

PAG XXI Workshop

The International Arabidopsis Informatics Consortium (IAIC) and development of the Arabidopsis Information Portal (AIP)



Co-organized by:

Blake C. Meyers, IAIC Interim Director

Chris Town, AIP Principal Investigator

Jan. 14, 2013

We're on the web at:

www.arabidopsisinformatics.org

The agenda for today



Part I: progress on the AIP and core components

12:50 pm: *Workshop Introduction and Overview:*

The International Arabidopsis Informatics Consortium

Blake C. Meyers, University of Delaware

The AIP Scientific Advisory Board

Mark Estelle (SAB chair), University of California, San Diego

1:00 pm: *Plans for the Arabidopsis Information Portal*

Chris Town, The J. Craig Venter Institute

1:20 pm: *Transitioning TAIR to iPlant*

Bob Muller, Carnegie Institution for Science

The agenda for today



Part II: Community-contributed modules for the Arabidopsis Information Portal: What will these look like?

1:40 pm: *Informatics Modules from BAR*

Sylva Donaldson, University of Toronto

2:00 pm: *Arabidopsis Phenotype Representation Using Ontologies*

Eva Huala, Carnegie Institution for Science

2:20 pm: *Arabidopsis Community Resources For Comparative and Network Analyses*

Doreen Ware, Cold Spring Harbor Laboratory

Gramene Consortium, KBase Consortium

2:40 pm: *Recent Developments at COGE and EPIC*

Eric Lyons, University of Arizona

2:58 pm: *Wrap up, feedback, next steps (see you in June in Sydney!)*

Blake Meyers & Chris Town

- Arabidopsis informatics needs are growing quickly, with new data types and a rapidly increasing rate of data generation.
How to consolidate and integrate these data?
- Individual investigators are devising new data handling and visualization tools that have broad utility.
How to make these widely available?
- The Arabidopsis community is global, yet most of the current informatics support is funded on a national level.
How to leverage funding across borders?
- To summarize: what is the best way to internationalize the Arabidopsis informatics efforts, to integrate new tools, to maintain long-term stability, and to address the needs of the users?

...and do this in a way that enhances the position of Arabidopsis in the top tier of model organisms.

The Challenges are the Opportunities

- The Arabidopsis community should or could lead the way in restructuring model organism databases and systems to meet the new bioinformatics needs in the plant sciences
- We must develop a system that will be sustainable, both scientifically and financially, for the next 10-15 years
- The US National Science Foundation and other funding agencies are open to provocative and risky initiatives

Now is the time to be innovative, far-reaching, and inspirational, and be the model for other organisms that are (or will be soon) facing similar challenges

2009: Discussion initiated during Arabidopsis meeting, Scotland

2010: Two workshops to discuss Arabidopsis informatics launched the consortium (IAIC) to coordinate efforts and help address Arabidopsis informatics needs

- The Multinational Arabidopsis Steering Committee: UK, April
- The North American Arabidopsis Steering Committee: US, May

2011:

NSF RCN: An International Arabidopsis Informatics Consortium

- Funded in June, 2011 for \$500K over five years.



“Design Workshop” to brainstorm for the AIP

- Atlanta, in December, 2011

Continuing Community discussions like this one

- 2009/10/11/12/13 “ICAR” (Arabidopsis) conferences.
- 2011/12/13 Plant & Animal Genome conferences.

The Plant Cell, Vol. 22: 2530–2536, August 2010, www.plantcell.org © 2010 American Society of Plant Biologists



COMMENTARY

An International Bioinformatics Infrastructure to Underpin the *Arabidopsis* Community

International Arabidopsis Informatics Consortium^{1,2}

The future bioinformatics needs of the *Arabidopsis* community as well as those of other scientific communities that depend on *Arabidopsis* resources were discussed at a pair of recent meetings held by the Multinational Arabidopsis Steering Committee and the North American Arabidopsis Steering Committee. There are extensive tools and resources for information storage, curation, and retrieval of *Arabidopsis* data that have been developed over recent years primarily through the activities of The Arabidopsis Information Resource, the Nottingham Arabidopsis Stock Centre, and the Arabidopsis Biological Resource Center, among others. However, the rapid expansion in many data types, the international basis of the *Arabidopsis* community, and changing priorities of the funding agencies all suggest the need for changes in the way informatics infrastructure is developed and maintained. We propose that there is a need for a single core resource that is integrated into a larger international consortium of investigators. We envision this to consist of a distributed system of data, tools, and resources, accessed via a single information portal and funded by a variety of sources, under shared international management of an International Arabidopsis Informatics Consortium (IAIC). This article outlines the proposal for the development, management, operations, and continued funding for the IAIC.

