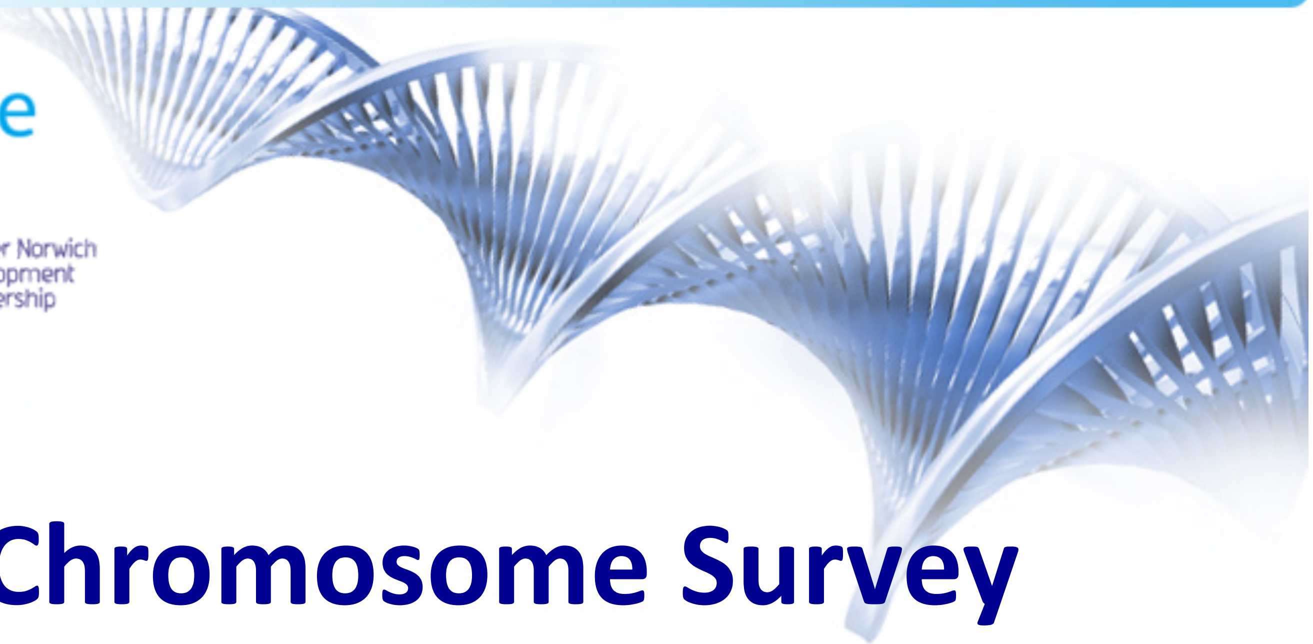
BBSRC Genome Analysis Centre

In partnership with EEDA East of England Development Agency Greater Norwich Development Partnership



