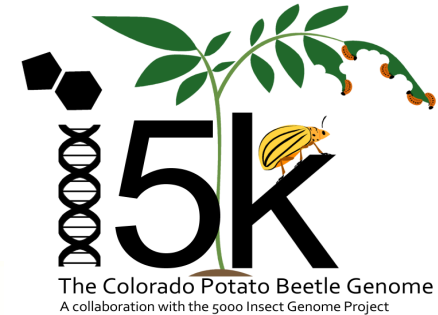# i5K Initiative

Current efforts:

- ◉ i5K Pilot: NIH/NHGRI
- ◉ Ag100Pests: USDA-ARS
- ◉ Cornome: USDA-ARS, Smithsonian
- ◉ i5K Workspace@NAL: USDA-ARS
- ◉ Aphid Plus Project: INRA
- ◉ Ants (GAGA): Chinese Academy of Sciences
- ◉ Bees and Wasps – Roots of Sociality: China
- ◉ Beetles: U. Memphis FedEx Institute, NSF

The Colorado Potato Beetle Genome
A collaboration with the 5000 Insect Genome Project

The Asian Longhorned Beetle Genome
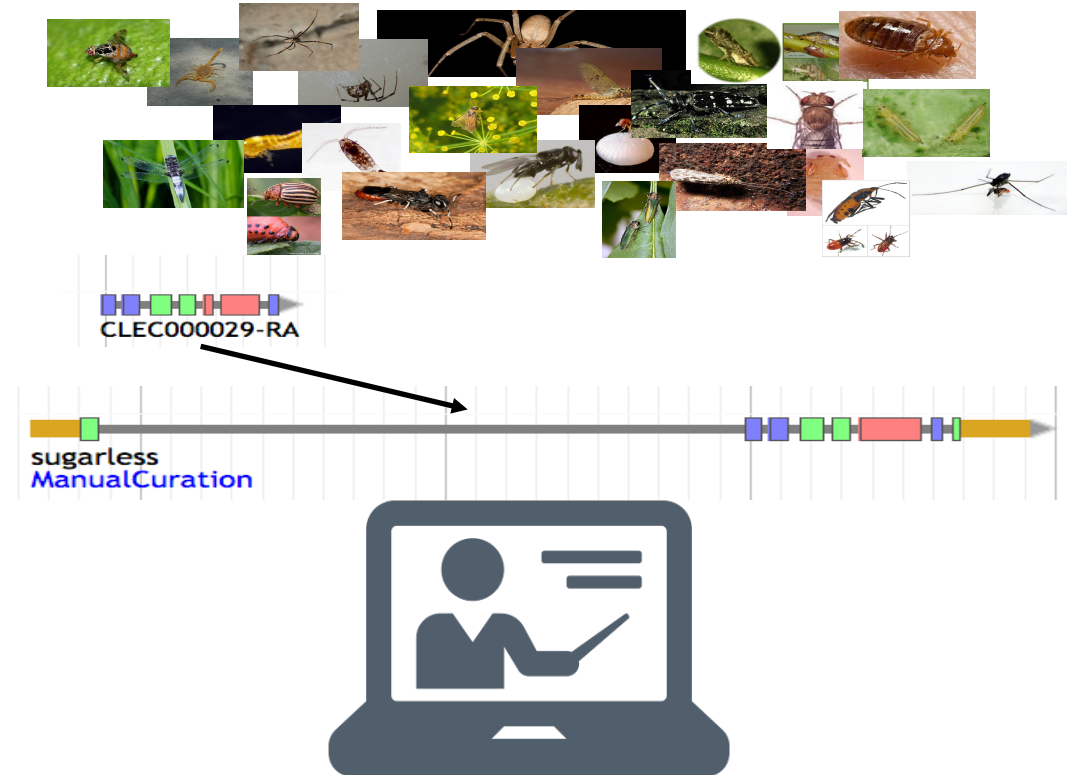A collaboration with the 5000 Insect Genome Project

The i5k Workspace@NAL

i5k.nal.usda.gov

Provides access to **70** arthropod genome projects and counting

Facilitates community-driven, manual gene annotation **curation** of over **15,000** gene models

Conducts webinars, tutorials, and training for the i5k community

CLEC000029-RA

sugarless
ManualCuration

i5k

USDA Agricultural Research Service

Cornome Project: Genomics of Beneficial Insects in the Maize Agroecosystem

USDA-ARS and Smithsonian collaboration

- **Aims and goals**
  - Sample beneficial non-target arthropods common in maize fields
  - Scan genome assembly with novel algorithms for putative RNAi targets
  - Identify microbial associates

