

Multinational Brassica Genome Project (MBGP)

Steering Committee meeting at *Brassica 2018*, St Malo, France

Date: Tuesday July 3rd 2018

Time: 6 – 7.30 pm

Location: St Malo Convention Centre

Minutes (prepared by Jacqui Batley and Graham King)

Present: Graham King, Dave Edwards, Jacqui Batley, Nathalie Nesi, Anne Marie Chevre, Rod Snowdon, Cyril Falentin, Sebastien Faure, Lenka Havlickova, Ian Bancroft, Yong Pyo Lim, Steve Robinson, Mathieu Rosseau-Guentin, Chaobo Tong, Isobel Parkin, Annaliese Mason, Geoffrey Wagner, Eric Schranz, Jenny Lee, Ning Guo, Dandan Hu, Jun Zou, Sarah Turner, Guusje Bonnema, Rachel Wells, Shenyi Liu, Asjad Ali, Jian Wu, Fan Liu (*apologies to those having participated and not appearing on the list*)

Chair: Graham King

1. Introduction and welcome

General introduction for participants.

2. Approval of minutes from January 2018

Participants unanimously approved the Minutes of the last MBGP meeting (PAG, San Diego, January 2018)

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

- Graham King suggested only presenting new work on top of what was presented at the meeting in January. These are resources available to **share**, not private data.
- Anne -Marie Chevre: Genoscope have just published a **new assembly** of *B. rapa* and *B. oleracea* using nanopore and OM. This will be on their website, and there is a publication submitted. The genomes are registered in NCBI. (https://www.ncbi.nlm.nih.gov/assembly/GCA_900412535.1/ and https://www.ncbi.nlm.nih.gov/assembly/GCA_900416815.2/)
- Graham has asked that all genomes are **registered** in NCBI as '**assemblies**' so that they are ready to go into ENSEMBL Plants.

Action (Anne-Marie): Send Graham King an email about it so it can go onto Brassica.info

- Dave Edwards: Has **pan genomes** for *B. napus*, *B. rapa*, and an expanded *B. oleracea* set. This work is being prepared for publication – happy to collaborate on it (either to do the work or have collaborators do it). e.g certain types of genes or analysis. Are ~70 accessions for each spp.
 - Graham King noted that this fits well with the role of this committee shares, with an aim to try and improve publication, as is suggested here.

- Rod Snowdon: Has a **new reference genome** for a winter *B. napus* genotype – ‘Express’. He is happy to share the data prior to publication, although this has yet to be annotated. There are Illumina data, 10X for scaffolding and 70X PacBio data, which may also help other assemblies. Seeing some significant differences to Darmor-*bzh*, they are currently assessing whether it is error in one or other assembly or real. The assembly is currently being fixed with Bionano and genetic mapping data. Also working with Dave to publish with the NRGene assembly. Specific information is below.

Table 4.1 Statistics of genome sequencing

Sample	Library	Insert size	Total data (G)	Read length (bp)	Sequence coverage (X)
BNS14013	Pair-end libraries	350bp	84.47	150	78.46
		450bp	94.69		87.95
		2k	51.78		48.10
		5k	55.45		51.50
		10k	34.67		32.20
	10XG	10XG	143.37	-	133.17
	PacBio	pacbio	58.68	-	54.50
	Total	---	523.10	-	485.87

Table 4.2 Statistics of genome assembly

Sample ID	length		number	
	Contig**(bp)	Scaffold(bp)	Contig**	Scaffold
Total	928,225,725	929,482,120	2,204	1,547
Max	11,217,797	19,639,753	-	-
Number>=2000	-	-	2,140	1,490
N50	2,366,204	6,480,219	113	45
N60	1,869,218	4,460,738	158	62
N70	1,374,369	3,340,131	215	86
N80	827,522	2,322,464	299	118
N90	262,834	540,500	492	200

** Contig after scaffolding

- Ian Bancroft: Lenka’s **radiation mutagenesis panel** unfortunately didn’t get funded, so they have put together what funding they can, but there is no industry partner involved and therefore it can **be made available**. This will not be as a service. If used commercially then there will need to be a **license**. If wanting to work on it, contact Ian, state which gene interested in. It will have a small re-amplification charge and will possibly ask for seeds back. They will use the ABC pan genomes to re-map sequence data back. Diploid genomes are being used as reference. Earlham institute (EI) pulled out of *Brassica* data management. BIP is currently running on autopilot.