Wheat Chromosome Survey Sequencing Project

Jane Rogers

The Genome Analysis Centre

Norwich, UK

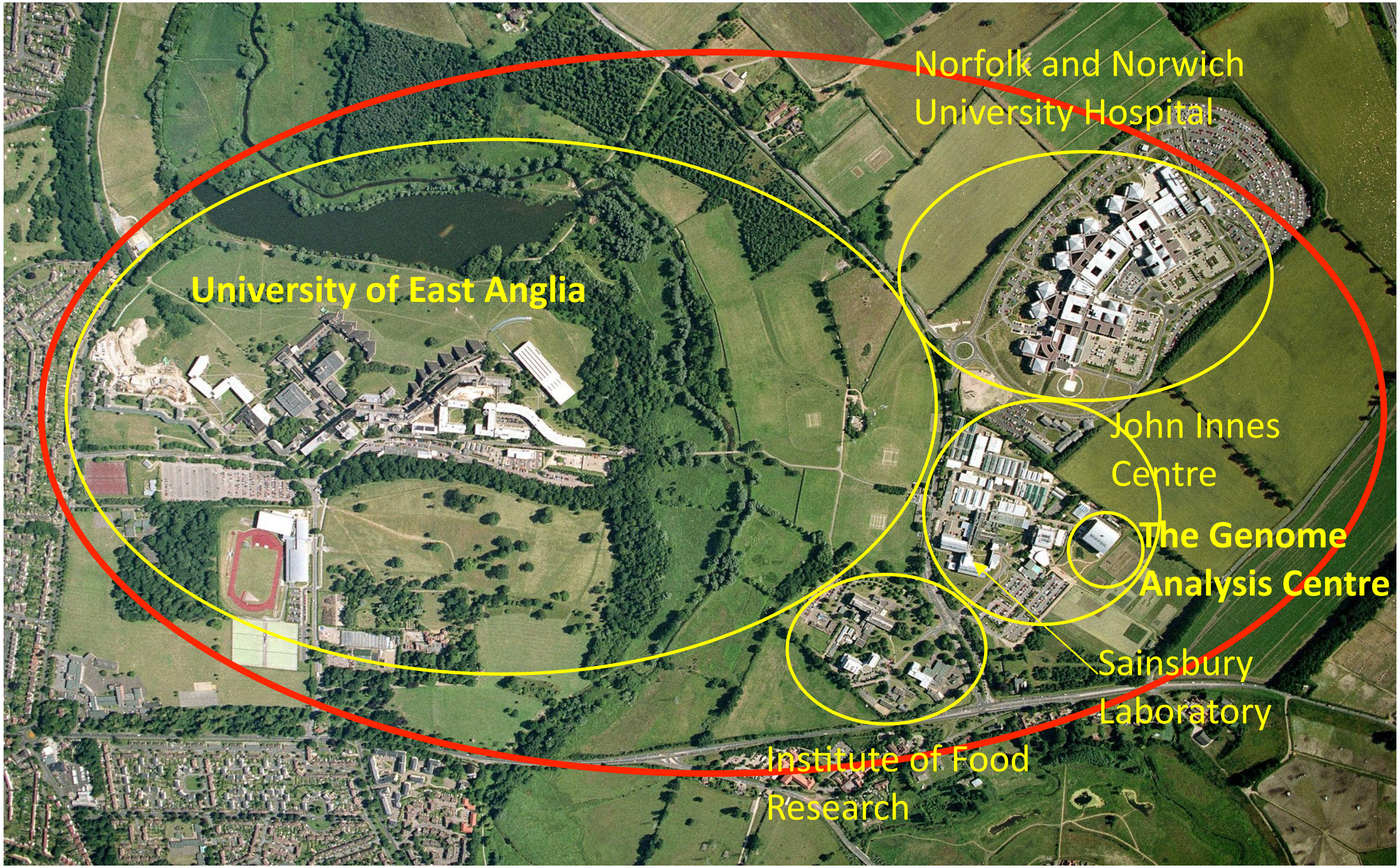
The Genome Analysis Centre

TGAC was launched in 2009 as a national centre of excellence in genomics for plant, animal and microbial research in the UK.

- **High throughput sequencing and data analysis**
- **State of the art technology platforms**
- **Advanced Bioinformatics –** innovative approaches to data interpretation and exploitation.
- **Partnerships with commercial and academic groups in the UK and internationally.**
- **Focussing on applications of non-medical genomics**:
  - Enhancing food and nutrition
  - Improving health
  - Improving production of sustainable energy and in biomanufacture

The Genome Analysis Centre

Norwich Research Park



History of the Survey Sequencing Project

- ITMI Meeting, Beijing, September 2010

  IWGSC agreed to accelerate the production of genome sequence associated with specific chromosome arms in order to :

  (i) assess the gene content of each the 21 chromosomes

  (ii) derive a virtual gene order based on syntenic alignment with other grass genomes e.g. using the *GenomeZipper* (Klaus Meyer)

