

**Minutes of the Multinational *Brassica* Genome Project (MBGP)
Steering Committee meeting**

**PAG XVII, Windsor Rose Room, Town & Country Hotel, San Diego
Sunday 11 January 2009**

Chair: Isobel Parkin

Attendees:

Isobel Parkin	isobel.parkin@agr.gc.ca
Andrew Sharpe	Andrew.Sharpe@nrc-cnrc.gc.ca
Martin Trick	martin.trick@bbsrc.ac.uk
Faouzi Bekkaoui	Faouzi.Bekkaoui@nrc-cnrc.gc.ca
Raju Datla	Raju.Datla@nrc-cnrc.gc.ca
Ian Bancroft	ian.bancroft@bbsrc.ac.uk
Dave Edwards	Dave.Edwards@acpfg.com.au
Bart Lambert	bart.lambert@bayercropscience.com
Chris Town	cdtown@jcv.org
Chris Pires	piresjc@missouri.edu
Reno Pontarollo	rpontarollo@genomeprairie.ca
Shusei Sato	ssato@kazusa.or.jp
Pierre Fobert	Pierre.Fobert@nrc-cnrc.gc.ca
Jeff Parker	jparker@genomealberta.ca
Jerome Pauquet	Jerome.Pauquet@biogemma.com
Carlos Quiros	cfquiros@ucdavis.edu

1. Minutes of previous meeting

The minutes of the January 2008 MBGP meeting at PAG, together with those from the meeting held in Lillehammer in September 2008, had been previously circulated and these were formally approved. There were no matters arising not covered by the scheduled agenda items.

2. Inventory of public domain resources

Graham King (RRes, UK) had emailed a bulletin to the committee summarising the latest updates to the inventory on www.brassica.info.

3. Updates on various resources

- The *B. napus* diversity sets were progressing, with RRes finalising the MTAs. They would be available on a cost-recovery basis. Chris Pires said that he'd been quoted ~2000 USD (for 200 lines).
- The *B. oleracea* set was not yet available.
- Clarification was needed on the plans for a *B. rapa* set.
- The *B. rapa* TILLING population developed by Lars Ostergaard at JIC was now established. Screening would be available as a service from some time in 2009 (see www.revgen.co.uk for details).
- Isobel said that a *B. napus* TILLING population was being developed by George Haughn at UBC with Genome Canada funding.
- The status/availability of the new KBr BAC libraries (EcoRI and Sau3A) made by NIAB Korea for gap-filling was unsure at this time, as was that for a planned fosmid library.

4. Update on genomics funding

- Canada/Europe – An ERA-PG project to carry out eQTL and association mapping in *B. napus* has been funded, this will generate a public set of SNP markers for *B. napus*. A decision on a proposal for development of public access SNP markers was expected in March 2009.
- Australia – Jacqui Batley had been awarded a 5 yr programme grant for a project on Blackleg-host interactions.

5. Upcoming international meetings

The Canadian Plant Genomics meeting would take place August 24-27 2009 in Saskatoon (www.cpgw2009.ca) and the Crucifer Workshop meeting would be held 5-8 September 2010, also in Saskatoon. Isobel said that it was likely a one day genome sequence workshop would be held during the latter meeting. It was also mentioned that the IPMB meeting (resurrected from the old ISPMB) was to be in St Louis, September 2009.

6. Information dissemination

Biogemma had developed an Affymetrix chip based on the 95k JCVI/JIC Unigene set released into the public domain in September 2007. The chip design was proprietary but the company was open to collaborations in using the resource.

The issue was raised of the mapping between gene probes on the various platforms to enable comparability studies to be made. This was an action outstanding from last year. Chris Town volunteered to take this on. The open Agilent and Combimatrix platforms could be reconciled quite easily.

7. AOB - Chinese Illumina sequencing initiative

It was decided to discuss the item during the MBGP, ahead of the BrGSP meeting to follow.

Professor Xiaowu Wang of Beijing CAAS Institute of Vegetables and Flowers (IVF) in Beijing had been due to present to the committee a joint programme between IVF, CAAS OCRI (Oil Crop Research Institute) and the BGI to produce a Chiifu sequence by WGS assembly using Illumina reads (and the public BES data). Unfortunately Prof Wang could not attend due to visa problems but had sent his presentation for consideration by the committee members.

