

Multinational Brassica Genome Project (MBGP)

Steering committee meeting 1 (2 meetings held due to time zones)

Date: February 3rd 2022

Time: 9am (Perth)

Location: Online (due to COVID)

Present: Graham King, Dave Edwards, Jacqui Batley, Yong Pyo Lim, Matthew Nelson, Xiaowu Wang, Evgeny Glazov (*Apology to those having participated and not appearing on the list*)

&

Date: February 3rd 2022

Time: 5pm (Perth)

Location: Online (due to COVID)

Present: Graham King, Dave Edwards, Jacqui Batley, Annaliese Mason, Rod Snowdon, Jun Zou, Cyril Falentin, Guusje Bonnema, Matthieu Rousseau, He Yuze, Srijan Shingon, Ulrike Miersch (*Apology to those having participated and not appearing on the list*)

(minutes below colour-coded accordingly)

Minutes (prepared by Jacqui Batley)

Chair: Jacqui Batley

1. Introduction and welcome

General introduction for participants.

2. Approval of minutes from January 2020

Participants unanimously approved the Minutes of the last MBGP meeting (PAG, San Diego, January 2020)

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

These are resources available to share, not private data.

Yong Pyo Lim: Will retire this month, probably moving to private university/research institute. The Chiffu BAC library will be donated to the government. Sequenced a core collection, also has a mapping population and has undertaken morphological characterisation. At the National Institute of Agricultural Science there are efforts into genome editing in *B. rapa*. Developing molecular markers for clubroot resistance, also has 17 mutants. Has a postdoc going to Alberta who sequenced ECD4, which was recently published.

Xiaowu Wang: Recently published *B. rapa* pan genome (16 accessions). Recently finished illumine re-sequencing of >1500 accessions of *B. rapa* (>30x), which is being de novo assembled and will be combined with the published data. Doesn't have good accessions of wild material for *B. rapa* and *B. oleracea* and would like accessions from the wider community. Interested in developing a *B. nigra* pan genome, would like to collaborate if anyone can contribute. Has a TILLING population of heading *B. rapa*, >2000 lines, all are sequenced, some will be available where plentiful seed.

ACTION: Please contact Xiaowu if you have available accessions and would like to collaborate for *B. rapa*, *B. oleracea* and *B. nigra*

Matthew Nelson: Undertaking an analysis in gaps in collections of wild Brassicas, there are many duplicates and feral lines that are not wild which needs to be tidied up. Has >200 wild relatives currently, looking to recruit a postdoc in this area. Has funds for Dartseq to study the genetic relationships. Chris Heliwell doing RNASeq/machine learning to identify genes for phenology, also looking at establishment genes. Have a TILLING population (Zircon), however not yet clear if will become publicly available. Will be undertaking targeted resequencing of this population

Dave Edwards: Working on pan genome development and visualisation tools, has undertaken in wheat and now extending to Brassicas. Also using machine learning in genomics for gene prediction models, association analysis and functional annotation. Establishing a Centre for Applied Bioinformatics and would like to link to other organisations. Looking for postdoc/PhD students.

Jacqui Batley: Working on blackleg gene identification -Rlm4 and 7 now available. Funding to undertake further sequencing and pan genomes in wild species with an interest on characterisation of resistance genes.

Matthieu: have an improved Darmor genome that was released last year (published in Gigascience). This should be available on BRAD.

Ian Bancroft noted that this is a nice assembly and he is using as new reference

Ian Bancroft: funding is quiet, there is no Brassica strategy. Has finally published paper on nomenclature.

Jun Zou: Assembling 2 *B. carinata* genomes, still in progress, using nanopore and HiC

Annaliese Mason: Noted that there is a *B. juncea* genome (RS also involved – pac bio, published in Nature Genetics), with large re-sequencing efforts, this was published last year. Is a *B. carinata* genome reference that was available in 2021. Shengyi Lui has sequenced synthetic *B. napus* (50x coverage) – approx. 40 lines. Has allohexaploid re-sequencing data, a mapping population and few other lines – illumina data.

Guusje: has a cauliflower and pointed cabbage genome. Undertaking SV detection to understand cauliflower domestication. The project is being finalised this year and the data will be released 6 months after completion. Also investigated recombination landscape in kohlrabi, cabbage and cauliflower, this will be published in the future. Has a genetic diversity paper in Hort. Science, describing *B. oleracea* collection.

Rod Snowdon: Express assembly published. Has a new project to sequence 100 diverse accessions from ASSYST. This will be 40x nanopore sequencing to develop SV atlas. It will consist of 25 spring, 25 winter, 25 swede and 25 semi winter lines. The data will available later this year. Noted that Agnieszka Golicz has moved to Giessen and is starting her own group. Working on 3D genome confirmation.

Graham King: Finished annotation of R-o-18, in Ensembl plants and Genbank. Annotation using Ian's pipeline. This currently is the only genome in genbank with the new annotation. Genbank have agreed a protocol for inserting the pan genome gene name as a /label in the feature table.

4. Nomenclature

Previously in MBGP steering committee meetings there was agreement to establish a consensus *Brassica* gene-model naming system based on the diploid-pan-genomes. Ian Bancroft, Dave Edwards and Graham King were working on this.