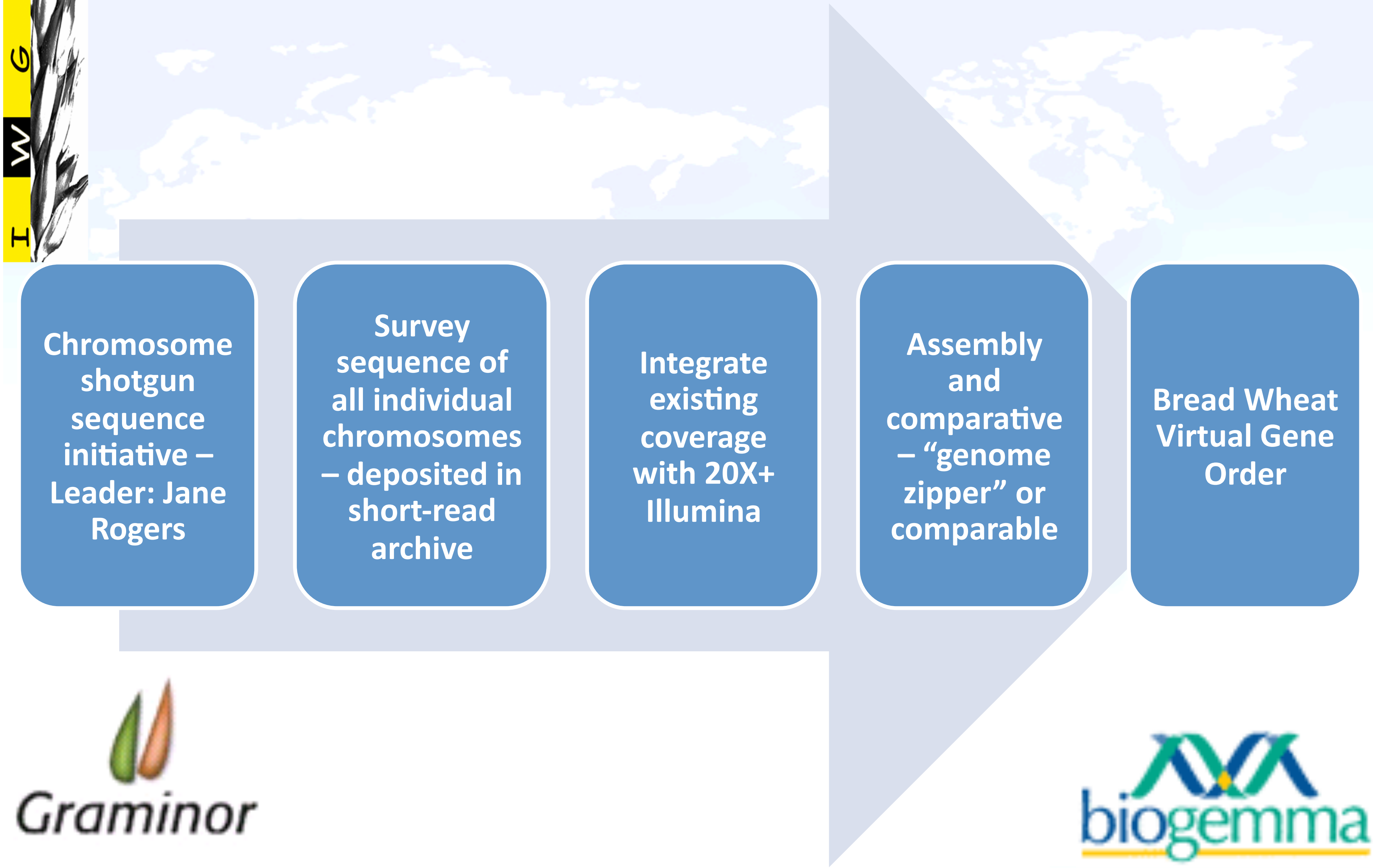
- A plan was developed to complement the ongoing next generation sequencing efforts on 10.5 chromosomes to achieve coverage of all chromosome arms with shotgun sequence to a depth equivalent to at least 20x paired end Illumina sequence reads. Goal: to complete chromosome arm survey sequencing by PAG / end of January 2011

- TGAC was asked by the IWGSC to lead and coordinate the survey sequencing project and to produce sequence of the chromosome arms for which funding for sequencing had not yet been secured or could not be delivered within the agreed timeframe.

- Sequencing was supported by Graminor, Biogemma and BBSRC(UK)

