# Questions for break-out discussion groups

**Concurrent Sessions**

The objectives of the concurrent sessions are to:

- Give ARS scientists a venue to discuss issues related to arthropod genomics research in ARS;
- Identify key goals and bottlenecks;
- Develop strategies to overcome constraints and better meet ARS mission goals.

There are three sessions. Workshop participants will break out into 5 groups for each session, each led by one member of the workshop committee. Each group will discuss the questions and topics listed in this document. At the end of each session, each group will articulate a vision or strategy for the session topic. Group leaders will compile notes from each session into a final document, which will serve as a memo for ARS administrators to inform ARS policy.

**Concurrent sessions-I: Defining projects, questions, goals and issues**
**Questions/Topics**

I.A What are the motivating factors, scientific questions, and project objectives that are being addressed?

- What types of data are required to address these goals? How will these data be acquired?
- What are the expected outputs and outcomes of research?

I.B What distinguishes arthropod genomics research in ARS from other similar research initiatives in ARS? Define how ARS arthropod genomics needs differ from those of other focus areas in ARS (such as compared to crop and livestock genomics research).

I.C Who are the Stakeholders in ARS arthropod genomics research?

- How will ARS arthropod genomics data and outputs be accessed for the benefit of s takeholders?
- In terms of meta-data, accompanying data, and/or data integration, what value added properties can/should accompany ARS arthropod genomics research to increase visibility, usage, and impact?
- Should ARS maintain topical databases and tools and if so for which taxa or research goals?

I.D How can ARS and the community of SYs address current gaps?

- Can shared resources, expertise or group knowledge be leveraged to address bottlenecks? If so, how can community resources be implemented in the short term?

- How will recurrent training, visiting scientists or internal/external collaborations assist in achieving immediate goals?

**Concurrent sessions-IIA: Genomic data analyses: where, who, and how.**

**Sessions IIA and IIB represent the major part of the break-out discussions. Having identified major questions and current gaps, we know want to focus on some issues in more detail, on solutions and on the "where, who and how". The 'where' of this session should discuss ideas around where bioinformatics analyses should be performed (e.g., on local machines, on public service platforms such as iPlant, or on a cloud-based service such as the proposed ARS scientific DMZ). There need not be a one-size-fits-all approach; scientists should discuss what may work best for different experience levels and for different project needs. It would also be useful to propose the model of analysis 'silos' vs. shared resources. The "who" and "how" will focus on who should be performing the bioinformatics analyses (ARS scientists themselves, or trained bioinformatics specialists, or a combination of both), who receives bioinformatics training (and what kinds of training are necessary) and how will we accomplish our larger goals. Then, the discussion will focus on data standards within the context of data management: what standards are important and when, and what are the obstacles in implementing them?**

**Questions and topics.**

<u>IIA.A Where</u>

- Where should bioinformatics analyses be performed - for example, in individual labs or 'silos', using shared ARS resources, or in external, cloud-based platforms (e.g. iPlant)? Are there other models?

<u>IIA.B Who</u>

- What is our vision and model for bioinformatics personnel resources in ARS? For example, should we train existing scientists, hire new individuals, or do we need a combination of both training and hiring?
- If individuals are to be hired, should they be located in one (or more) core bioinformatics groups? Or, should specialists embedded in research groups?
- If existing scientists should be trained, what are their training needs? Define key knowledge domains needed to further genomics research. (Some possible examples: RNAseq analysis/assembly, metagenomic analysis, GWAS, genome assembly/analysis).

<u>IIA.C How</u>

Within-ARS support

- How do we communicate our cyber-infrastructure needs and determine if they are shared with other groups, or unique to us?
- How do we ensure support is available across the ARS, and not just for individual units?
- Should there be resources dedicated to Arthropod research?
- What specific services would we like to have included in the Big Data initiative that are likely to be overlooked?
- How do you perceive the role of the NAL? What data access issues have you encountered that are not served by existing resources (e.g. NCBI, iPlant) that the NAL may be able to provide?

