Multinational *Brassica* Genome Project Steering Committee meeting

10am Sunday 14 January 2007
Crescent Room, Town & Country Hotel, San Diego
Chair: Graham King

PARTICIPANTS (INCOMPLETE LIST)

MEMBERS

Graham King *Chair MGBP Steering Committee 2005-*	graham king@bbsrc.ac.uk	UK
Dave Edwards	Dave.Edwards@dpi.vic.gov.au	Australia
Jacqui Batley	Jacqueline.Batley@dpi.vic.gov.au	Australia
Gijs van Rooijen	vanrooijen@genomeprairie.ca	Canada
Bin Liu	liub@genomics.org.cn	China
Boulos Chalhoub	chalhoub@evry.inra.fr	France
Beom-Seok Park	pbeom@rda.go.kr	Korea
Myung Ho Lim	mlim312@rda.go.kr	Korea
Yong Pyo Lim	yplim@cnu.ac.kr	Korea
Derek Lydiate	lydiated@agr.gc.ca	Canada
Isobel Parkin	parkini@agr.gc.ca	Canada
Andy Paterson	paterson@dogwood.botany.uga.edu	USA
Chris Town	cdtown@tigr.org	USA
Govind Garg	govindgarg@krishidhanseeds.net	India

Other delegates:

Andy Sharpe	sharpea@agr.gc.ca	Canada
Raju Datla	Raju.datla@nrc-cnrc.gc.ca	Canada
Gopalan Selavaraj	Gopalan.selvaraj@nrc-cnrc.gc.ca	Canada
Faouzi Bekkaoui	Faouzi.bekkaoui@nrc-cnrc.gc.ca	Canada
Larry Pelcher	Larry.pelcher@nrc-cnrc.gc.ca	Canada
J. Chris Pires	piresjc@missouri.edu	USA
Anker Sørensen	anker.sorensen@keygene.com	NL
Jung Sun Kim	junkim@rda.go.kr	Korea
Yeon-Ki Kim	Kim250a11@gmail.com	Korea
Matthew Nelson	mnelson@plants.uwa.edu.au	Australia
Robert Tran	RKTRAN@UCDAVIS.edu	USA
Ian Bancroft	Ian.bancroft@bbsrc.ac.uk	UK
Federico L Iniguez-Luy	fliniguezluy@wisc.edu	USA
Gohataj	gohartajkhan@rediffmail.com	India
Rubens N. Tomita	rubens@affrc.go.jp	Japan
Christian Obermeier	Christian.obermeier@agrar-uni-giessen.de	Germany
Tae-Jin Yang	tjyang@snu.ac.kr	Korea
Soo-Jin Kwon	sjkwon@rda.go.kr	Korea
Mina Jin	genemina@rda.go.kr	Korea
Carlos Quiros	CFQUIROS@UCDAVIS.edu	USA
Reno Pontarollo	rpontarollo@genomeprairie.ca	Canada
Martin Trick	martin.trick@bbsrc.ac.uk	UK
Graham King	graham.king@bbsrc.ac.uk	UK
Lars Østergaard	Lars.ostergaard@bbsrc.ac.uk	UK

Participants briefly introduced themselves.

1. Minutes of previous meeting, matters arising

Minutes of previous meeting were approved without revisions.

2. Inventory of public domain resources

Graham King (GK) sent out a questionnaire to Brassica researchers in February 2006, and then in December 2006. He received few replies that will be compiled and shared with the Brassica research community. GK enquired if there are other resources the participants wish to share. He has accumulated additional information about other resources (including from original survey 2002), but he will contact the relevant people first to ensure that the information is accurate and up to date before dissemination. Graham indicated that Rothamsted Research is hiring a Brassica data curator (part-funded from BBSRC AdVaB project), with start date in March or April. The curator will get in contact with contributors to update relevant information on the Brassica.info website.

3. Data and nomenclature standards

As a result of the increasing convergence of information aligning linkage maps, chromosomes and genomic sequences, it was agreed that it was timely to assign consistent chromosome/linkage group nomenclature to the canonical diploid *Brassica* genomes in the 'triangle of U' as: 'A', 'B' and 'C'.

	A	**B**	**C**
	A01 - A10	**B01-B08**	**C01 - C09**
B. rapa	R1-R10		
B. nigra		G1-G8	
B. oleracea			O1-O9
B. juncea	J1-J10	J11-J18	
B. napus	N1-N10		N11-N19
B. carinata		?	?