IWGSC

Roadmap: Short-Term Goal – Spring 2011

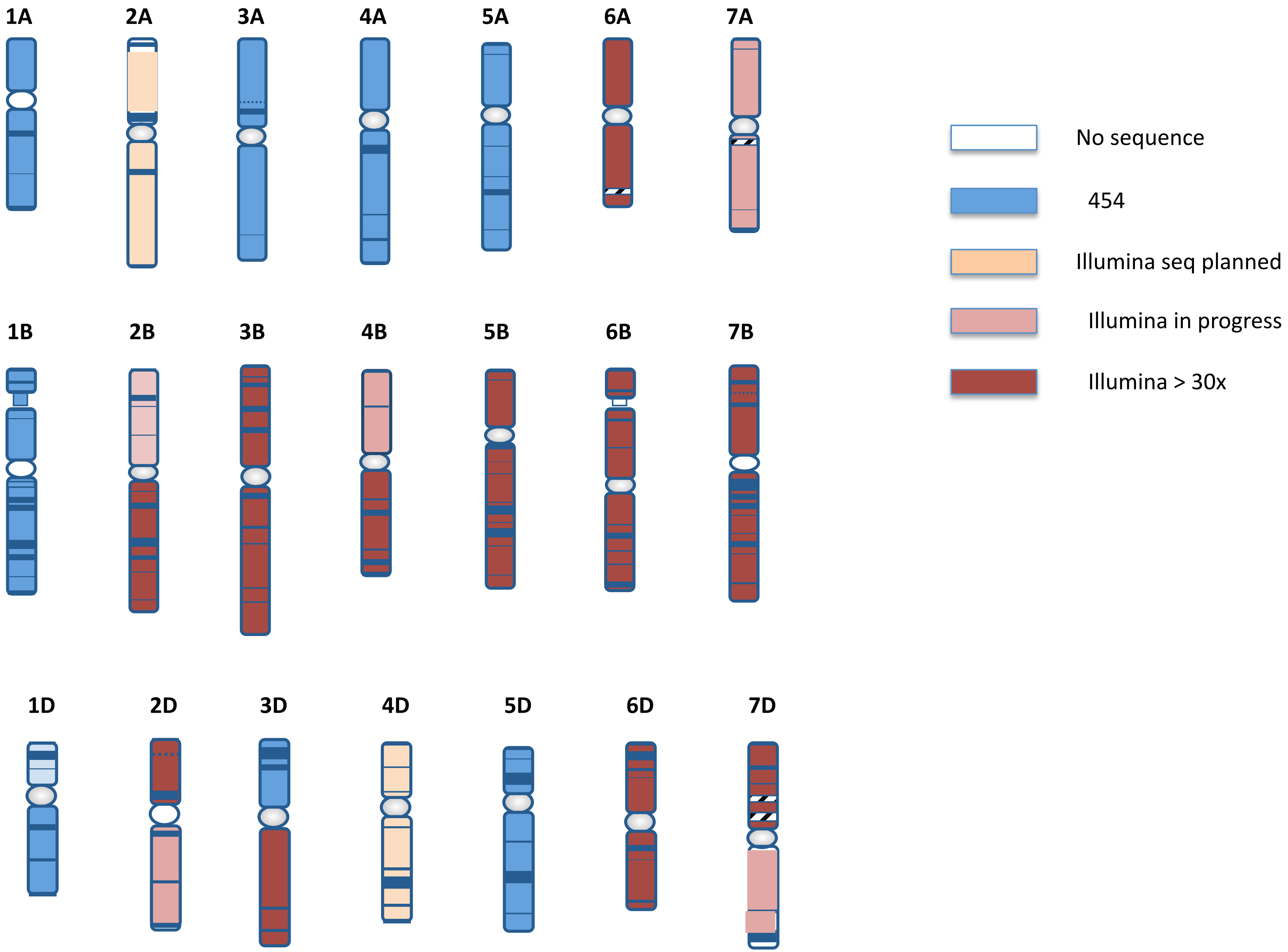


- Chromosome shotgun sequence initiative – Leader: Jane Rogers
- Survey sequence of all individual chromosomes – deposited in short-read archive
- Integrate existing coverage with 20X+ Illumina
- Assembly and comparative – “genome zipper” or comparable
- Bread Wheat Virtual Gene Order

