

**Draft minutes from the Multinational *Brassica* Genome Sequencing
Project steering committee meeting.**

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

Sunday 13 January 2008

Chair: David Edwards

List of Attendees:

Name	Email address
Isobel Parkin	parkini@agr.gc.ca
Andrew Sharpe	sharpea@agr.gc.ca
Andrew Sharpe	andrew.sharpe@nrc-cnrc.gc.ca
Faouzi Bekkaoui	faouzi.bekkaoui@nrc-cnrc.gc.ca
Raju Datla	raju.datla@nrc-cnrc.gc.ca
Graham King	graham.king@bbsrc.ac.uk
Chris Town	cdtown@jcv.org
Martin Trick	martin.trick@bbsrc.ac.uk
Ian Bancroft	ian.bancroft@bbsrc.ac.uk
Yong Pyo Lim	yplim@cnu.ac.kr
Dan Qiu	dqui@ucdavis.edu
Dave Edwards	dave.edwards@uq.edu.au
Jacqui Batley	j.batley@uq.edu.au
Yeon-Ki Kim	kim750a11@gmail.com
Genyi Li	g.li@umannitoba.ca
Chunyu Zhang	zhchy@mail.hzau.edu.cn
Cyril Falentin	cyril.falentin@rennes.inra.fr
Regine Delourme	regine.delourme@rennes.inra.fr
Carlos Quiros	cfquiros@ucdavis.edu
Gopalan Selvaraj	gopalan.selvaraj@nrc-cnrc.gc.ca
Pierre Forbert	pierre.forbert@nrc-cnrc.gc.ca
Reno Pontarollo	rpontarollo@genomeprairie.ca
J.Chris Pires	piresjc@missouri.edu
Jung Sun Kim	kimjs@missouri.edu
Jung Sun Kim	jsnkim@rda.go.kr
Ping Lou	ping.lou@dartmouth.edu
Guusje Bonnema	guusje.bonnema@wur.nl
Nathalie Riviere	nathalie.riviere@biogemma.com
Su Ryun Choi	srchoi@cnu.ac.kr
Jeong-Hwan Mun	munjh@rda.go.kr
Beom-Seok Park	pbeom@rda.go.kr
Noel Cogan	noel.cogan@dpi.vic.gov.au
Bernd Weisshaar	bernd.weisshaar@uni.bielefeld.de
Rod Snowdon	rod.snowdon@agr.uni-giessen.de
Tim Sawbridge	tim.sawbridge@dpi.vic.gov.au
Benjamin Laga	benjamin.laga@bayercropscience.com
Jerome Konecsni	jkonecsni@genomeprairie.ca

A sequencing update was provided during the MBGP meeting immediately prior to the MBGSP meeting. For clarity, the relevant section of the MBGP minutes is included below.

From MBGP minutes:

6. Sequencing update

More technical discussions would follow in the BrGSP meeting. Here, each group gave a brief status report.

- **Korea (NIAB)** – 93% of the 99,456 BACs from the H, B and S libraries had been fingerprinted by the SnapShot method. After removal of repetitive clones, 67,468 BACs were initially assembled with FPC into 4,726 contigs and then iteratively merged into 1,417 contigs. The FPC build comprises 1,428 contigs spanning 557 Mb (~1.5x genespace coverage). This was made available via WebFPC at www.brassica-rapa.org in June 2007. Build 3, using an additional 30,000 clones from a new library, is planned for 2008. A manuscript reporting the physical map is under review by BMC Genomics. Some experiments using 454 on 39 BACs proved OK with supplementary 2-4 x Sanger CE, was considered too costly at present.
- **UK/China/US** – the project officially started in July 2007 with the assigned chromosomes A1 and A8, using the original set of 523 seed BACs as starting material. The first 34 BACs sequenced had been recently annotated, these data being available now on brassica.bbsrc.ac.uk. This first batch comprised a mixture of extension clones for BACs robustly mapped to A1 or A8 and candidate rearrangement clones, identified using a novel bioinformatics approach. The quality control/validation was being carried out now. If passed, the sequences would be submitted to GenBank/EMBL by BGI as soon as possible. It was noted that the cost to the project was ~ USD1000 per clone finished to Phase 2. BGI was also about to embark on the end sequencing of the *B. napus* Tapidor library, data expected by mid-2008.
- **Canada** – had already sequenced ~30 BACs (using Sanger CE) as a pilot experiment. These data will be submitted to GenBank/EMBL soon. Funding had now been obtained to start the A2/A10 sequencing, between AAFC & PBI Saskatoon, will use the 454 FLX platform combined with Sanger CE–will start in Spring 2008. The strategy to sequence the two B. rapa chromosome will be similar to the Australian group. It is expected that the project would take 2 years to complete, pending additional funding.
- **Australia** – the A7 sequencing had been picked up by Tim Sawbridge following Dave Edwards' departure. 99 BACs had already been sequenced, with data at various stages, a further 22 were underway and 40-60 were yet to come. FLX reads indexed against Sanger CE was the adopted technology. 12 scaffolds had been FISH-painted. Older GS20 data needed to be de-convoluted. The release policy for all these data seemed problematic and was to be discussed end Feb 2008.