- The work followed the highly successful 358 Mb assembly of the cucumber genome by BGI using the same methods.
- Preliminary results from the *rapa* sequencing were obtained from three size libraries (150-200 bp PE 15 Gb 30X; 400-500 bp PE 7.8 Gb 15X; 2 kb PE 3.6 Gb 7X).
- At this combined 50X coverage, the N50 of fully resolved contigs was 4.3 kb (longest 45 kb) the global N50 (allowing gaps) was 30.7 kb (longest 294 kb).
- Using a reconstruction of 17-mer depths at varying coverages of *Erwinia* (of known genome size) the *rapa* genome at 50X was estimated to be somewhat smaller than other measures at ~ 500 Mb.

- Collaboration/contributions were invited, particularly in re-sequencing and transcriptome sequencing.
- Notice was given of upcoming sequencing projects – *B. oleracea* to be started by IVF/OCRI/BGI early 2009 and later *B. napus* (resequencing based on the A and C reference genomes) by OCRI/BGI.

There was considerable concern and discussion on the impact this *B. rapa* project will have on the multinational clone-by-clone sequencing project. Crucial questions were the quality of the assemblies and the true degree of collaboration/integration that could be achieved with the BAC by BAC approach. It was decided that a meeting should be sought with the Chinese principals at the very earliest opportunity. Ian Bancroft offered to drive this forward.

There were no additional items to discuss. This meeting adjourned briefly after which it resumed with a meeting of the BrGSP committee.

Minutes of the *Brassica rapa* Genome Sequencing Project (BrGSP) steering committee meeting

PAG XVII, Windsor Rose Room, Town & Country Hotel, San Diego

Sunday 11 January 2009

Chair: David Edwards

Attendees:

Isobel Parkin	isobel.parkin@agr.gc.ca
Andrew Sharpe	Andrew.Sharpe@nrc-cnrc.gc.ca
Martin Trick	martin.trick@bbsrc.ac.uk
Faouzi Bekkaoui	Faouzi.Bekkaoui@nrc-cnrc.gc.ca
Raju Datla	Raju.Datla@nrc-cnrc.gc.ca
Ian Bancroft	ian.bancroft@bbsrc.ac.uk
Dave Edwards	Dave.Edwards@acpfg.com.au
Bart Lambert	bart.lambert@bayercropscience.com
Chris Town	cdtown@jcv.org
Chris Pires	piresjc@missouri.edu
Reno Pontarollo	rpontarollo@genomeprairie.ca
Shusei Sato	ssato@kazusa.or.jp
Pierre Fobert	Pierre.Fobert@nrc-cnrc.gc.ca
Jeff Parker	jparker@genomealberta.ca
Jerome Pauquet	Jerome.Pauquet@biogemma.com
Carlos Quiros	cfquiros@ucdavis.edu

1. Minutes

The BrGSP minutes from PAG XVI and the *ad hoc* meeting at Lillehammer were tabled and both were agreed.

2. Actions from last meeting

- R7 status - Tim Sawbridge from DPI Australia was not present and so there could be no direct update on the progress of the R7 sequencing. It was noted that the www.brassica.info website had not been updated with the clones that had been sequenced by the Australian programme – this would be very useful to the community. It could not be confirmed but it was suspected that ~400 BACs had been pooled without indexing and thus the contigs could not be deconvoluted from the 454 data. The basis of their original assignment to linkage group A7 was also unclear – apparently FISH mapping undertaken by Chris Pires had shown a proportion were from elsewhere. It was agreed that rapid publication of the sequence data was needed, apparently the Australian funding agency had recently insisted on this.
- Korean funding – it was reported that the NIAB Brassica sequencing budget had been halved alongside a major restructuring of the

institute. Therefore there would be no Korean pickup of unallocated chromosomes and the actual completion of A3 and A9 could also be compromised.

- FPC data - Ian Bancroft had written to Beom-Seok requesting a copy of the FPC data file to disseminate via WebFPC from brassica.bbsrc.ac.uk as there had been some difficulty reported in accessing this resource from the Korean webserver.
- Genetic mapping data – some data had been received from the Korean group and had been placed on www.brassica.info, the raw scoring strings had not been supplied however. Ian Bancroft advised that there were numerous differences between the NIAB and CNU mapping data – Yong-Pyo Lim is now attempting to rigorously follow the Chiifu alleles and this was raising some issues apparently.
- Chris Pires asked why the sequence groups could not start sequencing BACs irrespective of chromosome. Ian Bancroft replied that this strategy would require real time updating of the BAC registry and, more importantly, immediate availability of sequence to all partners. He would ask this of Korea as, essentially, this would indeed be an effective way of working.
- Trace files – Martin Trick said that trace file submission for the UK/China BACs was imminent – it had been a protracted iteration with NCBI over the precise format coupled with difficulties in getting the relevant metadata from BGI. Andy Sharpe said that there were now defined standards for submitting 454 data, both assembled and short read data.
- Annotation – there was a request for a technical document to be written on the annotation pipeline developed by JIC. There were some details to be found on the GBrowse track listings and a readme on the Brassica Gateway site but Martin Trick would write an overview. Dave Edwards also asked for the SNAP gene prediction Brassica parameter file. It was agreed that this would be supplied.
- Graham King had received scoring data from some groups but not all – and not the Australians. The Australian seed BACs should be published.

