

ACTTGTGCATAGCATGCAATGCCAT  
ATATAGCAGTCTGCTAAGTCTATAG  
CAGACCCTCAACGTGGATCATCCGT  
AGCTAGCCATGACATTGATCCTGAT  
TTACACCATGTACTATCGAGAGCAG  
TACTACCATGTTACGATCAAAGCCG  
TTACGATAAGCATGAACTTGTGCATA  
GCATGCAATGCCATATATAGCAGTC  
TGCTAAGTCTATAGCAGACCCTCAA  
CGTGGATCATCCGTAGCTAGCCATG  
ACATTGATCCTGATTTACACCATGT  
ACTATCGAGAGCAGTACTACCATGT  
TACGATCAAAGCCGTTACGATAGCA  
TGAACCTTGTGCATAGCATGCAATGC  
CATATATAGCAGTCTGCTAAGTCTA  
TAGCAGACCCTCAACGTGGATCATC  
CGTAGCTAGCCATGACATTGATCCT  
GATTACACCATGTACTATCGAGAG  
CAGTACTACCATGTTACGATCAAAG  
CCGTTACGATAGCATGAACTTGTGC  
ATAGCATGCAATGCCATATATAGCA  
GTCTGCTAACCTCTATAAGCAGACCCT  
CAAGCTGGATCATCCCTAGCTAGCC  
ATGACATTTGATCCTGATTTACACCA

# The international effort to sequence the 17Gb wheat genome:

## Yes, Wheat can!



**Catherine Feuillet**

**INRA Clermont-Ferrand, France**

**Genetics, Diversity & Ecophysiology  
of Cereals**



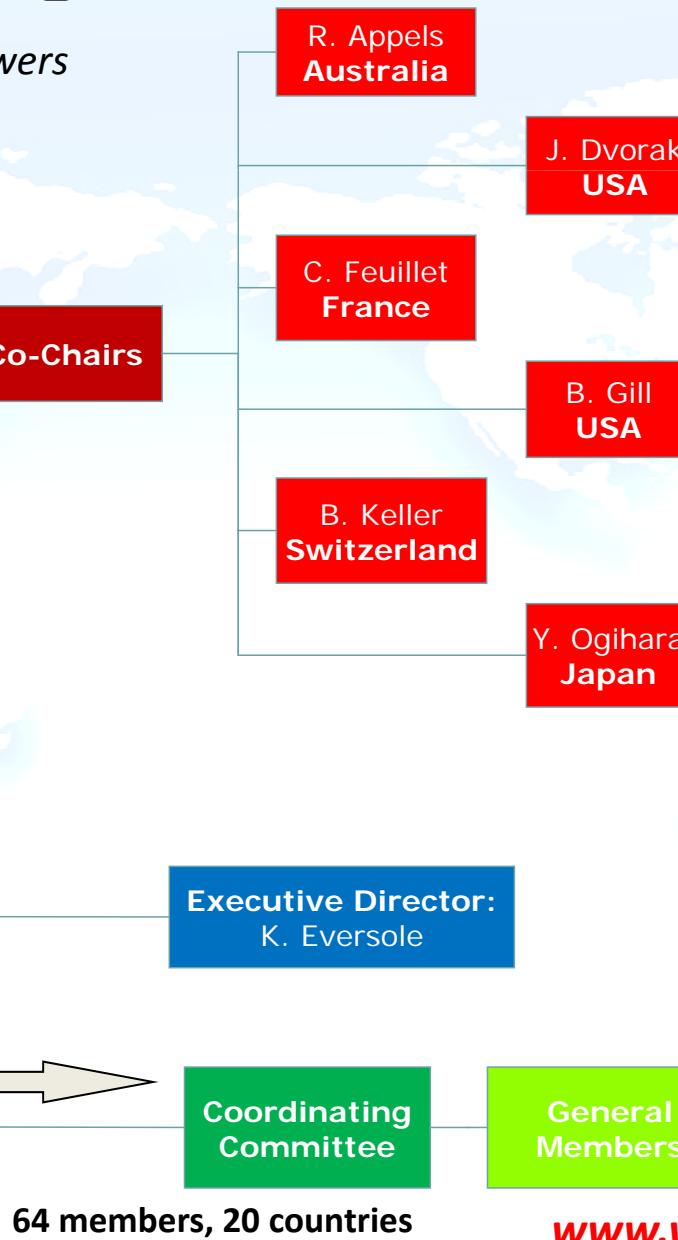
# The International Wheat Genome Sequencing Consortium