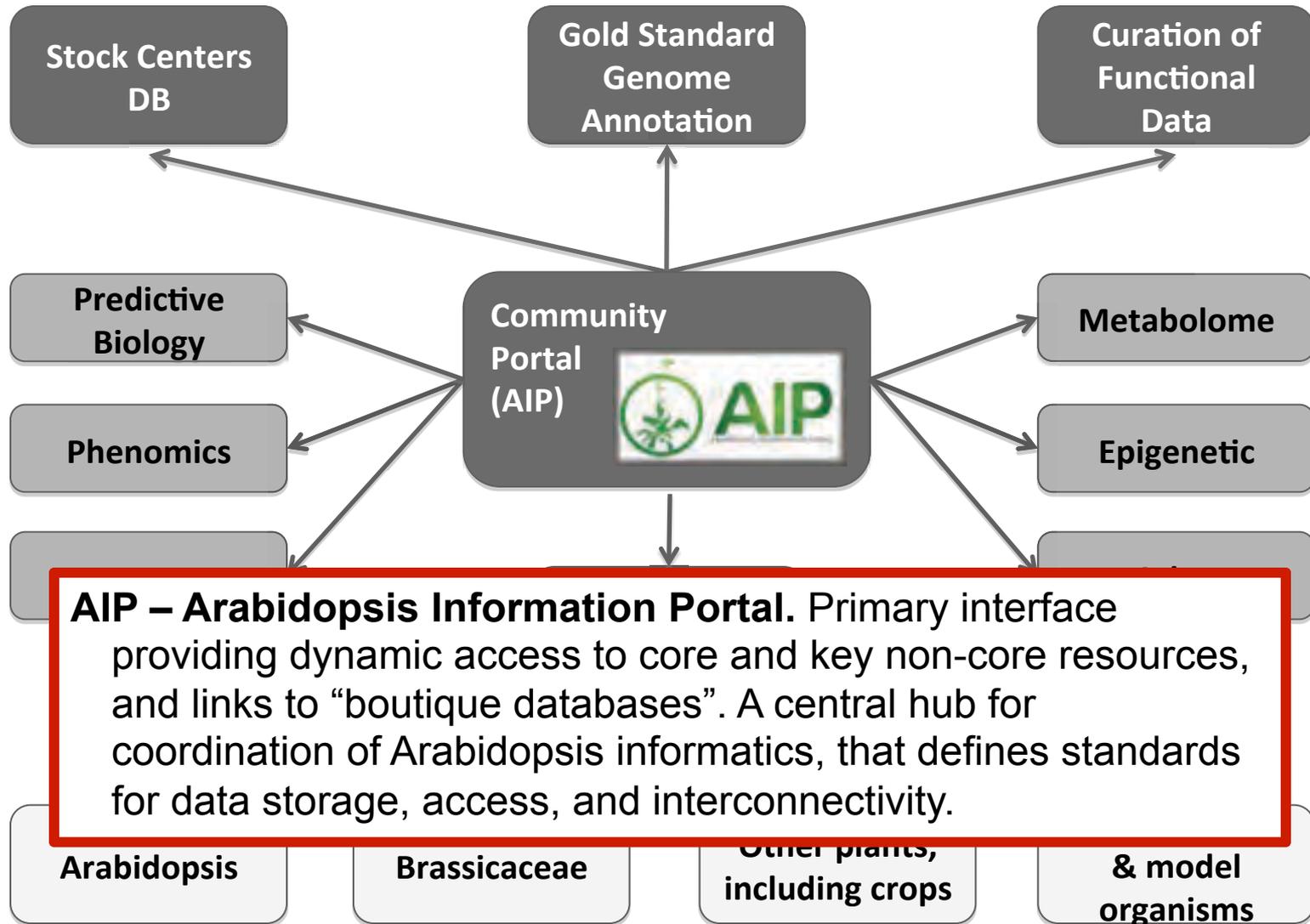
The Multinational Arabidopsis Steering Committee (MASC) and the North American Arabidopsis Steering Committee (NAASC) hosted workshops in Nottingham, UK (April 15 to 16, 2010) and Washington DC (May 10 to 11, 2010) to consider the future bioinformatics needs of the *Arabidopsis* community as well as other science

recent years primarily through the activities of The Arabidopsis Information Resource (TAIR), the Nottingham Arabidopsis Stock Centre (NASC), and the Arabidopsis Biological Resource Center, among others. However, the *Arabidopsis* community and funding agencies recognize the need for a single data management infrastructure:

