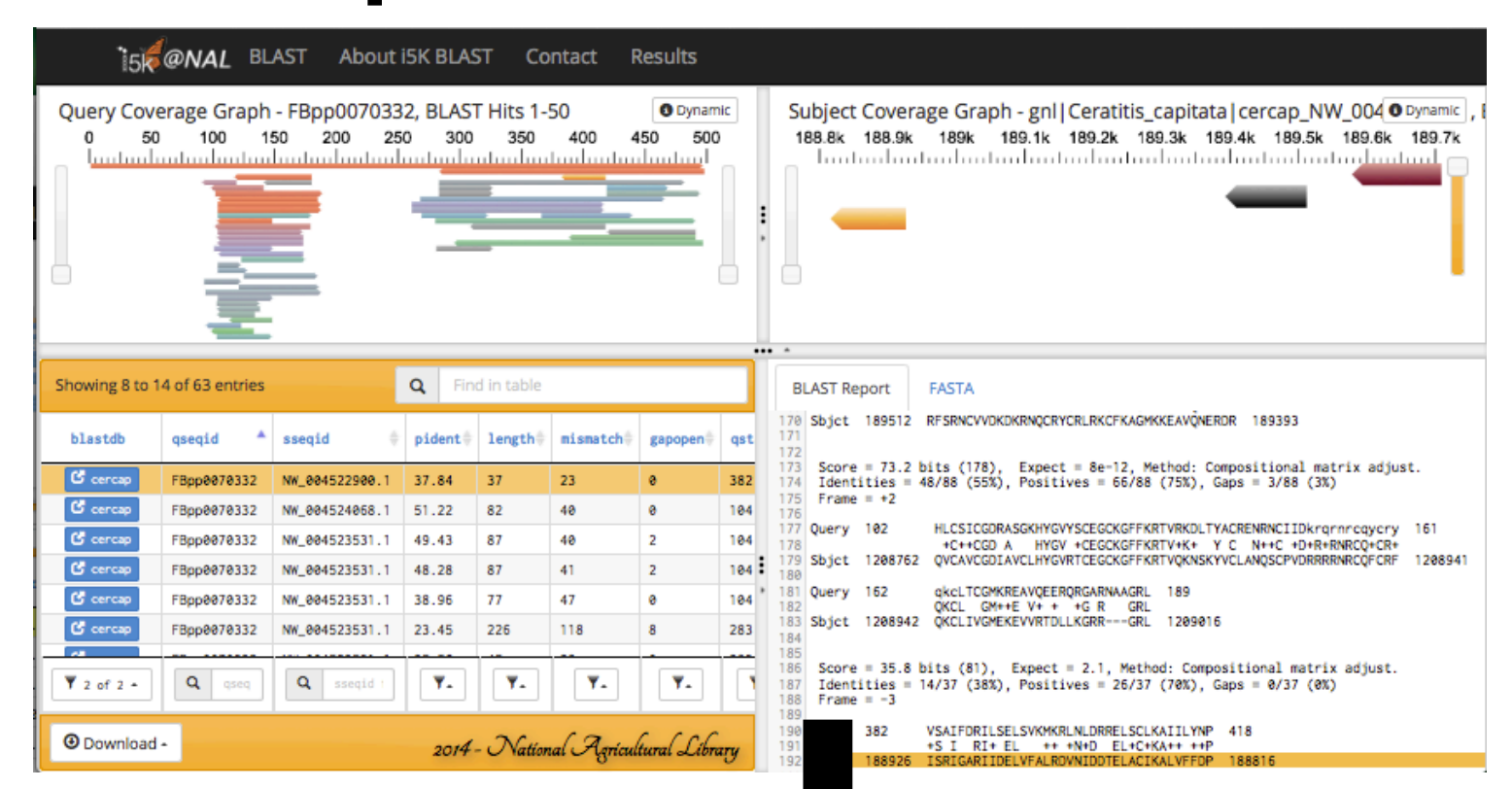


Tutorials

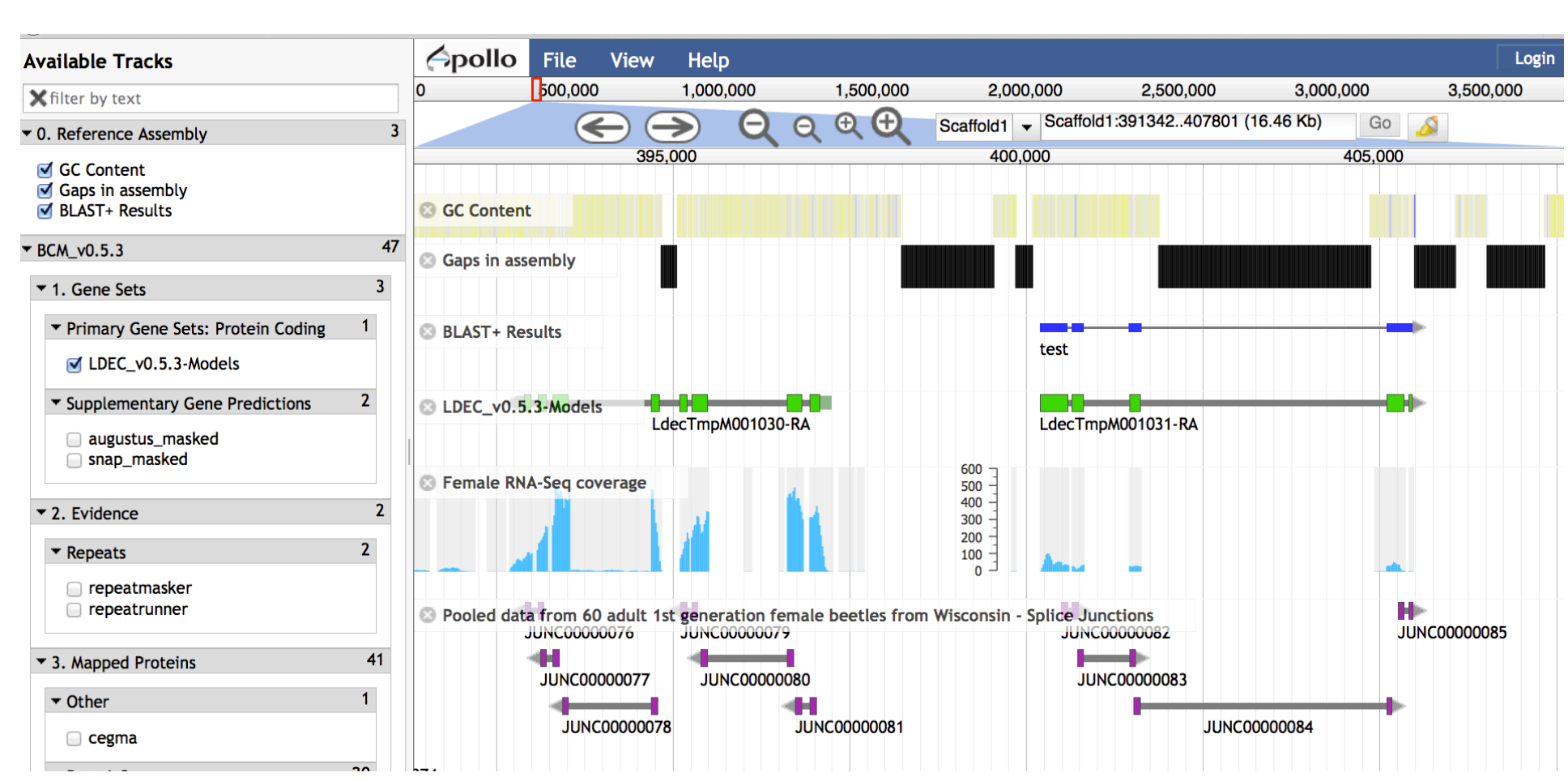


Tools

Unique BLAST+⁴ interface



Jbrowse⁵ Genome Browser



Web Apollo⁶ manual curation tool



Individual help and consultations for data producers

Upcoming Developments

- New interface;
- Gene pages for assemblies with an official gene set (OGS);
- Improved search tools (e.g. InterMine).

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.

References

- i5k Consortium (2013) The i5k Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. *J. Hered.*, **104**, 595–600.
- Poelchau, MF, *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
- <https://www.hgsc.bcm.edu/arthropods/i5k-pilot-project-summary>
- Camacho, C., *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
- Skinner, M.E., *et al.* (2009) JBrowse: A next-generation genome browser. *Genome Res.*, **19**, 1630–1638.
- Lee, E., *et al.* (2013) Web Apollo: a web-based genomic annotation editing platform. *Genome Biol.*, **14**, R93.