Launched in 2005 on the initiative of Kansas Growers

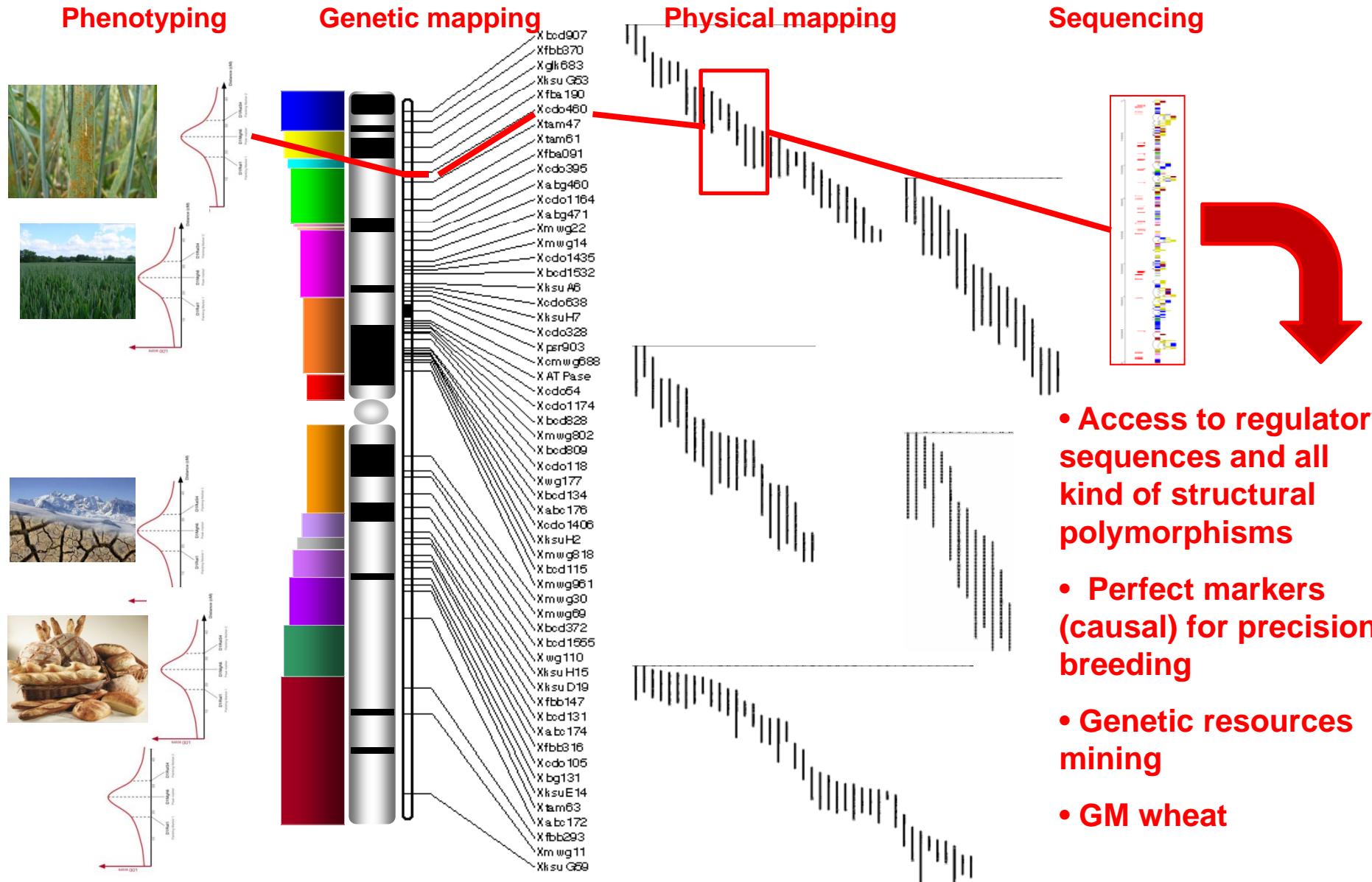


**21 Sponsors**

Funding or  
Scientific  
Contributors