- **Genome sequencing and assembly**
  1. Spined soldier bug, *Podisus maculiventris* - PacBio Sequel; Oxford Nanopore; Illumina HiSeq
  2. Green lacewing, *Chrysoperla rufilabris* - PacBio Sequel; Oxford Nanopore; Illumina HiSeq
  3. Ladybird beetle, *Coleomegilla maculata* - PacBio Sequel; Oxford Nanopore; Illumina HiSeq
  4. Parasitoid wasp, *Macrocentus grandii* - Illumina HiSeq; PacBio Sequel planned
  5. Minute pirate bug, *Orius insidiosus* - PacBio Sequel; Illumina HiSeq
  6. Rove beetle, *Dalotia coriaria* - PacBio Sequel; Illumina HiSeq
  7. Squash bee, *Peponapis* - PacBio Sequel; Illumina HiSeq
  8. Parasitoid wasp, *Trichogramma* - Illumina HiSeq
  9. Parasitoid wasp, *Aphelinus* - planning stages
  10. Caddisfly, Psychomyia flavida - done by Smithsonian
  11. Harvestman, *Phalangium opilio* - done by Smithsonian

- **Applications**
  - Predict risks of RNAi to non-target arthropods at molecular level
  - Determine resistance and susceptibility to crop protection products
  - Identify subspecies or races with improved biocontrol potential
  - Identify microbial associates that might be vectored to pests

Ag100Pest Initiative

Goal: generate high quality genome assemblies and annotations for top 100 US agricultural arthropod pests

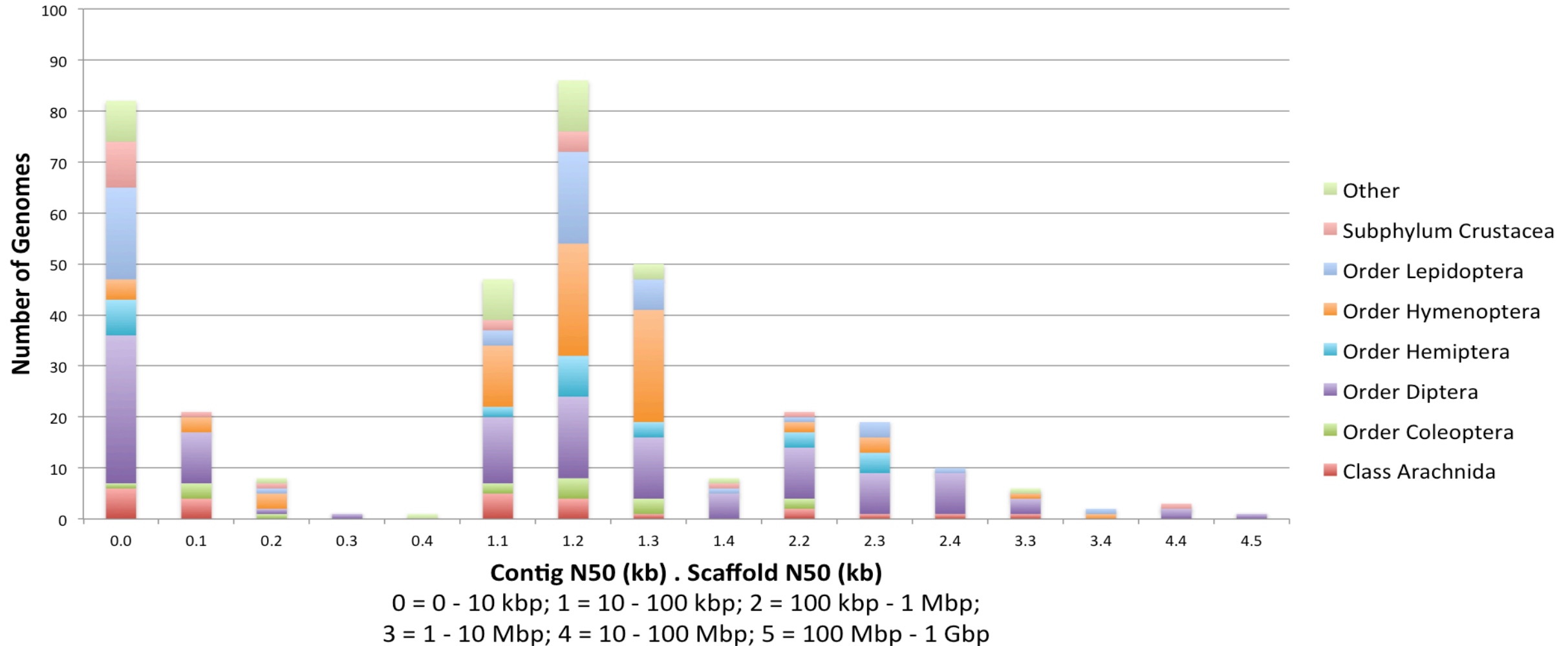USDA-ARS's commitment and contribution to the Earth BioGenome Project