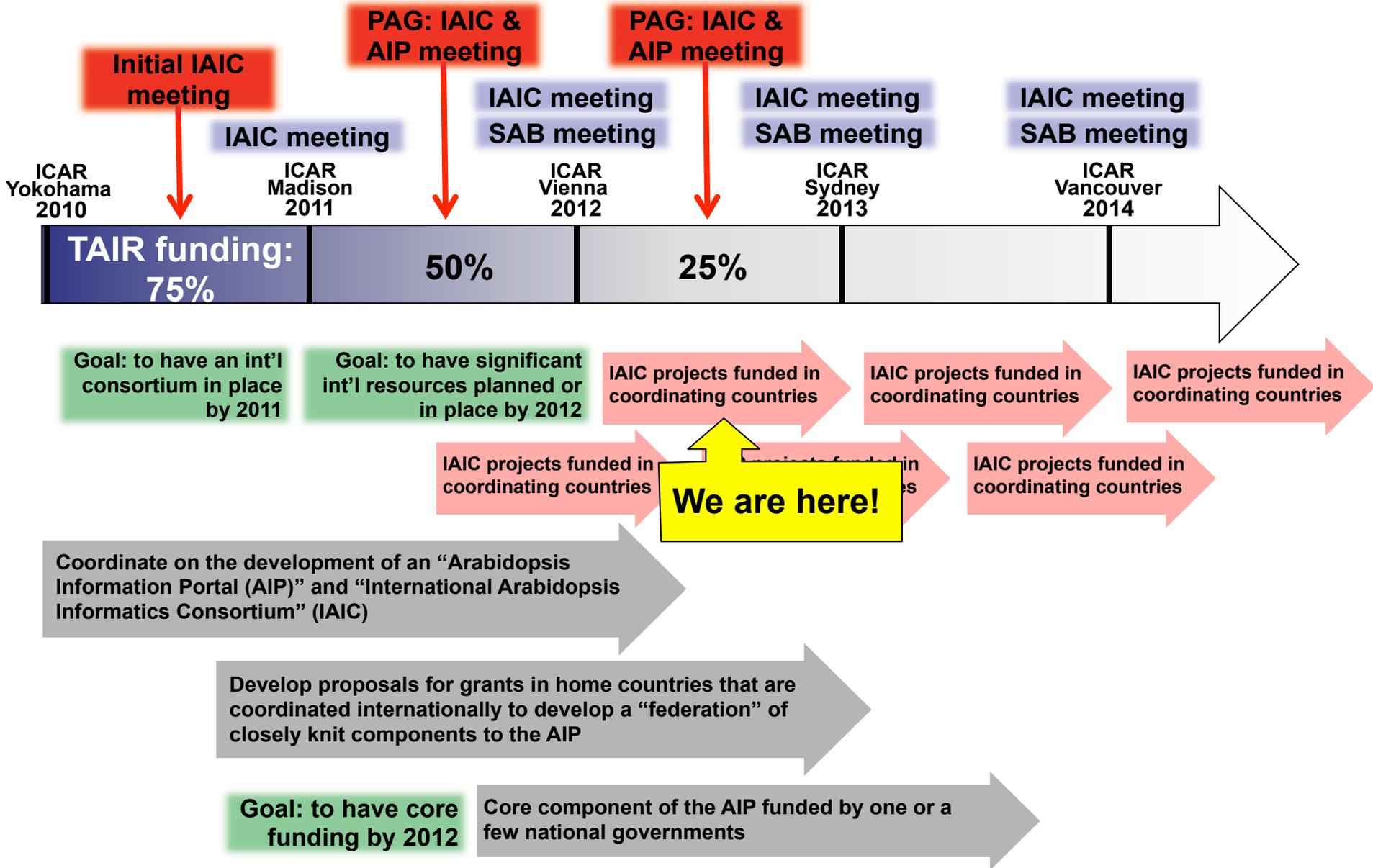
level modeling of plant processes and ultimately to translate these findings to crop plants. To achieve these goals, we must develop novel approaches to data management, integration, and access.