In order to ensure that this is universally agreed and adopted, it was agreed that it would then be necessary to a) publish a short paper describing the definitive nomenclature system in a recognised journal; b) notify key Journal editors; and c) also to publicise this more generally on websites/mailing lists etc.

- Derek Lydiate agreed to follow-up with concerned participants to draft the note on nomenclature for publication.
- This would include collation of existing major maps where available, and describe the resources necessary and available to verify chromosome identity. This would include the relevant populations and key sequence-tagged markers/allelic information required to verify.
- Derek would contact the Rennes group to confirm their agreement, and also to enquire re: status of existing *B. juncea* map/population (Lionneton *et al.*, 2002).
- Andrew Paterson suggested that any publication should include a translation table relating to existing published maps where these had been integrated.

4. Diversity sets

This topic was addressed in the TILLING meeting.

B. oleracea

Warwick-HRI (UK) developed *B. oleracea* fixed lines. The process is being finished though it is taking longer than expected. WHRI is working on 188 lines.
Seed and DNA available for founder lines. Contact: graham.teakle@warwick.ac.uk

B. napus

188 fixed lines are being developed at HRI in collaboration with Rothamsted Research (UK) as part of the OREGIN project. DNA is available publicly. 50 of 94 lines seeds are available to date.
Seed and DNA available for founder lines. Contact: graham.teakle@warwick.ac.uk

B. rapa

Guusje Bonnema (WUR, The Netherlands)/Xiaowu Wang (CAAS, Beijing) and Graham King (Rothamsted, UK) have identified 188 *B. rapa* founder lines. Graham invited people interested in helping with process or those developing diversity lines to contact him or contact Guusje Bonnema to update information on resources. Dave Edwards indicated that *B. napus* and *B. juncea* diversity lines are also available from Australia.

In the wider context: Chirs Pires is assembling a collection of material representing a wider distribution within the tribe Brassicaceae.

5. Report on TILLING consortia

The MBGP TILLING consortia held a meeting prior to this one. Lars Ostergaard gave an update on the TILLING consortia including consortia members, aims, distribution of resources and web design. Minutes from the meeting and contact info will be posted on the TILLING website. (www.brassica.info/resources/tilling/mbgptc.htm)

6. ESTs update

Currently there are approx. 163,750 Brassica ESTs in GenBank, an increase of ~40,000 during 2006. The number of Brassica ESTs to be available in GenBank will increase significantly during 2007, as a result of discussions during 2006 associated with providing sufficient information to develop a good representative transcriptional array. By June 2007, there will be 447,000 *B. napus* ESTs and a total of 623,000 total Brassica ESTs from *B. napus, B. rapa, B. juncea* and *B. carinata.*

GK indicated that there will be a link on the brassica.info website to provide updates on Brassica ESTs (http://www.brassica.info/mbgp/mbgp1.htm). Martin Trick agreed to provide additional link to current sequence data held in BrassicaDB (which also has a BLAST server that allows any sequence to be searched against all current *Brassica* sequences, or the Arabidopsis genome). GK enquired if the participants are aware of other institutions that have not deposited Brassica ESTs in GenBank yet. He suggested that private companies (e.g. Bayer, BASF, Dow) may now be in a position to deposit ESTs now that there are is significant and good representation of sequences deposited by other institutions and consortia.

7. Transcriptional arrays: MBGP Affymetrix GeneChip®, and others

Graham reported that during 2006 there had been considerable discussion about the design and preparation of the Brassica Community Affymetrix GeneChip®. In order to ensure that sufficient ESTs were available for the design, the original timeline was extended. This has allowed several groups to adapt their release programme into the public domain, with a target date of June 2007. He informed the participants about

the Affymetrix meeting at PAG on Tuesday 8 pm. He invited the participants who are interested in the Brassica chip to interact with Affymetrix during the design to ensure a good quality chip.

Participants suggested inclusion of probes sequences from pathogen species and other probes to distinguish Brassica paralogues and/or homoeologues as a separate set. It was agreed that these sequences could be included if there is space on the 45k chip and if it does not increase significantly the cost of the Brassica array. This was discussed further in the subsequent meeting with Affymetrix (notes of meetings to be posted at (www.brassica.info/genechip/genechip1.htm)

8. Update on sequencing and genomics funding

Technical details on the *B. rapa* sequencing will be discussed at the meeting following this one. Yong Pyo Lim reported that the first phase of the BrGSP project was completed during 2006.