- Graham King asked if **BIP** data can be downloaded to something else in future? There is concern if this is left too long then it won't be available anymore. He will enquire to EI and BBSRC.
- Shengyi Liu: -Has sequence data for **1000 accessions of *B. napus*** -these data will be released, however if before release someone wants to look at linear variation across accessions then he is happy to collaborate. Some lines > 30x, with a minimum 7x coverage. He is working on developing a pan genome and is happy to collaborate. Material represented includes kale, rutabaga, ZS11, Westar, and six de novo assemblies (Pacbio, HiC, Bionano). He has also talked to NRGene and might work on one genotype with them. Can only get to 1.2 Gb assembled, and he wants to increase genome size as much as possible.
- Rachel Wells: – ‘Order’ database which is a **transcriptome database**, ordered over time. This should be available soon and will have Westar and Tapidor data to browse initially. If anyone wants to beta test ‘Order’ then contact Rachel/Judith/Richard - it would be very much appreciated! There will also be ‘Bravo’ db, which is the same thing with more lines.
- Guusje Bonnema: Has **GBS data for 180 lines**, with a subset re-sequenced. This data will be available. She can't distribute seeds, but they are all from gene banks that can be accessed, or are hybrids that are available to purchase. All material is in principle available. Has a new project – 4 *oleracea* morphotypes that will have *de novo* assemblies and then crossed to study recombination across all different combinations and get idea of SV. Material will not be available, as companies made the crosses.
- Sarah Turner: Working on **225 *B. oleracea*** lines. Have glucosinolate and amino acid profiles, RNASeq data. This will be available through supplementary data
- Graham King: ***B. rapa*** genome – SCU Illumina + JIC PacBio data for **R-o-18** (used by JIC as TILLING population) and now **anchored** to dense 12,000 marker map – used RI population between Chiifu and R-o-18,– in collaboration with Kede Liu (HZAU).
- Isobel Parkin asked if we have consensus on **B genome nomenclature**?
 - Ian Bancroft suggested that we use Isobel's original nomenclature. His paper has been accepted on quality of genomes associated to linkage maps (suggest graphical genotypes as a check). *B. juncea* will now be v1.5. Ian's paper has information that suggests what to undertake in these QC checks for genomes.

Annaliese Mason asked if anyone minds that the centromeres are near the bottom.

Graham King indicated that it was agreed long ago

(<https://dx.doi.org/10.4226/47/5afb8519d194c>) that we wouldn't use cytology information.

- NB. brassica.info site can now **assign DOI to any of its page** as a reference (request initiated by Ian) allowing the page to be stable. Can even assign to data records in associated reference databases.
- Jun Zou: Have a **Ningyou7 assembly** in collaboration with Zhejiang. This will be the first genome to go through Ian's QC process. It is very well organised relative to TNDH linkage map.

- Jian Wu: There is now an **update of the Chiifu genome** to v.3 (was 2.5 in Dec), v3 will be registered, and is much better, it combines HiC, Bionano and PacBio. They are also doing a *de novo* assembly of *B. oleracea* with PacBio, on the same A2 accession as published by Shengyi Liu in 2014. Same strategy as for Chiifu will be used. There will be an additional 15 accessions *de novo* of *B. oleracea*
- Ian Bancroft asked if it was possible to make the Chiffu v3 **available for the QC?** It is already submitted, but GK suggests that it is done for Chiifu at the same time as Ro18. IB can fit this in reasonably soon.
- Eric Schranz: Has re-done a **genome of Aethionema** with PacBio. There are also **20 new BMAP** lines sequenced. If access is required, contact Chris Pires or Rod Wing. The paper is currently being written and some should be in JGI.
Action (ES): Send Graham King an update about the BMAP for Brassica.info (perhaps send the paper)
- Fan Liu: Have **introgression lines** between many different *B. oleracea* morphotype varieties in an introgressing population. Sequenced a cauliflower genotype, *de novo* assembly (PacBio with Bionano and HiC, very good results). Results reported on day 1 of the conference (Ning Guo). The purpose is to characterise the genetic background of introgression lines, and to study the mechanism of why the difference between cauliflower and heading cabbage.

Ian Bancroft asked if there was sequence data from the introgression lines. As there is, it will be possible to see what introgressions are represented (can do from DNA or RNA sequence).

- Graham King asks that **reference genomes** are **registered** with NCBI – this gives the genome a different status and it is then picked up by other database pipelines. Dave Edwards thinks there are pipelines available for this (he may have one available)

Action (GK): Add description of what is currently out there into the minutes

species	genome	genotype	version		size (Mbp)	coverage (%)	ena/ncbi assembly	Ensembl	
<i>B. rapa</i>	A	Chiifu-401	1.0	2011	115	92	GCA_000309985.1	IVFCAASv1	Xiaowu Wang
<i>B. rapa</i>	A	Chiifu-401	1.5		115	92	?		Xiaowu Wang
<i>B. rapa</i>	A	Chiifu-401	2.1	2017			GCA_000309985.2		Xiaowu Wang
<i>B. rapa</i>	A	Chiifu-401	2.5	12/2017					Xiaowu Wang
<i>B. rapa</i>	A	Z1	1		401		GCA_900412535.1		Genoscope
<i>B. oleracea</i>	C	TO1000	2.1	2014	489	75	GCA_000695525.1	v2.1	Isobel Parkin
<i>B. oleracea</i>	C	A2 (02-12)	1.0	2014	540	85	GCA_000604025.1		Xiaowu Wang or Shengyi?
<i>B. oleracea</i>	C	A2 (02-12)	1.1	2016					Xiaowu Wang or Shengyi?
<i>B. oleracea</i>	C	HDEM (wil)	1.0	2018	555		GCA_900416815.2		Genoscope
<i>B. napus</i>	AC	Darmor-bz	4.1	2014	850	79	GCA_000751015.1	AST_PRJEB5043_v1	Boulos/Genoscope
<i>B. napus</i>	AC	Darmor-bz	8.1	2017			not yet		Philipp Bayer/ Dave Edwards
<i>B. napus</i>	AC	Tapidor	1.0	2017			not yet		Philipp Bayer/ Dave Edwards
<i>B. napus</i>	AC	Zhongshuang 11		2017	976	82	GCA_000686985.2		BGI
<i>B. juncea</i>	AB	Tunida T84-64		2016	784	85	GCA_001687265.1		Zhejiang University
<i>B. nigra</i>	B	YZ12151		2016			GCA_001682895.1		Zhejiang University