This leaves 3 *B. rapa* chromosomes unallocated. Chris Town (JCVI, US) would not be resubmitting a US proposal – the NSF had been very cool on the previous one, the only possible route might be via *B. oleracea* sequencing. There was no news from France (although Boulos Chaloub had previously thought something might be possible) and Guusje said there was no prospect of funding in the Netherlands. Germany had tried several times before and could try again in March 2008 (maybe a Franco-German collaboration?). India perhaps could prepare a proposal for two chromosomes. Yong-Pyo Lim said that there was a possibility of Korea funding a further 3 chromosomes – the current funding ceases in December 2008, he was currently at a very critical time in discussions with maybe some news in May/June 08. *All the attendees offered letters of support if this would help.*

Following are the minutes of MBGSP 13/1/08

Progress Updates:

1. Sequencing progress

Australia – Update from Tim Sawbridge: Funds are available to sequence a further 60-80 BACs based on the 454 FLX platform (3-4 more plates at 20 BACs per plate – 5 BACs per 4 gaskets). One Sanger library per BAC will be prepared and sequenced one plate at a time until they have enough sequence to assemble. The Korean project has also sequenced a further 50 seed BACs on A7 with 23 finished. Selection of BACs is based on overlap with FISH mapped BACs. Some (12) of the scaffolds have been FISH mapped to other chromosomes and the location of these BACs will be released as seed BACs for other groups.

Action: Tim Sawbridge to distribute names and locations of 12 non-A7 BACs.

Some scaffolds have no FISH mapped BACs, some of these will be sent to Chris Pires for FISH mapping and these are also being genetically mapped. The BACs being sequenced and in progress will be updated on Brassica.info.

Action: Tim Sawbridge to update BAC registry at Brassica.info with sequencing progress.

BAC selection is proving difficult when we can't find single BAC-end with 99% confidence in contig. Update may be provided in Norway (dependent on funding?). Need to do additional sequencing on "first" BACs to link to BAC – disorientated contigs. Re-assembling as Newbler comes out with a new version.

Canada – Expect a 454 FLX machine to arrive next month. They have identified BACs overlapping with seed BACs on A2 and A10, through assessment of new technology. These have been assessed through FISH mapping with Chris Pires, in order to check that identification method is

working. Expect SOLiD and 454 data for BACs (don't have SOLiD yet), have Sanger data for them all. First 30 BACs sequenced using Sanger will be out soon after finishing and QC. A list of the 30 BACs being sequenced has been sent to Graham King. Need the BAC repository to be up to date, as the Korean project appears to have sequenced some on A2 and A10 (e.g flowering time regions etc), so a list of these BACs need to be in repository. Strategy – using the 454 will use either gaskets or index system which is due to be released (as this will lead to greater throughput). May use a similar plan to Australia, think they will use the indexing, but need to see when it will be available. No modelling done on potentially longer 454 reads (400 bp) to assess how much sequencing required.

China/UK – The seed BAC sequences need to be in Genbank so that they can be used for assembly by other groups. Chris Town noted that even phase 1 will be useful for this. Progress covered in the MBGP meeting (see above). 173 BACs have been sent to BGI for sequencing, including ca. 80 BACs additional seed BACs. These were identified as potentially containing re-arrangements (bioinformatics approach using Arabidopsis – the BAC end sequences showing homology to different Arabidopsis chromosomes or the same chromosome, but strand is switched), and are being mapped by Guy Barker – all are listed in the repository. Continuing with Sanger sequencing in China as this is the most cost-effective (ca. \$1000/BAC funded by the BBSRC as part of the UK-China collaboration).

Korea – Update from Beom-Seok Park: Have sequence and fingerprint data for 2 chromosomes. Two BAC libraries are not enough for chromosome walking. Additional *Sau3AI* library ordered from Lucigen last spring, but problems with it. When the new library arrives (approx. 30,000 clones), the clones will be end-sequenced, fingerprinted and the information provided to the community. Must finish A3 and A9 by end of the year and currently trying to get funding for a further 4 or 5 years to continue sequencing. Expect a decision between May-July. (see also discussion on Release Policy)

Action: Beom-Seok Park to inform the steering group of the decision when available.

Ian would like raw data for FPC as can't search the Korean website for BAC names (seems to a BBSRC problem).