External partnerships

- How can we build stronger cyber-infrastructure collaborations within and outside ARS (e.g., partnerships for long-term resource development)?
- How can we integrate ARS efforts with larger efforts outside ARS to achieve greatest impact and visibility, and accomplish goals more efficiently and effectively?

**Concurrent sessions IIB: Genomic data analyses: where, who, and how.**

**Data management and data standards. Why is data management important, and what data management challenges does AGR in ARS face?**

Our goals for this section are to: 1) get feedback on what kind of data management challenges ARS scientists face; 2) identify the most common workflows that ARS scientists use in AGR, and discuss whether these can and should be standardized; 3) understand what kinds of data are being produced that need to be accessible beyond the life cycle of a project; and 4) identify what types of metadata ARS scientists deem particularly important for their data, but also for publically available data that they re-use.

**Questions and topics.**

IIB.A Data management

- Do you have a data management plan? Why or why not?
- Are there particular obstacles in place that make it difficult to develop or implement a data management plan?

IIB.B Workflow standardization

- At what levels in a bioinformatics workflow should data outputs and methods be standardized?
- What aspects require more flexibility?
- Would it be useful or constraining to write up standards for commonly-used bioinformatics workflows?
- Would it be useful or feasible to create a toolkit, or create/use an existing VM image that has commonly used tools?

IIB.C Data standards: data formats and metadata

- Have you identified specific cases where you need to standardize your data? How did this help you?
- How can we improve data reusability/accessibility?
- Online repositories such as NCBI Short Read Archive, GigaDB, or the Nature Publishing Group Scientific Data site recommend or require metadata to accompany data submissions. What standards do you currently use to capture metadata?
- What standards or guides would you like to have for defining and retaining metadata?
- Are there particular issues related to data formats and metadata that are specific to arthropod genomics?

- What problems are there in enforcing data standards (e.g. Submission to NCBI, documenting analysis pipelines)?
- Should there be a data exchange or knowledgebase to share protocols or project data specific to the ARS?
    - What form would you like to see this take?
    - How would this be different than public resources?

**Concurrent sessions-III**
**Common tools and approaches, issues and constraints, and solutions and vision for the future.**

For this section, there will be three groups broken up by subject/application area. Each group will address the same set of questions, but with the focus on their subject area.

**Group I:** full genome sequencing group
**Group II:** population and evolutionary genomics group
**Group III:** functional genomics group-RNA-Seq/gene expression/RNAi/CRISPR

**Questions and topics.**

III.A Current approaches

- What genomics and bioinformatics tools or approaches are being used to address current research questions? How does application of these tools or approaches advance US agriculture?
- Are these tool or approaches sufficient to fully and efficiently address mission goals?
- How can we ensure that ARS researchers are using the right tools for their needs?
- How can we work together as a group to accomplish common goals and ensure as a group that we are on the leading edge of science?

III.B Current Issues and Constraints

- Which technologies, informatics pipelines or expertise are currently lacking for ARS arthropods genomics researchers? What factor(s) prevent genomics research from being successfully completed?
- If genomics approaches and analysis pipelines are NOT readily accessible to ARS researchers, what technical or bioinformatic challenges prevent this from happening? Prioritize the needed technologies, pipelines or expertise that are seen as currently lacking in ARS.
- What changes in ARS are needed to overcome the above stated constraints?

III.C Future Vision

- How can ARS arthropod genomics adapt to new innovations and seamlessly adjust strategies to expedite research output?
- How can HQ, NPO, and Area Offices help with genomics research projects? How can SYs assist themselves and each other?

- What political avenues (Federal or State) or Stakeholder-mediated directives might by relied upon to support and increase awareness of positive impacts obtained from ARS arthropod genomics research?
- What mission-driven vision can each group identify for the next 5-10 years?
- How will genomic needs fit with projected funding?