

# The *Brassica rapa* Genome Sequencing Project

Ian Bancroft

23<sup>rd</sup> May 2007



# Genome sequencing strategy

An international collaboration:

the “*Brassica rapa* Genome Sequencing Project”

Follows a defined strategy:

- Sequencing Chinese cabbage var. Chiifu
- BAC-by-BAC strategy with allocation by chromosome
- Seed BACs genetically mapped on CKDH and sequenced in Korea
- 110,000 BAC clones end-sequenced
- Walking from seed BACs *via* end-sequences
- BACs finished to Phase 2 standard (single ordered and oriented sequence contig, but some gaps and ambiguous bases permitted)
- All trace files to be deposited to enable “finishing” of specific BACs

# Genome sequencing strategy

An international collaboration:

the "*Brassica rapa* Genome Sequencing Project"

Follows a defined strategy:

- Sequencing Chinese cabbage var. Chiifu
  - BAC-by-BAC strategy with allocation by chromosome
  - Seed BACs genetically mapped on CKDH and sequenced in Korea
  - **110,000 BAC clones end-sequenced**
  - Walking from seed BACs via end-sequences
  - BACs finished to sequence contig, (oriented permitted)
  - All trace files to specific BACs
- Sequencing completed by groups in Korea, UK, Canada, Germany, France, Australia, USA**

# Genome sequencing strategy

An international collaboration:

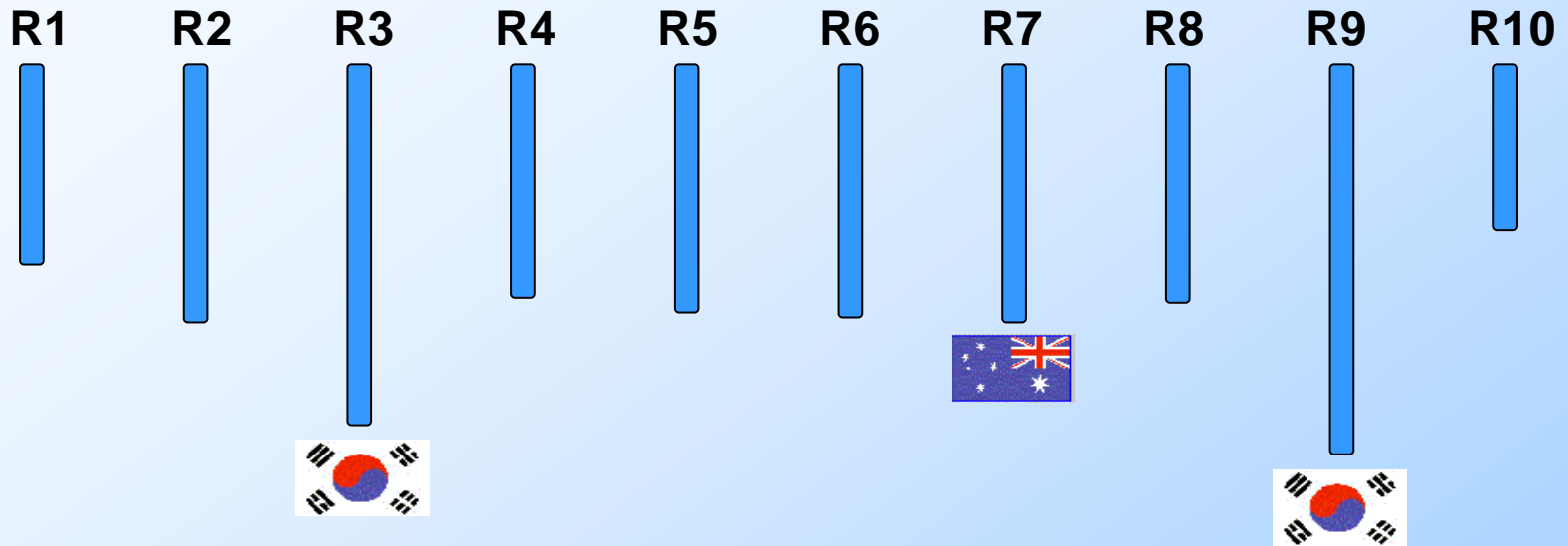
the "*Brassica rapa* Genome Sequencing Project"

Follows a defined strategy:

