

Multinational Brassica Genome Project (MBGP)

Steering committee meeting at PAG XXV, San Diego

Date: Sunday, January 15, 2017

Time: 1:30 pm until 3:30 pm

Location: Garden Saloon, Town & Country Hotel

Minutes (*prepared by Rod Snowdon, Graham King, Anne Laperche and Boulos Chalhoub*)

Present: Ian Bancroft, Graham King, Yong Pyo Lim, Dave Edwards, Jacqui Batley, Phillip Bayer, Bhavna Hurgobin, Chris Town, Felix Dreyer, Venki Pegadaraju, Wiktor Jurkowski, Anne Laperche, Marie-Françoise Niogret, Steve Robinson, Janina Braatz.

(Apology from those having participated and not appearing on this list)

Apologies: Lixi Jiang, Shengyi Liu

Chair: Boulos Chalhoub

Agenda

1. Introduction (Boulos)

General introduction for participants.

Aims and missions of MBGP discussed at previous MBGP meeting at Melbourne (October 6, 2016) have been briefly discussed and confirmed.

2. Approval of minutes from Brassica 2016 meeting (<http://brassica.info>)

Participants have unanimously approved the Minutes of last MBGP meeting (Melbourne, October 6 2016).

3. Scope for multinational projects – What works well in other systems and what doesn't?

- Compare MBGP with ongoing consortia for e.g. Arabidopsis (MASC), Wheat (IWGS), Rice, Rosaceae (GDR) and others. What about BMAP?
- Suggest develop a “road-map” document of resources that would benefit research and breeding community (build on minutes from Melbourne)

MBGP works very well. It is a good system. Strengths of MBGP: Relaxed relationship, good knowledge flow about who is doing what internationally.

Coordination is important: Interoperable structures with individual countries delivering different components and data?

Brassica Information System: First requirement is establishing a coordination – Graham King as organiser, people who are interested in participating and in data contribution should contact Graham.

Earlham Institute (Viktor) able to integrate data into their data warehouse system and has possibilities to get BBSRC funding to expand and host an extended BIS. BBSRC is very supportive of programs for data management/

How to communicate and exchange on different parts of data?

-After the MBGP meeting, interested ‘specialized’ people sit down and discuss further, define a strategy depending on where and how to ask funds.

Ian & Dave: Learn from the wheat experience. Generate the MBGP strategy inspired from wheat.

4. Quality assurance of Brassica genome sequences and resources (I. Bancroft, I. parkin)

- Quality standards to get MBGP support?

Significant quality problems with the recently published *B. nigra* and *B. juncea* genome assemblies. These have now been removed from the BRAD website. Assembly procedure appears to be flawed, Ian Bancroft has communicated with the journal to point out the problems and is communicating with the authors to ensure that the sequences are corrected before being released again.

This example showed the importance of establishing defined community standards for new resources and discussion in the community regarding new resources. Does MBGP have a mandate to approve the quality of a new genome sequence or annotation?

Ian: There is a need for some kind of community endorsement prior to publication/release of new genomes (one idea would be to implement e.g. "MBGP-endorsement" prior to public release). Getting together of good quality standards. MBGP can say we approve?

Graham trying to get some collective statements to recommend to Nature Genetics for quality assurance....

Action: Ian will communicate with the authors and negotiate to have the revised genomes quality-checked before release.

Steve Robinson: AAFC *B. nigra* genome from Canada is expected to be released soon.

5. Brassica - Information System & platform

- **Engagement with URGI, etc (update from last MBGP at Melbourne. R. Delourme & Rod Snowdon)**

Anne Laperche: URGI system is established and used in RAPSODYN project. Trait ontologies are freely available.

G. King suggests that URGI/Rapsodyn should communicate with Earlham Institute (Viktor) to coordinate and standardise.

- **Progress on creation of inventory (Rod Snowdon, Dave Edwards)**
- **Development of data warehouse of datasets,**
- **to include:**

- Repository for comparative analysis of re-sequenced genomes
- . inventory outlining genotype (well-defined), coverage, quality etc

G. King and D. Edwards: Get started working on it and need a way to share.

What about comparison with wheat

- *Dense linkage maps*
- Repository for public Brassica SNP array data
- *Trait/trials – links from field phenotype to genome*
- *Metabolomics*

G. King. Metabolomics data? Would be good to integrate. GK group is interested in *B. rapa*

A. Laperche: INRA-Rennes is doing some trait entology.

I. Bancroft: had 100 metabolites/ Raw data with spectra / Part of the Brassica data sets

Dave Edwards, somebody doing metabolomics in Australia/ Brassica specific contents

Rod Snowdon is involved in a big metobolomics project in rapeseed (led by Thomas Altmann at IPK Gatersleben).