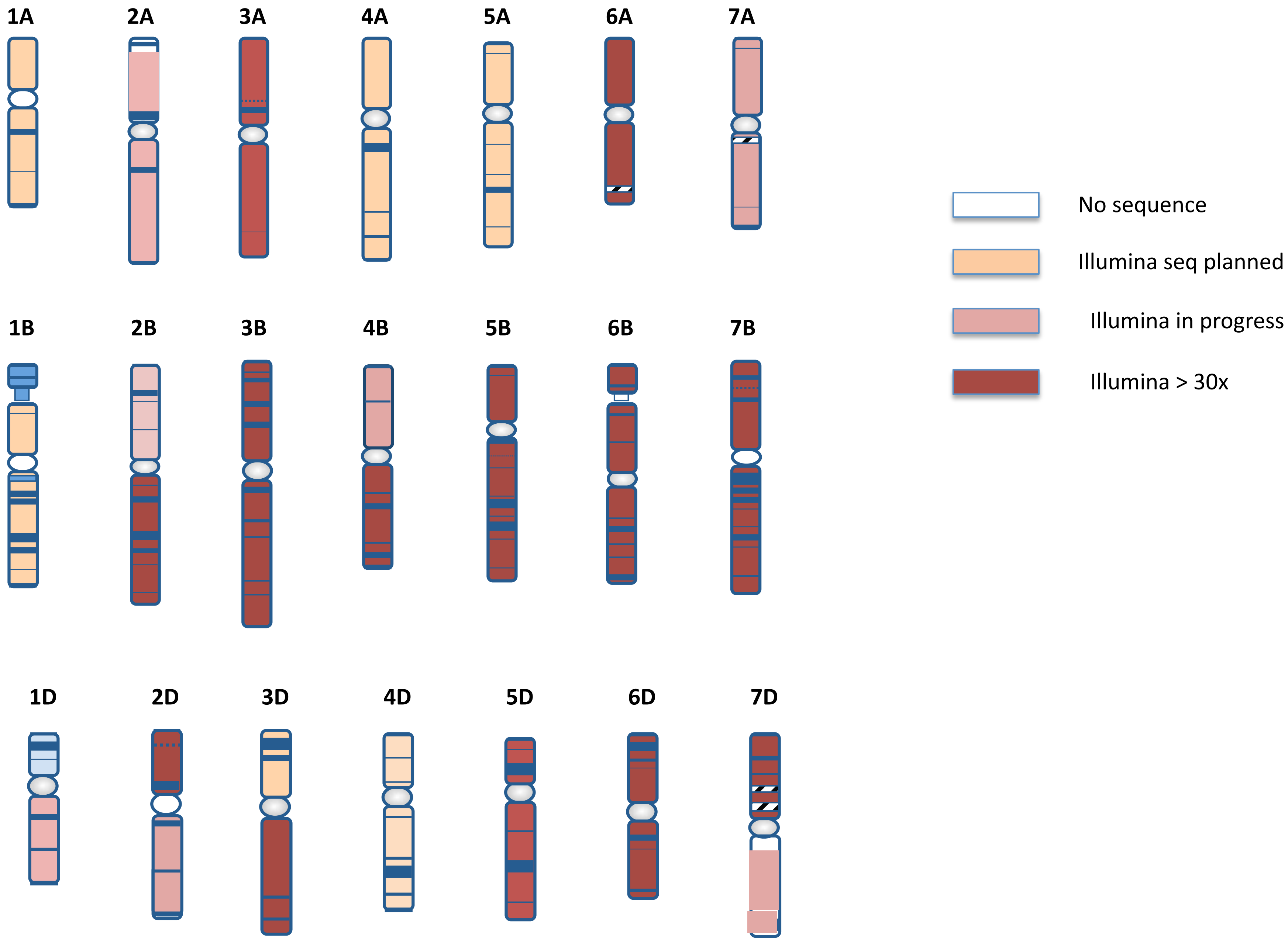
Graminor

biogemma

Wheat Chromosome Survey Progress January 2011



Wheat Chromosome Survey Progress April 2011



Chromosome Survey Sequence Coverage

Chromosome survey sequence depth (> phred 30) April 201

	1AS	**1AL**	**2AS**	**2AL**	**3AS**	**3AL**	**4AS**	**4AL**	**5AS**	**5AL**	**6AS**	**6AL**	**7AS**	**7AL**
454	1.4	1.5			9	9	2.3	1.5	8	2				
Ill’a	★	★	★	★	50	50	★	★		★	40	24		
	1BS	**1BL**	**2BS**	**2BL**	**3BS**	**3BL**	**4BS**	**4BL**	**5BS**	**5BL**	**6BS**	**6BL**	**7BS**	**7BL**
454	2.2	1.5												
Ill’a	★	★	17	25	45	45	56	57	44	40	91	74	76	45
	1DS	**1DL**	**2DS**	**2DL**	**3DS**	**3DL**	**4DS**	**4DL**	**5DS**	**5DL**	**6DS**	**6DL**	**7DS**	**7DL**
454	1.9	1.4			3.4	2	4.9	2.9	2	2				
Ill’a	★	★	50	30	★	150	★	★	48	39	40	40	34	

Assemblies generated for 11 chromosome arms (incl. 3B, 6B, 7B)

Wheat Chromosome Survey Project - DCC

Data Coordination Centre (DCC)

Aims:

Track and version the sequence data processed in the initial phase at TGAC.

Provide information about progress.

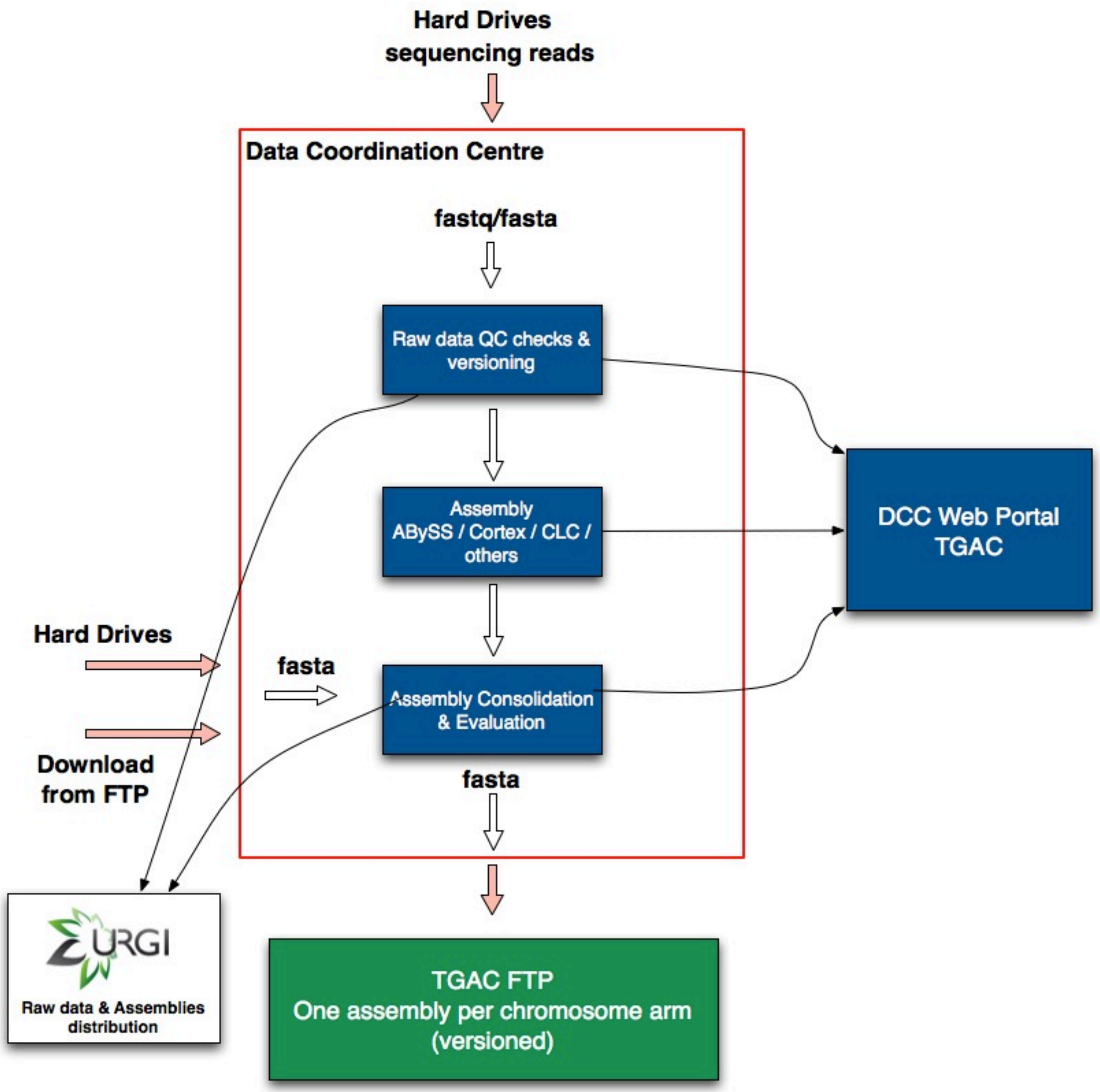
Track and assess the chromosome assemblies that will be fed into the next analysis steps.