2. Update from Lillehammer – nothing apart from an earlier announcement regarding the likely start of the Chinese Illumina initiative (now underway).

3. Sequencing progress

- **Korea** – in the absence of any representative from the Korean groups Ian Bancroft had given a summary at the Brassica workshop. Aspects of the programme are planned to continue until 2012.
A3 – 381 sequenced BACs formed 7 scaffolds spanning 34.9 Mb, with an estimated 4Mb left in gaps.
A9 – 289 sequenced BACs formed 15 scaffolds spanning 28.5 Mb, with an estimated 12.5 Mb left in gaps.
- **UK/China** – progress at end-December 2008

- 198 BACs sequenced to Phase 2 standard
 - 61 putatively on A1
 - 42 putatively on A8
 - 7 re-mapped top other linkage groups
 - 88 new seed BACs (to complement existing set)
- 24.1 Mb submitted to public databases in all
- 15 potential new seed BACs completed to Phase 1
- 74,607 *B. napus* BES submitted

It was noted that CE phase 1 in combination with Illumina PE data could produce phase2/3 data.

- **Canada** – Andrew Sharpe reported that 46 BACs from A2/A10 had been sequenced by tagging and pooling on a 454 instrument. The sequences had been submitted to GenBank as Phase 1, but enhanced to “Phase collinear” by rearranging contigs to maximise synteny with Arabidopsis.
- **Australia** – the situation with regard to the A7 BACs was summarised by Dave Edwards in the absence of a representative from DPI Victoria. It would seem that the extra Sanger CE sequencing required to deconvolute the pooled (and unindexed) 454 data had not been done. Data in this form could not be accepted as HTG by GenBank. It was suggested that the new Chair of this committee should write to DPI to ask for the raw data. Dave has produced a substantial quantity of *B. rapa* sequence using the Illumina GAII and has developed software to use this data to bring phase 1 BACs to phase 2 and also correct gaps in phase 2 BAC data.

4. Updates on sequencing project funding

- India A6 – may resubmit their proposal with a revised, substantially lower costing for outsourcing of sequencing (via Ian Bancroft, from TGAC UK).
- France A4 – Boulos Chalhoub has had a very positive response to a pre-proposal and was invited to submit a full proposal. Given this, it seemed funding was quite likely.
- Japan A5 - (Dr Shusei Sato, Kasuza Institute) – has received funding to sequence 150 BACs by Sanger CE. A list of adopted clones will be sent to www.brassica.info.

5. Discussion of sequencing approach

There was a lengthy discussion on the progress of the international project, which had further to be re-evaluated in the light of the Chinese WGS initiative. It was accepted by all that it was behind schedule due to a number of reasons. The shortage of equally spaced seed BACs and a paucity of extension clone candidates, uncertainty over genetic mapping data made everyone agree that it was time to switch from a chromosome-based allocation to a clone-based one, notwithstanding the potential impact from the Illumina data. An attractive possibility was to use the

Korean group's FPC contigs as a basis of distributing non-overlapping BACs to the various groups. It would follow that the identities of the FPC contigs used by the Koreans, together with their sequence data, would need to be obtained. The new Chair should write to seek this.

NGS will facilitate finishing of partially completed chromosomes. Deep transcriptome sequencing would discover genes not predicted from completed BACs.

It was also agreed that more frequent meetings between the active sequencing partners by video or phone conference, combined with rapid and coordinated sharing of data would be required to make this a success. It would be incumbent on the new Chair to organise this.

6. AOB

Chris Pires asked about the status of *Brassica oleracea* sequencing and should he continue with his selected BAC approach? The genotype that BGI intended to sequence was not known.

Andrew Sharpe said that a WGS of the TO1000 line could be undertaken by 454 Titanium.

Chair of committee

After serving for two years, Dave Edwards stepped down and was thanked by all for his efforts. Ian Bancroft was nominated to take over and this was agreed by all.