EARTH
BIOGENOME
PROJECT
sequencing life for the future of life

USDA Agricultural Research Service

i5K

Ag100Pest Leadership

◎ **Executive Coordination Team**: Kevin Hackett, Anna Childers and Brian Scheffler

◎ **Core Leadership Group**: Kevin Hackett, Anna Childers, Brian Scheffler, Brad Coates, Scott Geib, Monica Poelchau and Chris Childers

◎ **Prioritization Team Lead**: Brad Coates
◎ **Extraction/Pre-Sequencing Team Lead**: Scott Geib
◎ **Sequencing Team Leads**: Brian Scheffler and Tim Smith
◎ **Assembly and Analysis Team Lead**: Anna Childers
◎ **Post-Assembly Team Leads**: Monica Poelchau and Chris Childers

Arthropoda Genome Quality (n = 366)

X-axis: **Contig N50 (kb) . Scaffold N50 (kb)**
0 = 0 - 10 kbp; 1 = 10 - 100 kbp; 2 = 100 kbp - 1 Mbp;
3 = 1 - 10 Mbp; 4 = 10 - 100 Mbp; 5 = 100 Mbp - 1 Gbp

Y-axis: **Number of Genomes**

Legend:
- Other
- Subphylum Crustacea
- Order Lepidoptera
- Order Hymenoptera
- Order Hemiptera
- Order Diptera
- Order Coleoptera
- Class Arachnida

*as of October 2018

"Reference Quality Genomes"

Aim to meet EBP
2.3.2Qv40 standard

PacBio long-read
sequencing where
possible

Scaffolding for top
priority species as
possible

Supplementary Figure S1. Proposed metric system for describing genome assemblies. The system avoids classifications such as draft, complete, platinum, gold, etc., whose definitions may change over time due to technical advances.

x.y.z score for each assembly

10^x N50 contig size in kb

10^y N50 scaffold size in kb

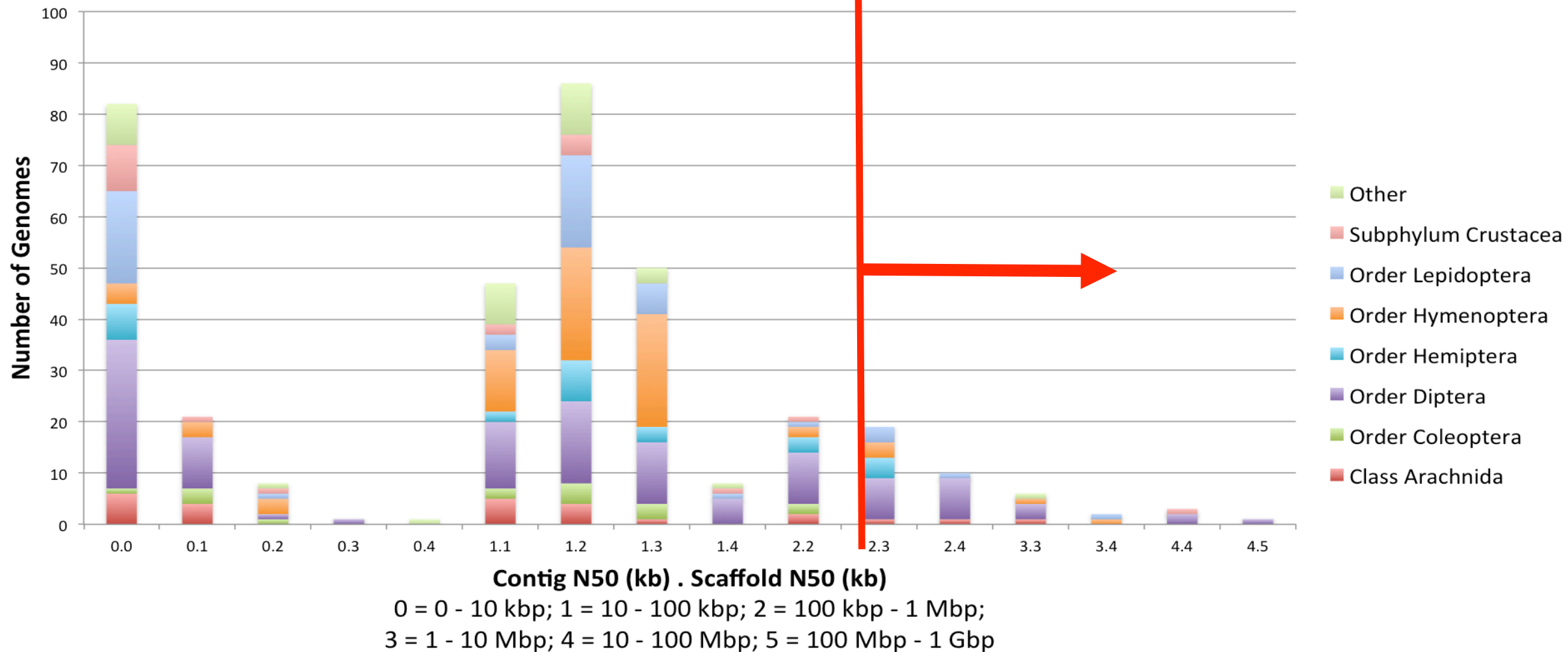**z**   chromosome map status as at right