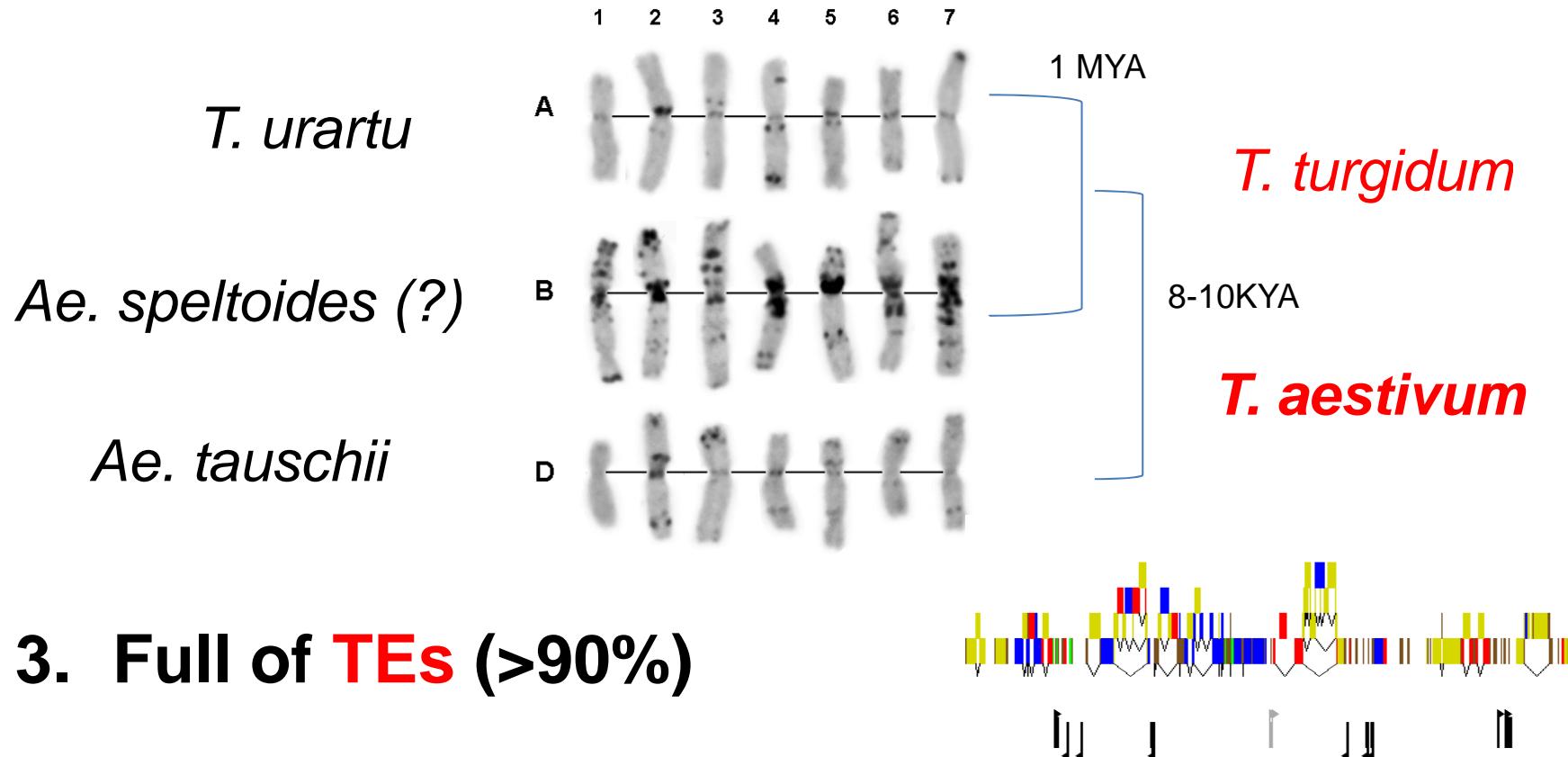


# An integrated and ordered wheat genome sequence

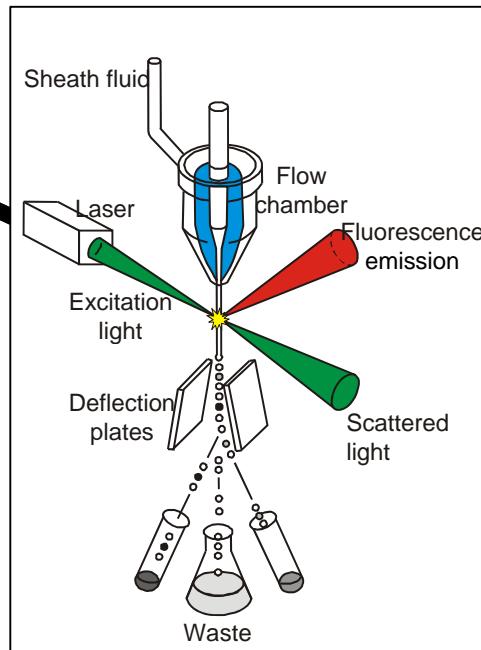
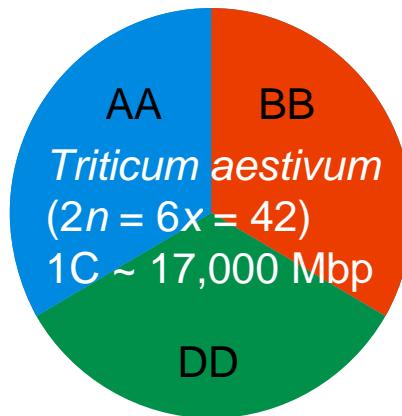