This has included:

- 200,000 BAC ends sequences from 110,000 clones have been completed thanks to the contributing institutions from: Australia, Canada, Germany, Korea, USA and UK.
- 510 + 60 seed BACs have been sequenced (Korea, with a few from Australia)

BAC by BAC sequencing of individual chromosomes has now commenced:

R1	UK + China consortium, funded, 2007-2010
R3	Korea, funded, completion end 2007
R6	USA, proposal submitted, awaiting outcome N. hemisphere spring 2007
R7	Australia, funded, complete end 2007 (using 454 shotgun of BAC pools + Sanger finishing)
R8	UK + China consortium, funded, 2007-2010
R9	Korea, funded, complete end 2008

(status at http://www.brassica.info/b_rapa_sequencing_project/bac_sequencing.htm)

- Dave Edwards (Australia) indicated that Australia has completed the sequencing so far of 68 BACs, and has to complete the remaining BACs by the end of this year.

- GK suggested compiling the DNA sequencing statistics and posting on the Brassica.info website.

- USA: Chris Town indicated that sequencing of R6 chromosome proposal was submitted. The review panel will meet by the end of this month. He will update the review panel about the recent deposition of Brassica ESTs in GenBank which will likely help the proposal.

- Canada: Isobel Parkin indicated that no funding is available yet to sequence R2 and R10. However, a proposal will be submitted in 2007 with possible funding start in 2007 if proposal is successful.

- France: Boulos Chalhoub indicated that he is a funding call, and will be submitting a proposal in May 2007 to sequence 200, BACs with funding

results to be announced in September. Information on BAC positions within chromosomes was not available.

- Germany: no update was available. (check Bernd)

- India: Gocind Garg indicated that India is interested in contributing to the sequencing effort. Gopalan Selvaraj indicated that he visited recently India and funding of the sequencing of 1 chromosome is promising. It was suggested to contact Dr M.K. Bahn, Secretary, Department of Biotechnolgy, Govt. Of India, CGO ComplexII, Lodhi Road, New Delhi.

Other sequencing efforts:

- UK: Ian Bancroft informed the participants that the UK-China consortium will sequence 72,000 *B. napus* BAC ends (in pSAC vector), and make these publicly available during 2007.

- USA: Chris Pires indicated that his group have received NSF funding to sequence x BACs from *B oleracea* as part of a comparative analysis

9. Summary of publicly available data from BrGSP (see also BrGSP management committee following this meeting)

This item was discussed in the following meeting.

10. Upcoming international meetings

- Rapeseed Congress, Wuhan, China in March 2007
- PAG, San Diego, USA, January 2008
- ISHS Brassica meeting & Crucifer Genetics Workshop in Norway, Sep 2008 (Scientific Committee and Programme to be announced March 2007).
- Crucifer Genetics Workshop in Saskatoon 2010

11. Information dissemination

- As indicated above, the BBSRC 'AdVaB' and sequencing projects in the UK will provide a data curator to update the brassica.info website. This work will include updating the inventory of public domain resources, as well as data for genetic map integration and consensus markers. Discussions are ongoing amongst several groups (e.g. AgVic, Melbourne - BASC; Georgia -) for collating brassica data sources, and to make sure there are definitive data sources.

- It was suggested that a mailing/contact list be made available. For this to be possible, each member of existing list would need to agree for their information to be public. GJK indicated that this may be possible to manage as part of the data curator work at RRes.

- The possibility of collating bibliographic information was suggested. However, this is an onerous task and probably sufficient added value compared to individuals carrying out online searches. There may be several text-mining approaches that could be suitable if funded. Alternatively, key papers could be collated and highlighted as part of the data collation on brassica.info relating to reference resources and information. The possibility of sharing End-Note libraries was suggested.

10. Election of Chair (agreed remit states every two years)
A committee chair is elected every 2 years. Ian Bancroft (UK) was the first chair of this committee. Graham King (UK, current chair, 2005-7) was due to step down, and called for nominations. Isobel Parkin (AAFC, Canada) was nominated. All participants were in favour of the nomination. Ian Bancroft and the participants thanked Graham King for serving as chair in the last 2 years.

