

## Multinational Brassica Genome Project (MBGP)

Steering committee meeting at PAG XXVI, San Diego

**Date:** Sunday January 14<sup>th</sup> 2018

**Time:** 1.30 pm – 3.30 pm

**Location:** Stratford Room, Town and Country Hotel

**Minutes** (prepared by Jacqui Batley, Graham King and Natalie Nesi)

**Present:** Graham King, Lenka Havlickova, Sarah Turner, Haitao, Agnieszka Golliicz, Birgit Samans, Yong Pyo Lim, Scott Woody, Isobel Parkin, Andrew Sharpe, J Chris Pires, Joerg Schondelmaier, Magalie Leveugle, Rod Snowden, Steve Robinson, Dave Edwards, Jacqueline Batley, Sebastien Faure, Nathalie Nesi, Ori Weisshaus, Yonghe Bai, Michael Barker, Boulos Chalhoub, Priyakshee Borpatra Gohain + 3 representatives from Genome Canada/Prairie.

**Chair:** Graham King

### 1. Introduction and welcome

General introduction for participants.

### 2. Approval of minutes from 2017

Participants unanimously approved the Minutes of the last MBGP meeting (PAG, San Diego, January 2017)

### 3. Reports on progress/notable milestones/announcements from members

- Chris Pires (with Mike Barker): Working on *B. rapa* USDA collection. Have found 15-20% accessions are miss-identified. Generating transcriptomic data for 100 lines. Sequencing *Crambe* with Pacbio and phase genomics, plan to do other Brassicaceae. Will contribute the data to pan genomics projects
- Rod Snowdon (Giessen): Has a number of public and private projects. Resequencing 100 *B. napus*, also generating transcriptomic, bi-sulphite and miRNA data for all. Unsure when/if this will be made public, but at least 1 reference should be released. Resequencing a TILLING population (Express 617)
- Dave Edwards (UWA): Managing Infinium array data in a database (<http://appliedbioinformatics.com.au/index.php/SNParrayDB>) and has a synteny viewer (<http://appliedbioinformatics.com.au/index.php/DaisyChain>). Also public assembly (can email and ask for data). With Jacqui Batley has developed an episeq capture set which can be used by others if interested.
- Jacqui Batley (UWA): Been working on R gene characterisation in pan genomes of the different *Brassica* species. Structural variation analysis across different genomes. Re-sequencing data can be made available for some Australian cultivars
- Natalie Nesi (INRA): Has been developing phenotyping tools, including drones. Projects on NUE and drought tolerance/resistance

- NRGene: Interested to work on and provide services for *Brassica* pan-genomes
- Mike Barker (Arizona): Investigating hard and soft sweeps and domestication. Sequencing new *B. rapa* in USA (gone feral and recombined)
- Graham King (SCU, Australia): Finished high quality re-sequencing of R-o-18 (TILLING line) – PacBio data generated by JIC, Norwich added to existing assemblies. Linking with bisulfite data from John Hammond (U. Reading). Working with Jun Zhou (HZAU, Wuhan) who has re-sequenced all BC4 from TNDH. [www.brassica.info](http://www.brassica.info) will be updated by March.
- Lenka Havlickova (York): Investigating structural variation in Brassicas and working on pan genomes. Mutation breeding panel developed to replace previous populations. This should be available to the community (Ian Bancroft to confirm details)
- Sarah Turner (Missouri): Working on *B. oleracea* and linking to morphology data.
- Boulos Chalhouh (Monsanto): Leading breeding and oilseed research at Monsanto. Will continue to support MBGP and be member of consortiums
- Yong Pyo Lim (CNU, Korea): Offering microspore culture of *B. oleracea*, *B. rapa* and *Raphanus sativa* as a routine service. Performing genome editing using agrobacterium transformation. GWAS for morphological and functional traits in *B. rapa* – collaboration opportunities. Marker development for traits for companies. Sequencing clubroot isolates.
- Scott Woody (Wisconsin): Education in *B. rapa* (*Wisconsin FastPlants*). Nine new RILs, have 4 in rapid cycling with yellow sarson as common parent. F2 pairwise intercrosses and then SSD to fix. Four additional populations (2 cabbages and 2 turnips) at S3 and S4. Getting a lot of sterility. FBSC seems to have novel FT allele. FPSC genome on phytozome (JGI sequenced).
- Isobel Parkin (AAFC, Saskatoon): Releasing B genome sequence, and a spring *B. napus* –will be a JBrowse available in a few weeks. B genome is Illumina reference, re-sequencing of 10 *B. nigra* with nanopore. Brassica 90K SNP array cluster file for *B. carinata* completed. Will work with Jacqui Batley on *B. juncea* cluster file
- Andy Sharpe (IFS, Saskatoon): Has moved back into Brassica research. There is a Genome Canada genomics assisted partnership program to develop pan genomes (with NRGene), haplotype databases and genotyping platforms. Characterising structural diversity in the A and C genomes. See also re: item below re: seed Genome Canada/seed protein.