- Ian B is about to submit a rearranged (**corrected**) genome version for *B. juncea* (Tunida), which is currently not accepted. Dave E has DOI for genomes which he knows will be updated soon as they won't go into NCBI, but is a fixed DOI, so always something to reference (is DOI per version).

4. Scope for multinational collaborative projects

Opportunities for co-funding and international funding were discussed.

- Ian Bancroft has a collaborative project with India on mustard rape, it is focussed on technology transfer and transfer of traits. The project should deliver a **diversity panel of *B. juncea***. Ian will re-sequence Surinder Banga's *B. juncea* radiation mapping panel. Can't guarantee material will be released internationally, but hope that there will be some, or at least it will provide information about the lines already available outside India

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

- Not much happened regarding BIS since January. GK has asked research groups that indicated resources, whether they have resources online that be searched; to date no one has anything ready at present. DE questioned whether Emsembl plants had any concrete plans. They will hire someone with responsibility to take over *Brassica* work as Paul and Dan are leaving).

Action (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 (Dave Edwards/Graham King): Determine whether the query page can be set-up in Brassica.info

- Dave Edwards: Has a **synteny viewer** (daisychain) – people can download and run themselves, DE wants feedback about how useful it is. He is happy for people to host versions. He also has a SNP database with Infinium data, which includes a lot of public data from supplementary files etc), he wants to know if it's of use, if it is, can we find a way to encourage people to upload?

Action (Dave Edwards): Send links for synteny viewer and SNP database to GK to post on Brassica.info.

- Rachel Wells asked if the Infinium data has all been through same pipeline? DE wants the raw data uploaded if possible, but it isn't currently. If the data is re-analysed then it will be uploaded under a different name. Rod S agrees it is a good idea to upload raw data. Annaliese M asked if it would upload information about species, population type etc? Isobel P mentioned that if uploading Idat files, then can also upload sample sheet.

6. Standards, ontologies and Brassica trait dictionary

- There is no update on BraTO.
- **Brassica Trait Dictionary:** This was managed by Wicktor. Rachel Wells did a lot of work trying to get it right. Anthony Hall is now nominated 'curator'. The pipeline to enter new trait ontology was clear. Currently it is more focussed on *B. napus*. Could be developed far more than has been to date.
- Isobel P asked if there was a **decision for naming genes?** Graham pointed out (again) there is a recommendation which is available online and has been agreed on in three previous

meetings. Correct nomenclature has not been used in some recent genomes (e.g Darmor v8) and we have a responsibility to ensure correct nomenclature is used. IP asked about the **problem of adding extra genes** and GK indicated that some stable genes should ideally be fixed between versions and new genes just get a unique identifier. DE stated we also need to make rules about translocations. GK questioned whether it should be a DOI per gene? ES asked whether we could just make links to the old gene names? GK suggested that it cannot only be a problem with Brassicas. ES suggested Solanaceae and Legumes could be good examples (tomato is good, potato was bad and now is good, so how did they solve it?).

- GK suggested that now we have reasonable coverage with later generation/version of genomes and an understanding of the pan-genome, we start implementing a rule-based system from 2019. IB suggested have a space of 10 between gene names to allow for new genes (as currently). DE suggested that as an interim there could be a look up table to enable ease of gene identification between different assemblies, his synteny viewer (daisychain) may be suitable.

Action: The issue will continue to be communicated. The community is open to clear models. Set target to come up with examples. Proposals/suggestions to be provided prior to, and then discussed and then to be agreed at next meeting (Jan 2019), and then disseminated.

7. Any other items/meeting announcements

- **Brassica 2020:** Mike Barker (with Chris Pires) to organise next meeting. Probably spring or autumn of 2020.
- GK questioned about the ISHS being involved in alternate meetings and they now seem to have set up independently! YPL is currently chair of the *Brassica* working group
Action: Yong Pyo Lim to try and resolve this at the ISHS meeting in Istanbul
- **Rapeseed Congress** June 16-19 2019, Berlin.
- **Polyploid meeting** 11-14 June 2019 Ghent (just before rapeseed congress)
- Oct 22-25 2018 **Crop genomics and heterosis** conference in HAU (<http://gcgi.hzau.edu.cn/>)
- **Agriculture and Climate Change** Meeting. -Budapest March 2019

Next MBGP Steering meeting – San Diego, @ PAG – 2pm Sun January 13th