Data delivery:

Hard drives

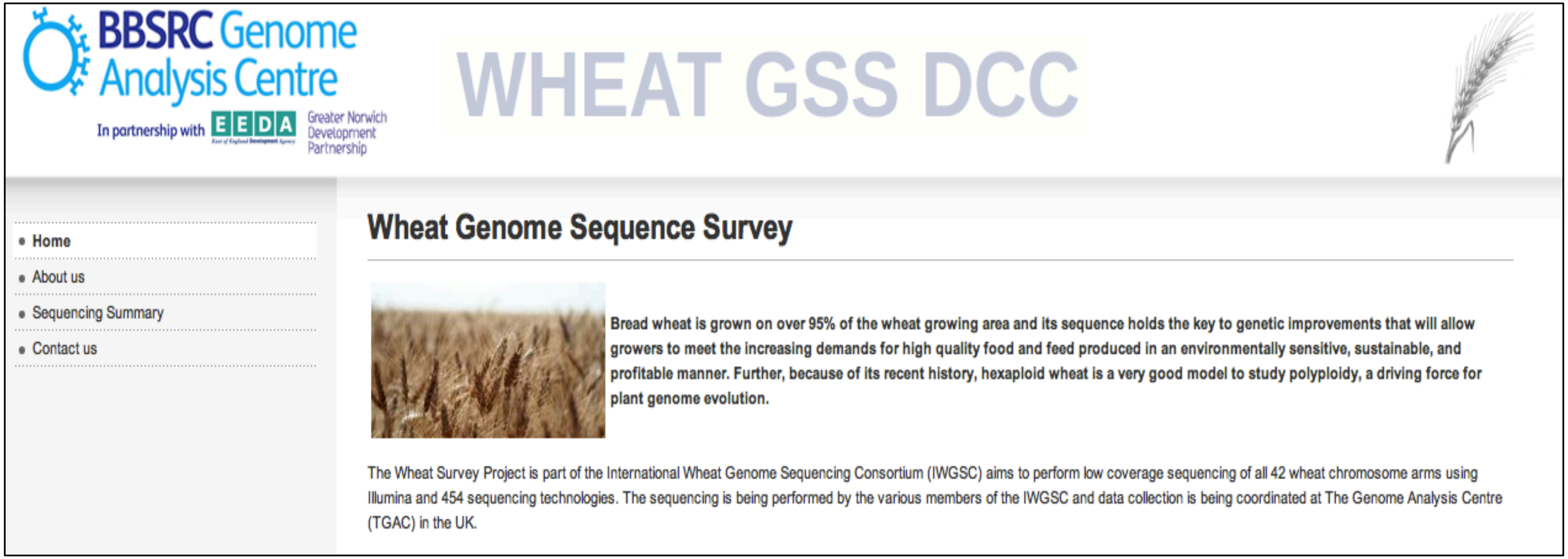
fastq sequence files but…

we are happy to assist with other formats and methods.