#### 4. Scope for multinational collaborative projects

- Opportunities for co-funding and international funding were discussed.
- Representatives from Genome Canada/Prairie also outlined that they would be funding a new initiative starting later this year focused on modifying seed protein content. This is an international collaboration opportunity, needs to have interaction with partners, and letters of intent are due September.

- Bioinformatics: The funding proposal of Rod Snowdon, which was supported by many MBGP members, but was unsuccessful, is now incorporated into Andy Sharpe's proposal to Genome Canada.

**Action:** *all to see if this can be used to leverage funding from other countries*

- Standardised genome annotation pipeline: Graham asked if there was an appetite to start this as a distributed process?
  - For example hand annotation of sub-groups of genes, e.g R genes and seed storage proteins, as well as annotation of splice variants and transcription start sites.
  - Dave suggested that there needs to be a mechanism to transfer these to new reference genomes when they become available. Graham sees this as an international collaboration opportunity – many individual labs are undertaking this on specific target gene classes, but how can it be linked out as a public resource?
  - Agnieszka noted that in rice and Arabidopsis they link IDs and traditional gene names. Graham suggested that look-up tables would be helpful.

**Action:** *to be discussed further at St Malo meeting in July*

## 5. Brassica Information System – briefing from sub-group and discussion of roadmap

- The Brassica Information System (BIS) doesn't yet exist! However, this meeting marked a step beyond its inception. A sub-group meeting was held on Saturday 13<sup>th</sup> January to progress gestation.
- Graham had circulated the proposed Road-Map to the sub-group in November 2017. There are 3 phases – inventory, collection/triage and detailed curation.
- There was a consensus at the sub-group meeting to adopt the approach taken by the Wheat Information system, which uses distributed databases, supports individual efforts and is linked by a common search tool. The Brassica community can learn from the mistakes made in wheat.
- Steps/capabilities required include:
  - Inventory of existing dataset sources.
  - Prioritise datasets to enable:
    - Centralised indices for existing databases and flat files, likely building on the Apache 'Solr' system.
  - Wider development and adoption of standards for brassica-related experimental entities (biosamples, trials, traits etc)
  - Addressing issues of synonyms/homonyms – scope for allocating DOIs (FAO have facilitated this for Ag-research). Moving towards establishing look-up registries and allocation of unique identifiers.
  - Accept the reality of pan-genomes and develop a means of managing gene-name referencing (see 4 above).

**Action 1** (Dave Edwards/Graham King): *A questionnaire will be distributed via Brassica.info. This will be based on the wheat information system questionnaire and will require datasets to be nominated for sequence, markers, phenotype etc.*

**Action 2** (Graham King) *Update BIS Road-Map to incorporate feedback above. Place on [www.brassica.info](http://www.brassica.info) to form White Paper for the MBGP community.*

**Action 3 (Dave Edwards/Graham King):** Determine whether the query page can be set-up to appear as part of the Brassica.info domain.