The UK workshop addressed three principal issues: the types of data generated by the *Arabidopsis* community, the types of

The proposed structure of the AIP



Proposed Timeline for IAIC Activities



- **Steps for developing the AIP, and an “action plan”:**
 - **What we need to build**
 - **What are the use cases?**
 - **How do we make it broadly accessible?**
 - **How do we integrate diverse datatypes?**
 - **How to make it robust to changes in data?**
 - **How to build it & fund it?**

***Working groups* worked in advance on specific issues, including:**

1) Engineering/architecture/infrastructure:

Countries represented: **Germany, US, UK, Japan, Canada, China**

2) Modules/use cases:

Countries represented: **Canada, Chile, US, Austria, Switzerland, Germany, UK**

3) Standards/ontologies:

Countries represented: **UK, US, Switzerland, UK**

Brainstorming transformative aspects of the AIP

Examples:

- **Modular design:** integrate global resources
- **Individualized user space:** ‘Dashboard’
- **“Extreme curation” and community annotation**
- **Full integration of social media tools:** App store
- **Allow direct access to databases, data stores, and other low level assets:** “power users” can develop new methods and applications

Identifying challenges for the AIP

Examples:

- **Achieving community buy-in**
- **Community agreement on data standards and their adoption**
- **Expensive software development**
- **Redirection to community development of apps**
- **Discrete fundable projects**
- **Managing data**
- **Leveraging existing cyberinfrastructure**

The Plant Cell, Vol. 24: 2248–2256, June 2012, www.plantcell.org © 2012 American Society of Plant Biologists. A



COMMENTARY

Taking the Next Step: Building an Arabidopsis Information Portal ^{DA}

The International Arabidopsis Informatics Consortium^{1,2}

The Arabidopsis Information Portal (AIP), a resource expected to provide access to all community data and combine outputs into a single user-friendly interface, has emerged from community discussions over the last 23 months. These discussions began during two closely linked workshops in early 2010 that established the International Arabidopsis Informatics Consortium (IAIC). The design of the AIP will provide core functionality while remaining flexible to encourage multiple contributors and constant innovation. An IAIC-hosted Design Workshop in December 2011 proposed a structure for the AIP to provide a framework for the minimal components of a functional community portal while retaining flexibility to rapidly extend the resource to other species. We now invite broader participation in the AIP development process so that the resource can be implemented in a timely manner.

INTRODUCTION

Increasing global demands for improved agricultural products including fiber, food, and fuel intensifies the need for a thorough understanding of the basic biology and ecology of plants. The growing challenges of climate instability, competition for arable land, and the need to transition to renewable and carbon-neutral energy sources will only

cross and other economically and agronomically important species. As in nearly all major scientific endeavors, the Arabidopsis community is global. However, in recent years, most of the financial support for the primary community informatics resource has been provided on a national level, specifically, the funding of The Arabidopsis

storage, access, and interconnectivity. It is worthwhile emphasizing that the Arabidopsis community in particular, and the plant community as a whole, needs the services currently provided by TAIR, and in the near future to be provide by the AIP, because innumerable non-Arabidopsis publications reference Arabidopsis genes. In the ab-

Arabidopsis Information Portal Development

- Several concepts for the format, structure and capabilities of the AIP generated at the design workshop culminated in a proposal to NSF.
 - *Submitted by PI Chris Town, Sept. 2012, (next speaker)*

The **Scientific Advisory Board (SAB)** will provide key AIP oversight and:

- Guide and shape future activities
- Help to encourage compliance with standards set by the AIP
- Point of contact with the community/contributors

Members elected February 2012

1. **Mark Estelle**, University of California, San Diego, USA (**chair**)
2. **Gloria Coruzzi**, New York University, USA
3. **Kazuki Saito**, RIKEN PSC, Chiba University, Japan
4. **Magnus Nordborg**, GMI, Austria
5. **Mark Forster**, Syngenta, UK
6. **Paul Kersey**, European Bioinformatics Institute (EBI), UK
7. **Xuemei Chen**, University of California, Riverside, USA

- Participate in the IAIC and in oversight of the AIP via the Scientific Advisory Board (SAB)
- Continue to develop open standards, to push for open data, good metadata, & open source requirements, w/ enforcement led by IAIC
- **Develop proposals for grants in home countries that are coordinated internationally to develop a “federation” of closely knit components to the AIP**
- **Continue to remain engaged in the community-led process to ensure that the AIP meets current and future needs** (e.g. via conferences such as PAG and ICAR, and the Multinational Arabidopsis and North American Arabidopsis Steering Committees)

We’re on the web at:

www.arabidopsisinformatics.org