- Sequencing Chinese cabbage var. Chiifu
- BAC-by-BAC strategy with allocation by chromosome
- **Seed BACs genetically mapped on CKDH and sequenced in Korea**
- 110,000 BAC clones **sequenced**
- Walking from seed
- BACs finished to **oriented** sequence contig, but some gaps and ambiguous bases permitted)
- All trace files to be deposited to enable "finishing" of specific BACs

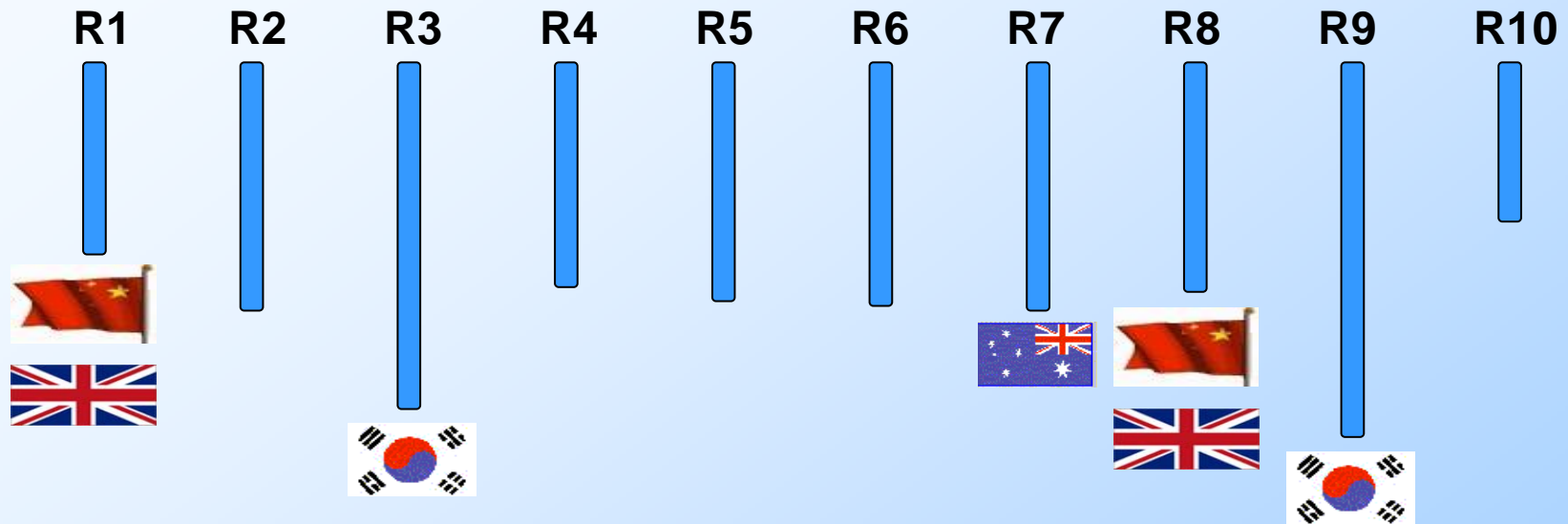
**522 sequenced seed  
BACs in public domain**

# Funded projects



- Gene space of chromosomes R3 and R9 (Korea)
- Gene space of chromosome R7 (Australia)

# The BBSRC-funded project



- 3-year collaborative project between UK (JIC, RRes, W-HRI), China (BGI, HAU) and USA (TIGR)
- Funding awarded January 2007
- Will sequence 700 BACs representing the gene space of R1 and R8
- Will end-sequence a 73,000-clone *B. napus* BAC library (JBnB)

# Availability of outputs

BAC sequence data will be released ASAP

Automated sequence annotation pipeline (based on TIGR tools) to be established in UK and used to annotate all BACs sequenced in the *Brassica rapa* Genome Sequencing Project

A web-accessible version of the annotation pipeline to be made available for annotation of any *Brassica* BAC (prototype at: <http://brassica.bbsrc.ac.uk>)

Sequences to be aligned between the *B. rapa* BACs and the Arabidopsis genome sequence, and between the *B. napus* BAC end sequences and the *B. rapa* and Arabidopsis genome sequences

A database of sequences and annotation will be developed, with multiple display interfaces tuned to community requirements



# Schedule (estimated)

Start of project: June 2007

Initial annotation of BACs in GenBank: Has been available for >6 months (<http://brassica.bbsrc.ac.uk>)

First sequenced BACs from project: September 2007

*B. napus* BAC library end sequenced: December 2007