Nathalie Nesi (INRA-Rennes) and Rod Snowdon (Giessen University) have been doing metabolomic and transcriptome analysis in association with germination and early vigour in rapeseed.

6. Development of standards to describe data to allow direct data comparison (Rod Snowdon)

- To allow integration of publicly available data
- **Metadata standards** – brassica-specific for non-sequence entities (eg plant lines, traits, trials, extend or validate those at brassica.info)

Standards for genotype names and tracing –important to know the providence of a line which carries a particular name because two inbreds with the same name may segregate from a common origin.

Standards for gene model nomenclature have been agreed to by most groups releasing new sequences (or updates), but still no response from Xiaowu Wang regarding the *B. rapa* update. Idea is to develop a “look-up table” which enables users to find all orthologs of a particular gene. Numbering system should enable insertion of new genes in updated annotations and in pangenomes, however it does not enable changes (or differences) in gene order.

G. King. listing these data files. Graham suggests that new proposals for renumbering of genes should be checked and approved by MBGP. Can MBGP learn from experience from Arabidopsis. They decided not to go for a single “pangenome” annotation, rather to number each gold-standard annotation individually and link them via a lookup table.

Wiktor: A list of data type and preferred way of presenting these. Wiktor can send to everyone. Map Brassica data and GAPs information data / Standard development currently.

Wiktor based on transplant project that Alexy is doing. Changing the standards

Ian: Numbering according to pan-transcriptome of diploid species? Problem is that every Bn accession has a different gene order and multiple translocations.

Specific format? Minimum set of information. What should be compensatory and what not

Action: I. B: Some aspects difficult/ Easy part is describing samples and putting descriptors in database, like for the ASSYST population from Rod. Ian suggests this is something that should be described in the general information management system.

➤ **Gene model** nomenclature

Reference genomes (Graham King, liaising with relevant research groups)

Graham King: Five groups are included in reference genomes. The adopted nomenclature (at Wuhan in 2014) was only used for *Brassica napus* cv Darmor-bzh sequence (Boulos Chalhoub). Dave Edwards, is using the nomeclature system for *Brassica napus* cv Tapidor Ordering index system

Ian Bancroft: use of the pan transcriptome reference.

Someone will propose renumbering. Very easy to have that done.

Chris Town Arabidopsis has 5 standards (and there will be discussion after the presentation)

G. King: it is valuable to have Darmor on genome, chromosome etc...

Adopt that nomenclature for *B. oleracea* cv TO1000 and *Brassica rapa* cv Chiifu

IB: another way of annotating that? AAA and slipt to CCC
Dave and Ian and Chris Town are happy and not happy
Relative low amount of data/ ordering the accessions.
Discussing with people (IB) no big bioinformatics?

NB: We didn't resolve the issue. Working with thousands would be difficult. In those particular accessions

➤ **Formal (functional) annotation** – An international annotation group/ (again)?

G.K. Different models, working on these and distribution Experience from wheat: not always best to enforce new standards, rather to establish workable guidelines for data standards and encourage people to use them.

7. Development and adoption of Crop Ontologies for Brassica (phenotypic trait etc)

Boulos have no specific notes on this issue.

8. Miscellaneous

J. Batley – ABC Infinium chip now in production, 90k at same price per sample as the 60K AC-array. Jacqui/Isobel to release probe sequences.

G. King: Brassica.info still running, needs updating with newest data. Possibility to add contact data for community? Is Brassica mailing list still the most effective communication medium?

New data resources. Ian (EvoGenapus mRNAseq data, JIC new project with lots of RNAseq data, plans to produce an expression browser.

AAFC plan to release Bni, Bj and Bcar. AAFC and JIC have carinata mapping pops, Ian could generate the transcriptome data for the JIC pop if the seeds germinate (soon!)

D. Edwards, Jacqui, Rod & Boulos have a new Darmor assembly generated with NRGene De novo Magic. Quality looks very good with ~200 Mb assembly size increase, will be released as soon as the last improvements on the pseudomolecule assembly have been completed to fix some collapsed regions. Open for use and collaborations already.

I. Bancroft: Need to better engage the Chinese groups in the MBGP.

9. Election / nomination of new MBGP chair

Graham King was suggested as new chair for the next coming two years to replace Boulos Chalhoub. He will be seconded by Jacqueline Batley.