# The Breadwheat genome is.....

1. Big: **17Gb** (5 x human genome, 40 x rice...)
2. Polyploid:  $2n= 42 = 6x$



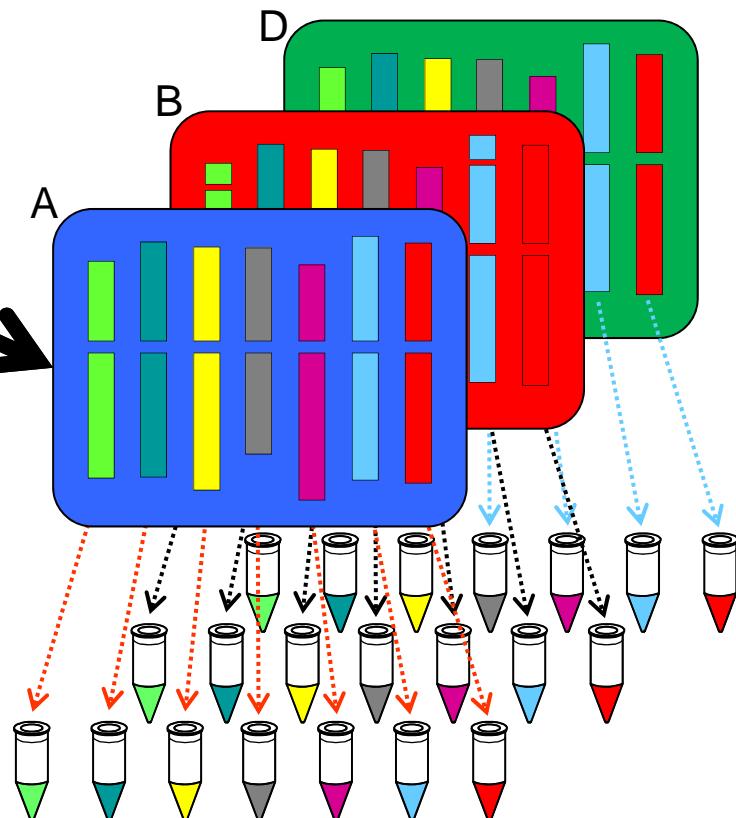
# A chromosome-based approach



Doležel et al., Chromosome Res. 15: 51, 2007

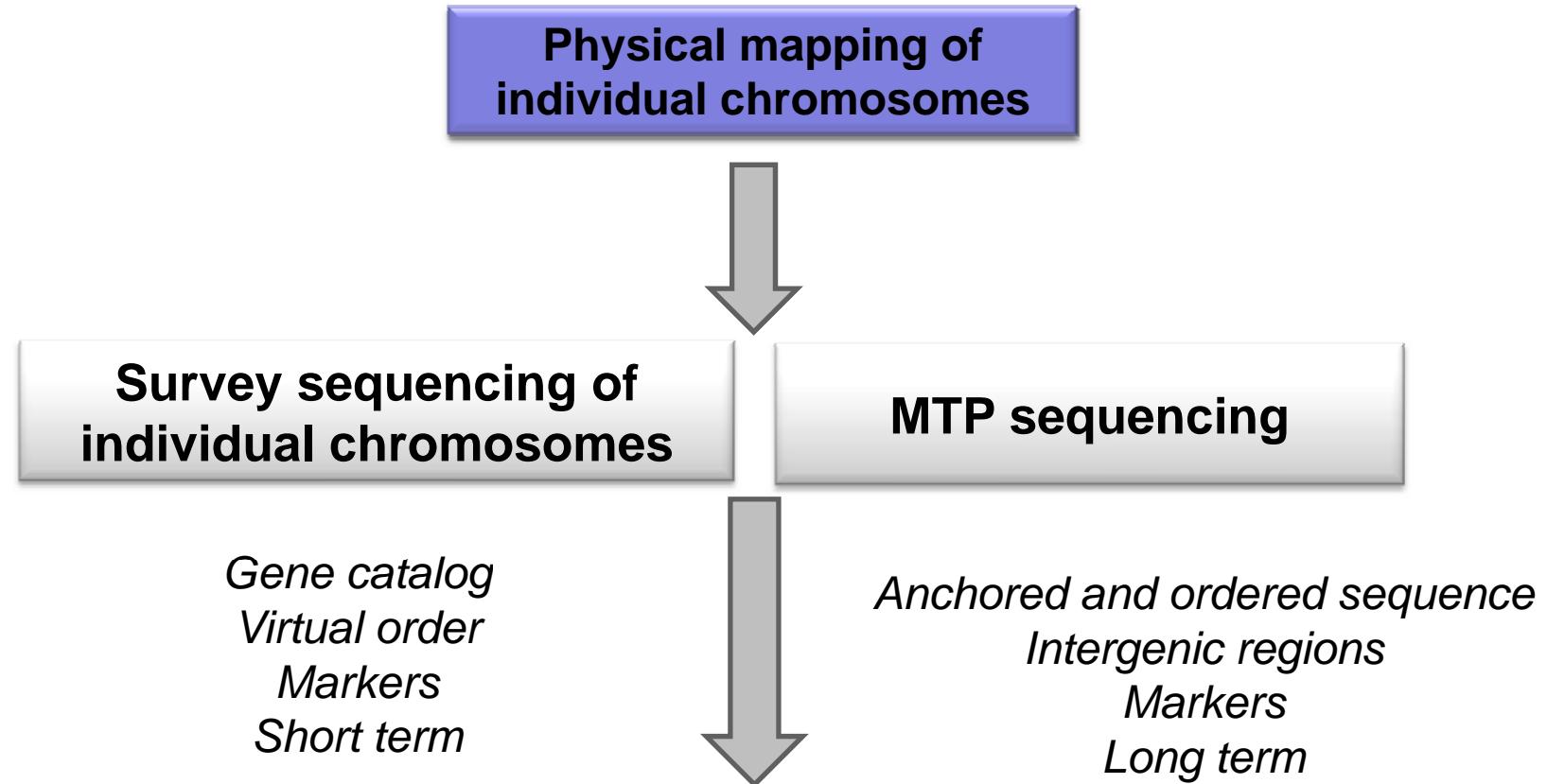
- Chromosomes: 605 - 995 Mbp (3.6 – 5.9% of the genome)
- Chromosome arms: 225 - 585 Mbp (1.3 – 3.4% of the genome)