e.g.

3.4.3   contig > 1Mb, scaffold >10Mb placed on in-species map

1.3.1   contig > 10kb, scaffold >1Mb aligned to inferred ancestral chromosomes
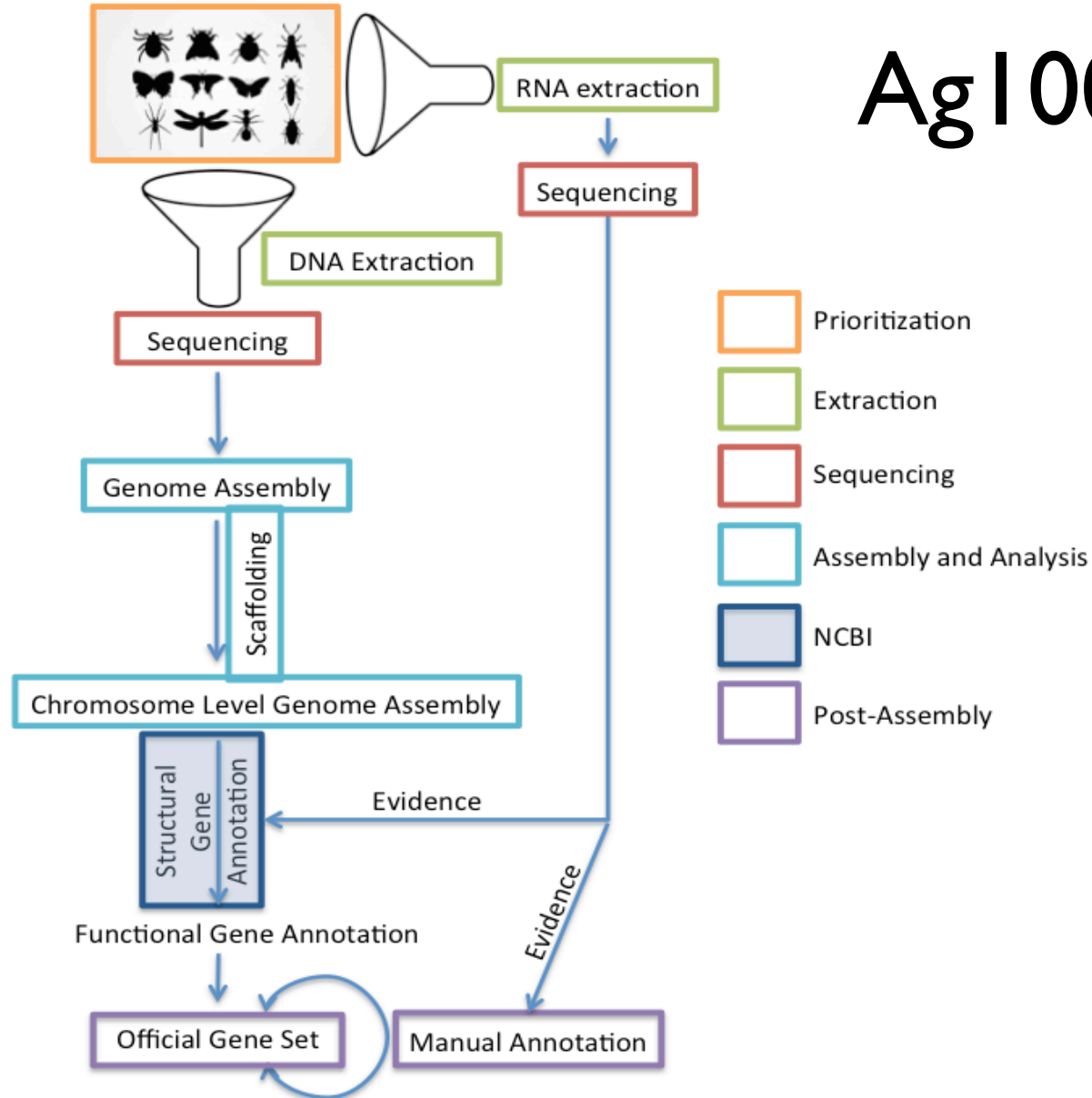
z   Chromosome assignment measure