DCC portal

wheatdcc.tgac.bbsrc.ac.uk



Wheat Genome Sequence Survey

Bread wheat is grown on over 95% of the wheat growing area and its sequence holds the key to genetic improvements that will allow growers to meet the increasing demands for high quality food and feed produced in an environmentally sensitive, sustainable, and profitable manner. Further, because of its recent history, hexaploid wheat is a very good model to study polyploidy, a driving force for plant genome evolution.

The Wheat Survey Project is part of the International Wheat Genome Sequencing Consortium (IWGSC) aims to perform low coverage sequencing of all 42 wheat chromosome arms using Illumina and 454 sequencing technologies. The sequencing is being performed by the various members of the IWGSC and data collection is being coordinated at The Genome Analysis Centre (TGAC) in the UK.

3DL strategy

Assembly

- Illumina reads (80x76bp CSHL, 70x150bp TGAC) assembled with SGA.
- SGA assembly extended with a single iteration of Curtain.
- 2x 454 3DL reads (TGAC) aligned to ends of Curtain-extended contigs to identify 454 reads which either extended contigs or bridged between contigs.

Analysis

- Curtain-extended assembly aligned to Brachypodium v1.2 exon set.

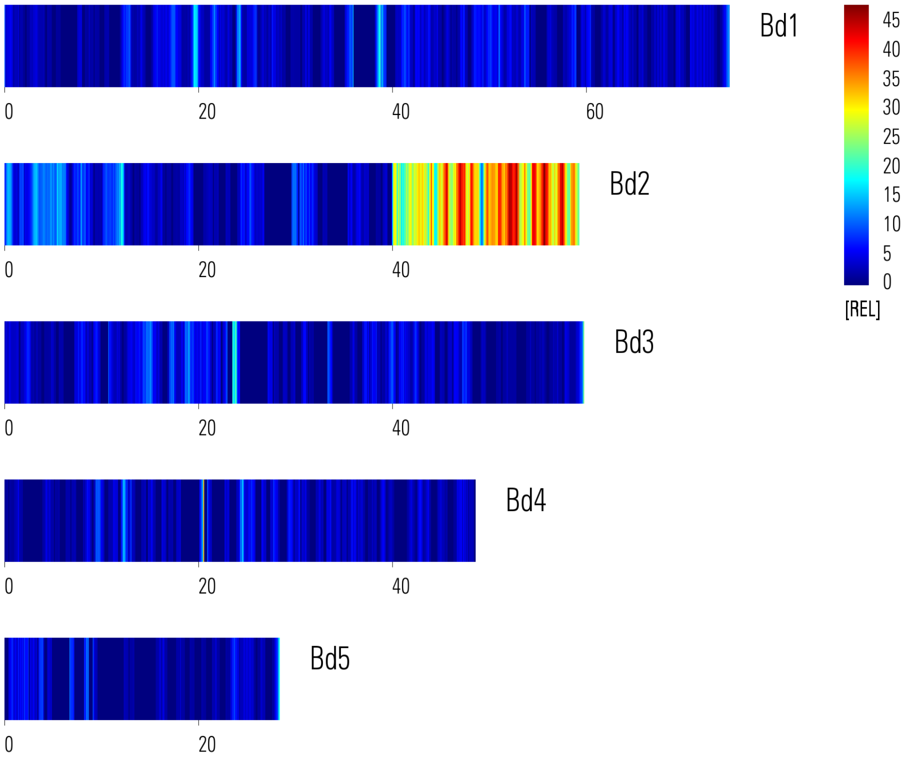
  At 70% translated sequence identity the assembly hit >95% of SYNTENIC genes on Bd 2L (at the exon level, we hit 76% of the syntenic exons), and 70% of ALL genes on Bd 2L (exon hits = ~54%).
- Aligned the Curtain-extended assembly to a set of *A. tauschii* gene sequences from 3DL, provided by Jan Dvorak. Hit 495 of the 541 sequences (91.5%).