Dissection of the genome to single chromosomes (arms) representing individual (sub)genomes



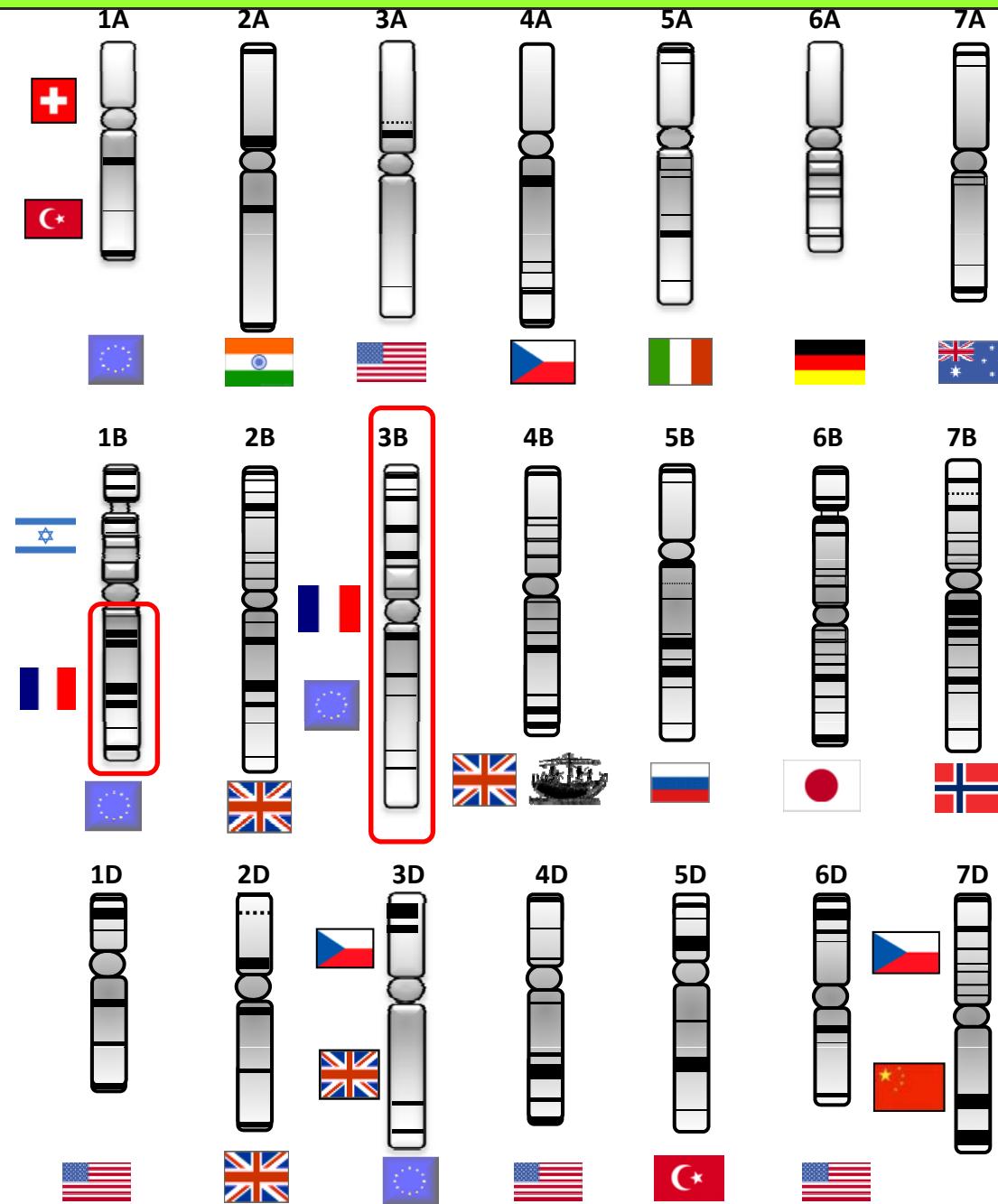
- Chromosome specific BAC libraries (June 2012)
- Amplified DNA for chromosome survey (Nov 2011)