0. Anything below 1.
1. >90% DNA assigned by *in silico* matching to reference or inferred ancestral full length chromosomes
2. + all interchromosomal rearrangements (fusions, fissions, translocations) validated by >2 data sources (if possible experimental)
3. + >80% directly assigned to within-species maps, plus experimental validation of all intrachromosomal and representative interchromosomal (deletions, insertions, inversions, transpositions) breakpoints

Arthropoda Genome Quality (n = 366)

Number of Genomes (y-axis, 0–100)

Contig N50 (kb) . Scaffold N50 (kb) (x-axis)

0 = 0 - 10 kbp; 1 = 10 - 100 kbp; 2 = 100 kbp - 1 Mbp;
3 = 1 - 10 Mbp; 4 = 10 - 100 Mbp; 5 = 100 Mbp - 1 Gbp

Legend:
- Other
- Subphylum Crustacea
- Order Lepidoptera
- Order Hymenoptera
- Order Hemiptera
- Order Diptera
- Order Coleoptera
- Class Arachnida

*as of October 2018

Ag100Pest Project Workflow

Long-read sequencing on single individuals when feasible

RNAseq to support gene prediction and manual annotation

Functional annotation pipeline tuned to arthropods (in development)

Data hosted at i5k Workspace@NAL

Ag100Pest Species Selection

Spotted Lanternfly

◎ Status: selecting top pests from ~400 nominations across at least 10 Orders

◎ Nominations gathered from:

   ◎ APHIS

   ◎ Federal Interagency Committee on Invasive Terrestrial Animals and Pathogens (ITAP)

   ◎ Cooperative Agricultural Pest Survey (CAPS)

   ◎ USDA researchers

   ◎ i5K community

◎ Diverse nominations include pests of US field crops, animals, bees, forests, and stored products

◎ Nominations include some foreign pest species considered high invasive threats to US Ag

EARTH
BIOGENOME
PROJECT
sequencing life for the future of life

USDA Agricultural Research Service

i5K

Goals Beyond Genomes

◎ Genomes as infrastructure for larger research questions

◎ Generation of best practices and decision trees for various stages including extraction, sequencing design, assembly, and annotation

   ◎ Realized within publications as well as scripts and documentation

◎ Capacity and expertise building within ARS

◎ Community building within and outside ARS

EARTH
BIOGENOME
PROJECT
sequencing life for the future of life

USDA Agricultural Research Service

i5K

Principal Investigators / Collaborators

i5K Pilot: S. Richards (Baylor College of Medicine and UC-Davis) and many others

i5K Workspace@NAL: C. Childers and M. Poelchau (USDA-ARS)

Cornome: J. Evans, B. Coates, A. Childers, K. Shelby and K. Hackett (USDA-ARS); D. Hawthorne (U. MD); J. Coddington (Smithsonian)

Ag100Pest: K. Hackett, A. Childers, B. Scheffler, B. Coates, S. Geib, C. Childers and M. Poelchau (USDA-ARS)