11. AOB
none

Meeting was formally adjourned at 12:25 pm

Minutes prepared by GJK from notes kindly taken by Faouzi Bekkaoui

[BrGSP notes follow on page 7]

Brassica rapa Genome Sequencing Project (*Br*GSP)
Management Committee Meeting

Sunday 14 January 2007
Crescent Room, Town & Country Hotel, San Diego
Chair: Yong-Pyo Lim

This meeting followed on directly from that of the MBGP Steering Committee. Reporting and discussion addressed the agenda items tabled by Professor Lim.

1. Current status of sequencing project

The current funding status of the project was briefly discussed. Future plans and timings of key funding decisions would be discussed at the end of the meeting.

2. BAC end sequence data

All were now complete and deposited with GenBank/EMBL.

Plates	Clones	Groups
KBrH001-015	5,760	NIAB; S. Korea
KBrH016-050	13,440	NRC/AAFC; Canada
KBrH051-062	4,608	JIC; UK
KBrH063-087	9,600	DPI; Australia
KBrH088-117	11,520	JIC/Bath/TIGR; UK/US
KBrH118-136	7,296	U Bielefeld; Germany
KBrH137-144	3,072	CNU; S Korea
KBrB001-096	36,864	NIAB/CNU; S. Korea
KBrB097-132	13,824	NRC/AAFC; Canada
KBrS001-015	5,760	NIAB; S. Korea

Issues arising:

There was an issue with the clone nomenclature for the sequences submitted by DPI (a leading 1 inserted into the plate number component). This would be rectified as soon as possible. Also, some of the sequences submitted by JIC (those conducted at Sanger) had not been quality-trimmed. This would also be corrected as soon as JIC made an appointment on the project (summer 2007). There was an agreed commitment to submit trace files to the GenBank trace repository. Pre-computed *in silico* mappings to the Arabidopsis sequence and real time BLAST facilities for the BAC end dataset were now being offered by several groups.

3. Seed BAC sequencing

In Korea, 1,007 seed BACs were being sequenced; 514 BAC sequences had already been submitted to GenBank/EMBL. A further 62 were due to be submitted as soon as possible (subject to approval by the parent Institute) and 60 more would be added later in 2007. A list of these and their status are available at http://www.brassica-rapa.org.

4. Genetic and physical mapping to support sequencing project

CKDH reference map

- In the reference genetic map, there are a total of 10 linkage groups. The map consists of 278 AFLP, 236 SSR, 25 RAPD, 18 ESTP, 3 STS, 5 PCR-RFLP and 3 CAPS markers.
- So far, 1,039 genetic markers have been generated by the CNU group. These comprise 740 AFLP, 202 SSR, 38 RAPD, 22 ESTP, 4 STS, 20 PCR-RFLP and 13 CAPS markers.
- 151 SSR markers developed from 514 sequenced seed BACs have been mapped by CNU, these datasets will be sent to the labs doing the sequencing. Dave Edwards' group at DPI have developed a further 150 BAC end SSR markers
- During 2007, a further 500 SSR markers will be generated from the end sequences of the *Hin*dIII and *Bam*HI BAC clone libraries. In a pilot experiment, 16 SSR markers had been developed (a success rate of 30%).

Development of a new genetic map from CKRI population

- 153 F_8, 83 F_7, and 5 F_6 lines of CKRI population have been produced.
- Using 52 F_7 plants of CKRI, SSR markers located at regular intervals in CKDH map are being mapped in the CKRI population.

JWF3 population

- 390 of the seed BACs have been genetically mapped by NIAB (250 by SSR markers; 50 by RFLPs ; 20 by CAPS and 70 by Ctg
- Screening was being carried out initially in JWF3 then sent to YPL (CNU, Korea) for mapping in CKDH.
- SSRs from the 'Suwabe' map have been used to identify and integrate with established *B. rapa* linkage maps
- There are ~100 common markers between the CKDH and JWF3 maps, 10 per linkage group. One consensus map has been constructed using JoinMap and is available from the NIAB home page.

<u>Issues arising:</u>

Scoring strings and association of markers to seed BACs needed to be made available to Ian Bancroft (and others) to assist in BAC selection.