# Combined strategies to establish a wheat reference genome sequence



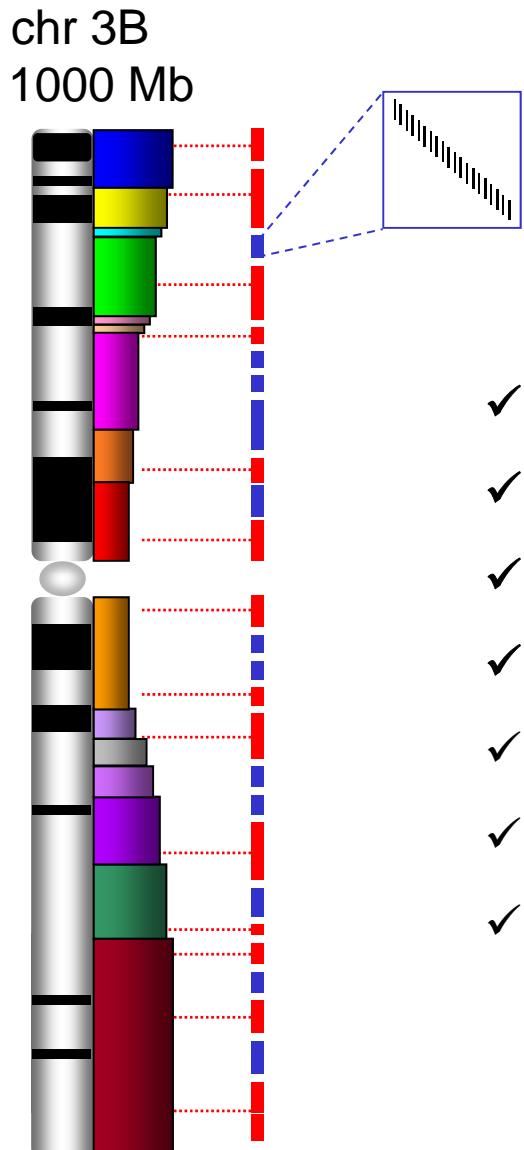
**A reference sequence anchored to the genetic and phenotypic maps**

# An international effort



*T. aestivum*  
cv Chinese Spring

# Physical map of the 1GB chromosome 3B



## A Physical Map of the 1-Gigabase Bread Wheat Chromosome 3B

Etienne Paux,<sup>1</sup> Pierre Sourdille,<sup>1</sup> Jérôme Salse,<sup>1</sup> Cyrille Saintenac,<sup>1</sup> Frédéric Choulet,<sup>1</sup> Philippe Leroy,<sup>1</sup> Abraham Korol,<sup>2</sup> Monika Michalak,<sup>3</sup> Shahryar Kianian,<sup>3</sup> Wolfgang Spielmeyer,<sup>4</sup> Evans Lagudah,<sup>4</sup> Daryl Somers,<sup>5</sup> Andrzej Kilian,<sup>6</sup> Michael Alaux,<sup>7</sup> Sonia Vautrin,<sup>8</sup> Hélène Bergès,<sup>8</sup> Kelley Eversole,<sup>9</sup> Rudi Appels,<sup>10</sup> Jan Safar,<sup>11</sup> Hana Simkova,<sup>11</sup> Jaroslav Dolezel,<sup>11</sup> Michel Bernard,<sup>1</sup> Catherine Feuillet<sup>1</sup>

SCIENCE VOL 322 3 OCTOBER 2008



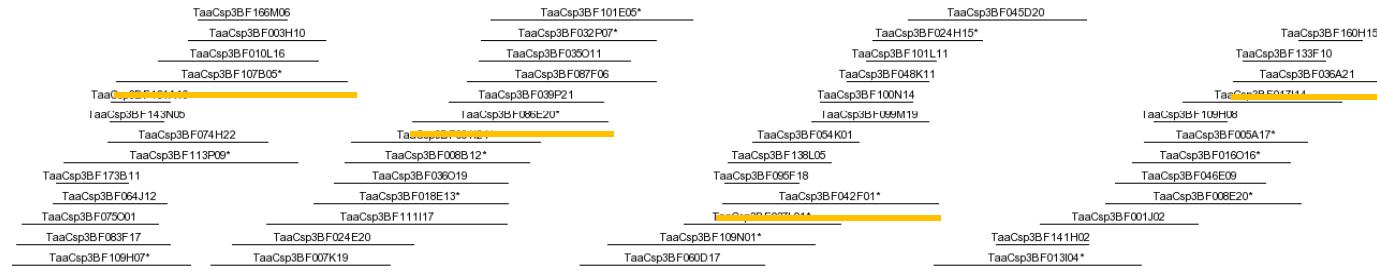
Paux et al, *Science* 2008.; Rustenholz et al, *Plant Physiol* 2011