D. Waite, M. Bevan, P. Kersey

3DL assembly

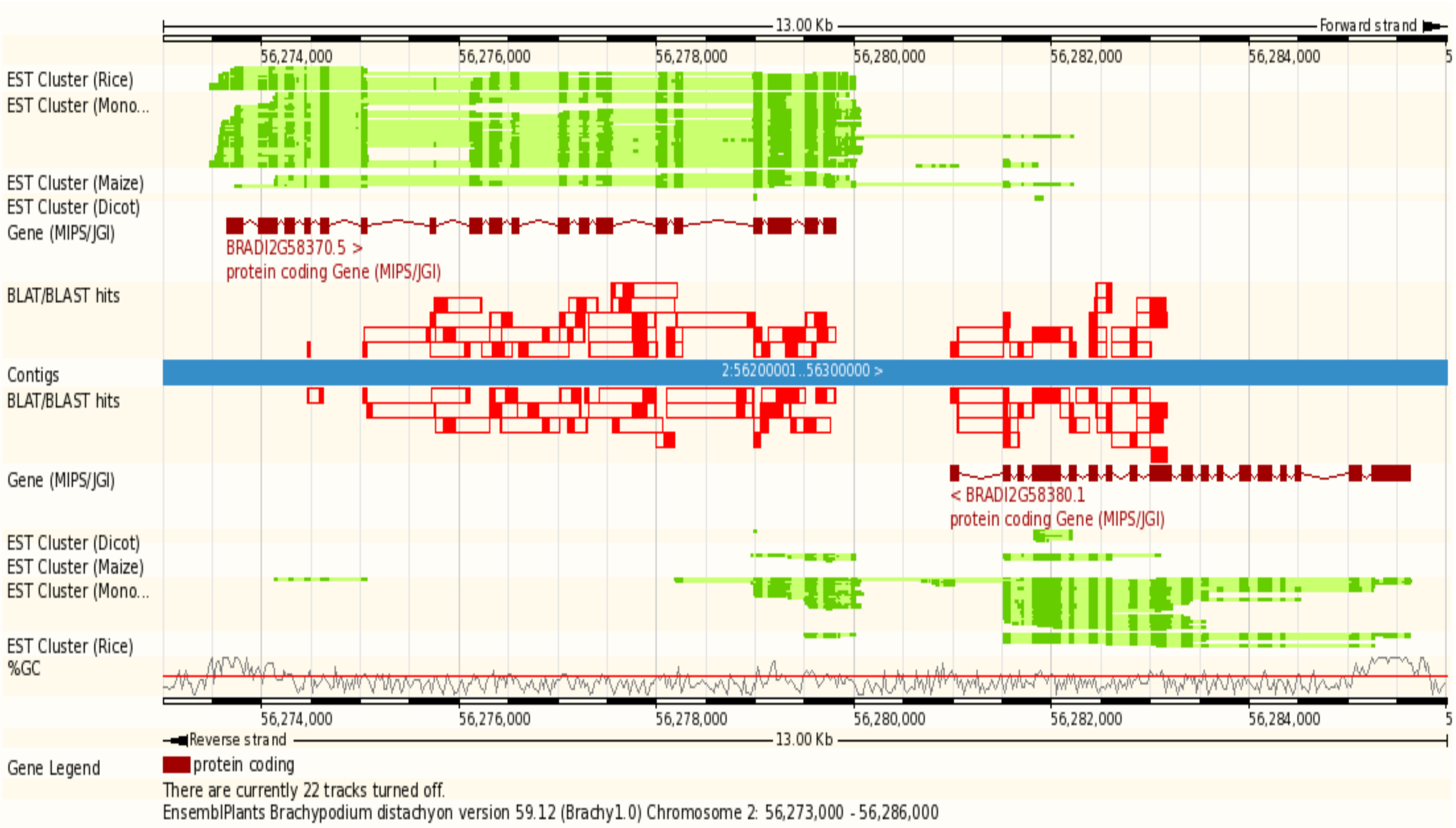
	SGA	SGA+Curtain (> 200bp) …1x iteration
Max contig length	17,949	20,729
N10	3,283	4,037
N50	348	625
N90	212	222
Min	200	200
Bases	220,065,841	265,196,493
contigs	581,136	561,927
Ns	0	17,826,728

Wheat 3DL vs *Brachypodium*



K. Mayer

Ensembl views of alignments



There are currently 22 tracks turned off.

EnsemblPlants Brachypodium distachyon version 59.12 (Brachy1.0) Chromosome 2: 56,273,000 - 56,286,000

Next Steps for the Survey Sequence Project

1. Complete chromosome arm survey sequencing (end May 2011)
2. Data QC + submission to URGI (completed June 2011)
3. Assemble data and make contigs available to K. Mayer to run Genome Zipper analysis
4. Build syntenic views of genes assigned to individual chromosomes
5. Refine
6. IWGSC publication / data release / integration with full chromosome sequencing efforts (Autumn 2011)

Acknowledgements

- **Chromosome DNA**
  Jaroslav Dolozel,

- **Wheat 3DL**
  - **JIC:** Mike Bevan, Darren Waite
  - **EBI:** Paul Kersey, Matthias Haimel
  - **TGAC:** Nizar Drou, Ricardo Ramirez, Jon Wright
  - **MIPS:** Klaus Meyer

- **Wheat Chromosome Survey**
  - **International Wheat Genome Sequencing Consortium**
  - **TGAC:**
    Sequencing: Melanie Febrer, Kirsten McLay, David Baker
    (DCC): Xingdong Biam, **Jon Wright**, **Mario Caccamo**

Funding support for Illumina survey sequence: Graminor, Biogemma, INRA, BBSRC