## 6. Standards, ontologies and Brassica trait dictionary

- Chris Pires asked if there was a standard for naming genes in pan genomes?
- Graham pointed out there is a recommendation which is available online ([http://www.brassica.info/info/genome\\_annotation.php](http://www.brassica.info/info/genome_annotation.php)) and has been agreed on in two previous meetings.
- It was suggested that MBGP should supply the standard to journals and research groups involved in sequencing? The issue is particularly problematic in pan genomes where data may alter as gene names are added, leading to discrepancies between different versions. There are also the issues of PAV (presence/absence), CNV (copy number) etc in different annotations.

**Action:** The issue will continued to be communicated, as an interim step there will be a look-up table to enable ease of gene identification between different assemblies.

The synteny viewer from Dave Edwards may be suitable for this :

<http://appliedbioinformatics.com.au/index.php/DaisyChain>.

- **BraTO (Brassica Trait Ontology):** Natalie Nesi updated the group on BraTO. Data can be contributed to this (contact Wiktor.Jurkowski@earlham.ac.uk)
- Brassica Trait Ontology (BraTO) hosts trait information to describe brassica crop data. Terms are collected from various projects including OREGIN, RIPR (UK) and RAPSODYN (France) at present, and any other project is welcome to add other trait information. BRaTO development is conducted by Earlham Institute (UK), Southern Cross University (Australia) and INRA (France). Information is available at [http://www.croponontology.org/ontology/CO\\_348/Brassica](http://www.croponontology.org/ontology/CO_348/Brassica)

### Pipeline to enter new BraTO terms:

- At first, each single variable is defined by a given research team as a combination of “a trait + a method + a scale”. All the variables are gathered in a Trait Dictionary template (xls file) whose versions are managed in a github repository. Wiktor Jurkowski (Earlham University, UK) is the curator.
- Graham mentioned this is still maturing as a system and Earlham (Wiktor) and his group are working to make it MIAPPE (<http://www.miappe.org/>) compliant. Isobel is also working towards ontologies and asked what to do with digital data. Isobel and Andy will review what has been done already to avoid multiple ontologies.
- The ASSYST data from Rod follows the standards used in the NAM population. However ASSYST data from other labs may not be following the same standards.
  - Graham asked whether it would be possible to get funding to enable all ASSYST data curated and into the public domain? Rod had spoken to someone at Pioneer who fund “public good” projects.

**Action:** Rod to talk with the Pioneer representative and also German breeders to see if there would be interest in funding this.

## 7. Quality Assurance of genome sequence

- Graham noted that it is very important for students new to brassica to understand what is the current genome version and where do they go to find this out?
- In the MBGP meeting in January 2017 it was asked whether there could be MBGP approved genome versions. Is it possible to say which is the primary or secondary version for example? Graham agreed to set up a table for this in Brassica.info with explicit version numbers and links to primary sources.

**Action:** Dave to suggest what basic stats should be included in this information (include BUSCO?)

- Agnieszka noted that on Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>) there are 5 sentences describing what has changed between different versions. It was agreed that this should be implemented for the different *Brassica* genomes.

**Action:** Graham to co-ordinate compilation of information about different genome versions for Brassica.info

- Options were presented as to how MBGP could 'endorse' a genome. One possibility is to have >1 person on MBGP look at the genome and these cannot be nominated by the producer of the sequence. In the cover letter to journals it could then state it has been endorsed.

**Action:** This proposal will be circulated as a recommendation.

## 8. Any other items/meeting announcements

**ABRC:** (Arabidopsis Biological Resource Center: <https://abrc.osu.edu/>) At the end of the Brassica workshop they announced they are taking *Brassica* seed resources. This is excellent for MBGP community, as currently there is no central repository. They are accepting donations and it would be key to at least add all sequenced lines.

**Brassica 2018:** July 1-4 St Malo (<https://colloque.inra.fr/brassica2018/>). Registration and abstract submission are now open.

It was suggested that the next MBGP meeting be held during the conference

**Rapeseed Congress** June 16-19 2019, Berlin. Tours include Giessen!

<http://gcirc.org/event/single-view/article/15th-gcirc-rapeseed-congress-berlin-june-16-20-2019.html>

**International clubroot meeting:** Alberta July 7-9 2018.

<https://albertacanola.com/event/2018-international-clubroot-workshop/>