- ✓ 131 792 fingerprinted BACs by **SNaPshot**
- ✓ 1 283 contigs (average size = 749 kb) with **FPC**
- ✓ **961 Mb coverage (97% chromosome)**
- ✓ 4367 molecular markers (SSRs, ISBPs, unigenes...)
- ✓ 919 anchored contigs (740 Mb) in 16 deletion bins
- ✓ **19.2 X coverage**
- ✓ **MTP (8448 clones)**



# Improving physical maps

- ✓ Pilot sequencing showed that:



→ 10% of BACs in contigs are mis-assembled, some chimerical contigs (0.6/10Mb) and an average contig size smaller than in small genomes

## Whole Genome Profiling

Sequence-based physical mapping of complex genomes by whole genome profiling

Jan van Oeveren,<sup>1</sup> Marjo de Ruiter,<sup>1</sup> Taco Jesse,<sup>1</sup> Hein van der Poel,<sup>1</sup> Jifeng Tang,<sup>1</sup> Feyruz Yalcin,<sup>1</sup> Antoine Janssen,<sup>1</sup> Hanne Volpin,<sup>1</sup> Keith E. Stormo,<sup>2</sup> Robert Bogden,<sup>2</sup> Michiel J.T. van Eijk,<sup>1</sup> and Marcel Prins<sup>1,3</sup>

<sup>1</sup>Keygene N.V., Wageningen, The Netherlands; <sup>2</sup>Amplicon Express Inc., Pullman, Washington 99163, USA



## Linear Topography Contig

Frenkel et al. BMC Bioinformatics 2010, 11:584  
http://www.biomedcentral.com/1471-2105/11/584