The issue is how to get existing reference genome annotation re-labelled with MBGP pan genome names that are all integrated as a reference. For this, we need permission of original submitters to NCBI. The first step is to get a list of all available genomes, to identify who to get permission from. There is a look up table in He et al paper, however it needs to be discussed who implements this: Ensembl are on board but first we need to have genomes in NCBI. Any new genomes needs to run alignment against the He dataset.

ACTION: Work with others to get list of genomes together and follow up asking for permission (GK)

Ensembl are happy to take in additional reference genomes with additional labels. A system needs to be developed to retro fit the information from Ian's paper. Graham collating a list of genomes, needs time and resources to put these through the pipeline. Ian had applied for money to do this, but was unfortunately unsuccessful. In order to go into Ensembl it needs to first be entered into genbank so they can pick it up.

Graham King's lab (Ramil Mauleon) has a workflow he thinks will work, but needs someone to do the grunt work and need to see how long it will take to determine what is realistically feasible to do.

ACTION: Given time, GK to write position paper on what is required. This can then be distributed to individuals to say will be done on the genomes they developed, and see who is willing to undertake.

5. Brassica Information System

There has been little movement on this since the last meeting. There is a model from the Wheat Information System (WIS) for indexing a range of remote databases. These are indexed in a standard way and the search comes back to the original database. This should be easy to implement in Brassicas.

ACTION: Dave Edwards to progress

GK mentioned that we could investigate possibility of joint funding with wheat and Brassica and it could be worth having conversation with Peter Langridge about how we could join with wheat effort in some way to get funding support. Divseek may have a little funding to get it done. Graham's mission is to get more funds use Wheat and Brassica as model for a generic platform

IB noted there is funding for wheat through institutes in the UK. He couldn't get resources from BBSRC for this, but perhaps institutes could, this could be investigated with Anthony Hall. Potentially a relevant DEFRA call.

ACTION: Graham King to progress (if time available)

It was acknowledged that someone would be required for setting up a front page and providing support for indexing.

6. BRAD update

Xiaowu Wang presented an update on BRAD. It is a useful data for non-bioinformaticians. It is organised in a table, designed for hosting the *B. rapa* genome. It has now been extended to all available Brassicaceae genomes. In the publication there are 26 species and 36 genomes/annotations. However another 4 genomes have been added since. The genes are organised syntenically, and the synteny was identified using a tool developed in house. Sub genomics syntenic relationships are also analysed. All genes can be linked to Arabidopsis corresponding syntenic genes (in a pop up window). This new version of BRAD (v3) also has a more user friendly interface and polymorphisms are integrated, which makes use of published variation data. It is possible to look at selection signal of published SNPs. There is a multiple sequence alignment tool and phylogenetic tree tool. Microsynteny can be analysed – 2 genomes can be viewed in a small region. Aim to add another 22 genomes in future. There is also an aim to help breeders by linking a KASP primer design tool. Expression data will also be integrated into the database, as well as new module to show pan genomes. For more information please contact Xiaowu.

7. Scope for multinational collaborative projects

IB noted there is strong support in UK for participating in EU framework projects. There will be a strong international program moving forward. DE noted that there is also access for Australia to participate in EU projects.

RS is preparing a grant for an upcoming call with Germany, US and UK – one of the focus topics is crop genomics.

GK noted that Australia is now a member of the Elixir network (EBI and Ensembl are also members).

8. Brassica 60K/90K array

Evgeny Glazov presented results of a survey conducted to determine what will happen with the product moving forward. At the time of the meeting there had been 10 responses, and 7 of these groups intend to keep using the array. Most users estimate >10,000 samples per year. Most users expect to keep using for >1 year more. When using alternative tools, these tend to be WGS or RRGs. Noone is interested in development of another high density array. Potential interest in lower density, that can support breeding and germplasm characterisation in genbanks. Illumina are still collecting data and are happy to receive feedback. A decision is likely to be made mid 2022 including if a new array will be developed. If so this should be available this year and the current array will overlap for this year and next year.

AM noted after presentation that it is a cost related issue, need to determine the price range and number of SNPs to get optimal use. Need to determine this to see how many people would be interested at the difference densities/prices. AM noted that the array is more useful than sequencing when there is less bioinformatics support. IB also noted that there will be a strong demand for the array to continue, especially from groups with no bioinformatics. RS noted that breeding companies had switched to the 15K from Trait Genetics, as more cost effective (cost is half that of 60K array). This has been optimised for single copy SNPs.

MR noted that in current arrays there are regions with no markers due to state of genome assembly at time of array design, so in these regions must design KASP instead currently. If another array is produced it needs to cover whole chromosomes. Don't want cost of >30-40 Euro.

IB noted there is growing interest in B genome polymorphism. EG noted that this could be included as they want the array to be of broad interest. A possibility is having separate arrays for different sub-genomes. AM noted that she is keen to cover Crop wild relatives, but breeders may not be, so 2 arrays may be an option. RS noted that the species are closely related and to a dual species chip may not work well.

ACTION: Please contact Evgeny for any further information/feedback. Evgeny Glazhov +61 419 867 480 eglazov@illumina.com

9. Any other items/meeting announcements

Brassica 2020: Unclear whether this will still go ahead.

IRC: September 2023 Sydney. This will be in person and potential for online also

Agriculture and Climate Change meeting, in person, May 2023, Dresden

PAG 2023: Ian will be running the Brassica workshop