Who could support the genetic mapping? Graham King would be able to do so by the end of 2007, requiring collaboration with CNU and NIAB. A consensus map could be created by end of March 2007.

One issue was that marker names used by NIAB and CNU differed. A lookup table needed to be made available as soon as possible.

Data from the CKRI mapping needed to be released to the community.

There was a question regarding the locations of distal genetic markers on linkage groups. Physical locations had been estimated by NIAB using FISH.

Robust, single-locus SSR markers from the AAFC proprietary database could be mapped *in silico* to the seed BACs and the results, in principle, could be made

available to the consortium as background knowledge. Isobel Parkin would pursue this.

Construction of integrated physical and genetic maps

- 3,443 contigs have been generated by CNU, using automated contig assembly from agarose gel fingerprints of *Hin*dIII-BAC clones.
- 2674 contigs have been produced by NIAB and CNU, using automated contig assembly from 4-enzyme SNaPshot® fingerprinting of *Bam*HI clones.
- Physical mapping of *Hin*dIII clones has been completed by 4-enzyme SNaPshot® fingerprinting and will be released soon.
- Using BAC-based SSR or SNP mapping in CKDH and CKRI, several thousand contigs from the *Hin*dIII and *Bam*HI libraries will be merged.
- All FPC data will be made public by end of March 2007 via the WebFPC Java interface – currently there are ~70,000 fingerprints available

Issues arising:

It was noted that the Coral program, developed by GSC Vancouver, could be used to re-order the FPC generated contigs. This had been used with success in the BBSRC IGF A genome project.

5. Bioinformatics

- BAC registry – Dave Edwards would establish a database for BAC sequencing status based on the Medicago schema. This would incorporate the information specific to the R3 and R9 projects already available from the www.brassica-rapa.org site. It was agreed that, once set up at DPI, this would be hosted on www.brassica.info as the Australian commitment came to an end.
- Sharing BAC data – it was noted that 454-generated sequence data posed a particular challenge. It was proposed that Phase I 454 data would be served from an FTP site and Phase 2 data for CE and 454 would be directly deposited with GenBank/EMBL. As noted above, where possible, trace files should also be deposited.
- Annotation – TIGR, in collaboration with key partners in the consortium, would be tuning up an annotation pipeline using Combiner to weight the results of a number of *ab initio* genefinders. This standard annotation would then be sent to GenBank/EMBL.

Issues arising:

It was noted that GenBank/EMBL do not normally accept Phase 2 data as “finished”. The Chair of this committee would need to advise the relevant database curators that the *Br*GSP was resourced only to this stage.

A first-pass annotation of the seed BACs, not to the standard described above, had been made available by JIC at brassica.bbsrc.ac.uk.

6. Next phase of BAC-by-BAC sequencing

Chromosome	Country	PIs	Status	Completion
R1	UK/China.	Bancroft/Meng	Funded	2009
R2	Canada	Parkin/Keller	Submitted 2006; to start 2007	2009
R3	Korea	Park	Funded	2008
R4	Unassigned			
R5	Unassigned			
R6	USA	Town *et al.*	Submitted, decision April 07	2009
R7	Australia	Edwards	Funded	2007
R8	UK/China	Bancroft/Meng	Funded	2009
R9	Korea	Park	Funded	2007
R10	Canada	Parkin/Keller	As R2	2009

Issues arising:

Other contributions: Netherlands was contemplating a possible submission as part of a new programme *(*G. Bonnema) and France was to submit proposal in March 2007 to sequence 250 BACS as part of their National programme.
India - unknown status

For the unassigned chromosomes, we could only keep advertising our efforts to the research community and to strongly support countries developing submissions.

7. AOB

Professor Lim had now completed his term as Chair and the committee thanked him for all his efforts. Although it had previously been agreed that this post should rotate every 2 years, the committee thanked Dave Edwards who agreed to assume the chair for one year, until the 2008 meeting, at which time this issue would be revisited.

8. Next meeting

The committee would meet again at PAG XVI in 2008 but there was also an identified need for an earlier meeting in 2007, especially for those collaborators directly engaged in the sequencing. It was suggested that the 5th Canadian Plant Genomics Workshop, to be held in Vancouver July 30 - Aug 2, might be a suitable occasion for this. Isobel Parkin would send out information regarding the meeting.