Action: Beom-Seok Park to try to provide the data.

Trying to isolate a single chromosome for phase 3 completion using SOLiD sequencing – are there any other groups who can isolate a chromosome? No known method may be able to use flow cytometry to isolate the largest chromosome, but otherwise problems of identifying which chromosome is which. Contamination is also a large issue. Chris Pires is interested in following this up, but the task is very difficult.

Dave Edwards – is there advantage to shot-gun sequence single chromosome over whole genome? We are sequencing 3-6x shot-gun of Chifu

in the next few months. Ian Bancroft suggests that this won't provide the data for the missing regions. Dave Edwards – suggests that it may help with regions that can't be cloned and can assist to order and orientate contigs.

2. BAC registry

Graham King - BACs genetically mapped to multiple populations can be priority seed BACs as they provide higher level of confidence in their location.

Action: Colleagues at NIAB and CNU Korea to send all mapping data to Graham King.

Ian Bancroft – single polymorphic band in one population may have paralogous loci, as they map to a different place in other populations. It was suggested that this may be due to PCR amplification dominance effects.

Action: Groups to provide all mapping information where possible.

Progress on the BAC registry: Monthly updates are to be provided from chromosome specific projects, checking data with seed BACs, and adding map data for 2 populations. FISH data is to be included in the future. After updating, the original data will be kept on archive, but deleted from public view, as will changes from in progress to completed etc.

3. Trace file repository

Agreement was that everyone would deposit trace files, so people can pick up and finish BACs they are interested in. Need a procedure and nomenclature. Sequences should go to the GenBank repository. Korean trace files and sequence of approx. 500 BACs submitted.

Action: Beom-Seok Park to distribute GenBank ID's of trace files.

4. Access to fingerprint data

See above

5. Funding updates

Covered in MBGP

Discussion:

6. Distribution of annotation activities

Annotation - Chris Town – doesn't have funding to annotate all rapa BACs. Are they all going to be annotated by same pipeline? It is important that annotators use the same parameters etc. in the different pipelines that people have to ensure people have same gene models. Ian Bancroft is setting up an automated method which will pick everyone's BACs from GenBank and annotate them, but this will be a first pass only. One pipeline for all BACs will be best outcome and can go between groups for different processes. Chris Town – may be able to distribute gene modeller, though may need licenses

for some of the software. Need to ensure that everyone has same data to reference against – ESTs etc. if are annotating in-house.
Methodology must be open and clearly defined.
Need to define how gene will be named etc.

Action: Chris Town will write up annotation process and make public.

7. Data release policy

UK/China adhering to release policy

Dave Edwards checked with GenBank about “sequencing in progress” statement, suggested formulate statement in the annotation section that the BAC is only being sequenced to phase 2 – we need uniform statement,

Action: Dave Edwards to draft and distribute a statement.

‘BACs sequenced as part of the Multinational Brassica rapa Genome Sequencing Project are assembled to phase 2 (ordered and orientated contigs) prior to release. We welcome collaborators who wish to finish these BACs to complete phase 3 assemblies. Please contact the owner of the BAC sequence accession. Details are available on www.brassica.info.’

Tim Sawbridge – will meet with GRDC in February to define release policy for BAC sequences. Tim Sawbridge states only 1 BAC released before he started (GenBank records 5).

Beom-Seok Park – over 500 BACs released, aiming for immediate release. Release of data by other groups will help. Workshop in Korea in December, hope to invite 2-3 from committee.

8. Deployment of next generation sequencing

Any new technology which may be relevant. UK/China will change if cheaper method comes along. Solexa, currently approx 300-600 bp mate pairs, trying 1-3 Kb. Dave Edwards testing SOLiD system, perhaps use throughput of 454 with this. Single molecule sequencing – no real information is available for this.

Benjamin Laga– Bayer interested in sequencing *B. napus* by a BAC-by-BAC approach using next-gen sequencing. The sequence will not be public, but may be open to use in common interest projects with individual groups. Sequencing not to phase 3 level, just a rough draft so cost-effective. Spring oilseed rape selected.

9. Robustness of genetic anchoring of seed BACs

Confidence could be improved for Ian Bancroft by looking at the gels/electropherograms *etc.* NIAB have provided information to Ian and may be able to provide to Isobel Parkin. No information from Australia about the BACs they are anchoring (Noel Cogan had left the meeting by this point). Graham King needs scoring data for bin mapping.

Action: scoring data to be provided to Graham where possible. (ALL)

10. AOB

Dave Edwards – chair only for 1 year initially, happy to do second year. – supported.

Actions:

Aim to have all seed BAC sequences submitted to GenBank (ALL)

Next meetings

Norway 2008?

PAG 2009