### METHODOLOGY ARTICLE



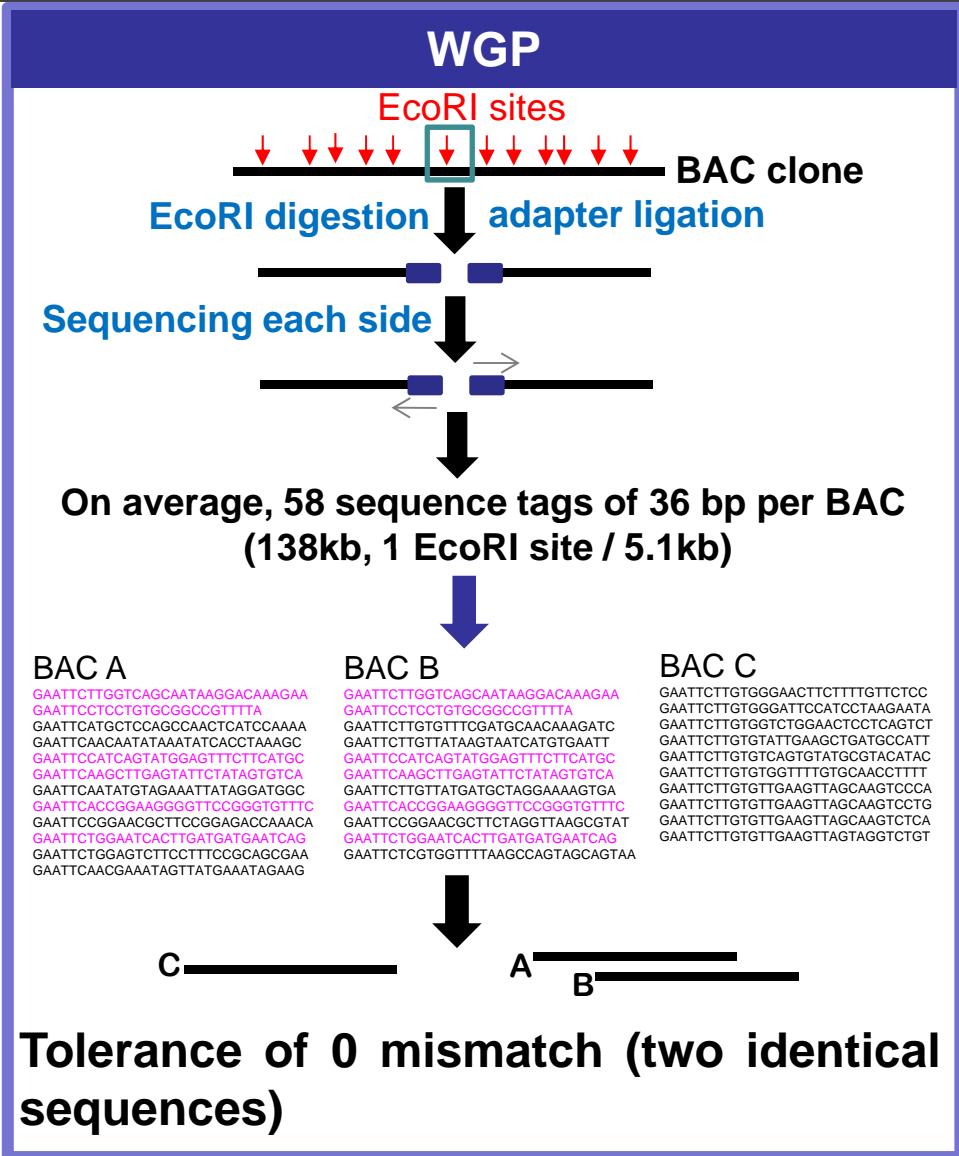
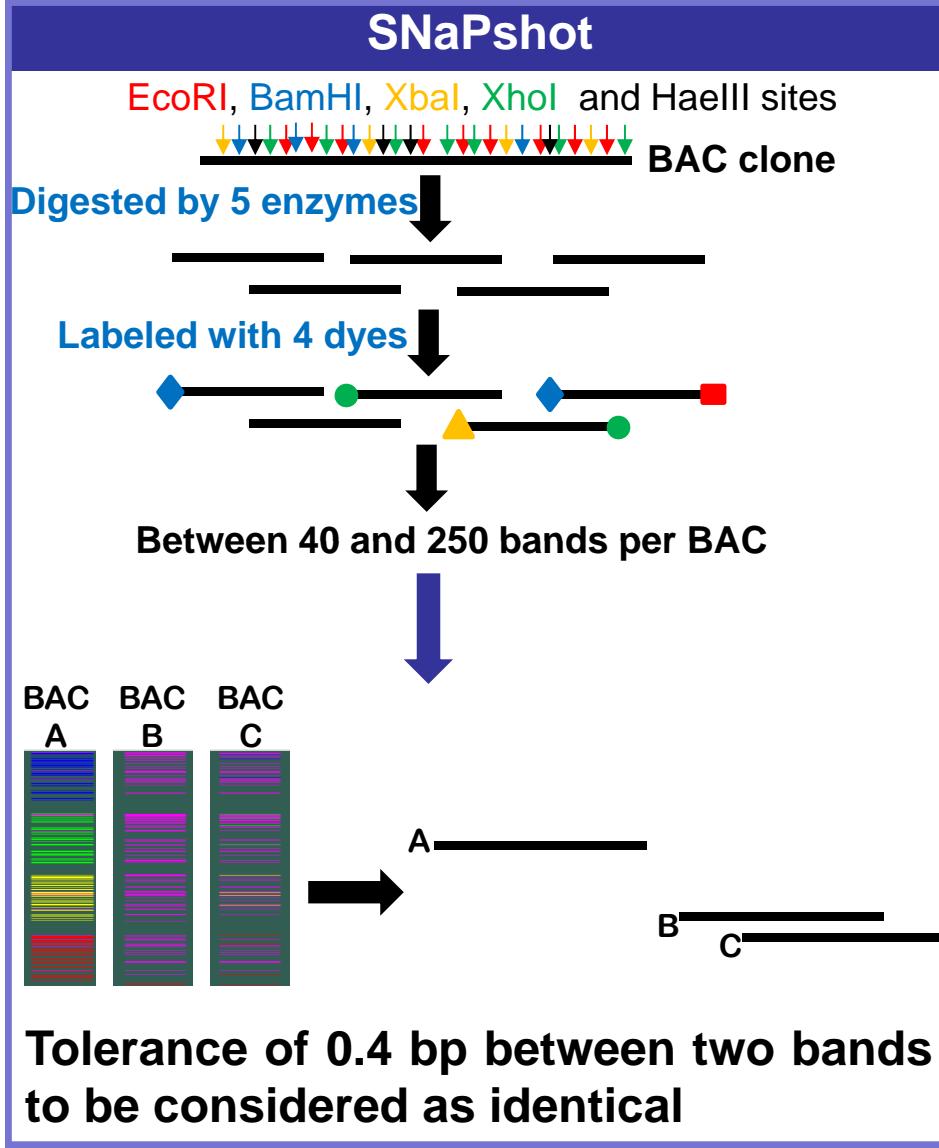
Open Access

LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes

Zeev Frenkel<sup>1\*</sup>, Etienne Paux<sup>2</sup>, David Mester<sup>1</sup>, Catherine Feuillet<sup>2</sup>, Abraham Korol<sup>1</sup>



# Whole Genome Profiling (WGP) : a new sequence-based physical mapping technology



➤ Pilot project on 16,128 FP (9.6X) of 3B

# WGP improves physical mapping in wheat

- ✓ Physical map assembly with FPC

	<b>SNaPshot (<math>1^{e-25}</math>)</b>	<b>WGP (<math>1^{e-11}</math>)</b>
Estimated coverage in length	236 Mb ± 65	199 Mb ± 42
Number of contigs	631	434
Average contig size (Kb)	374	<b>469</b>
N50 (Kb)	455	567

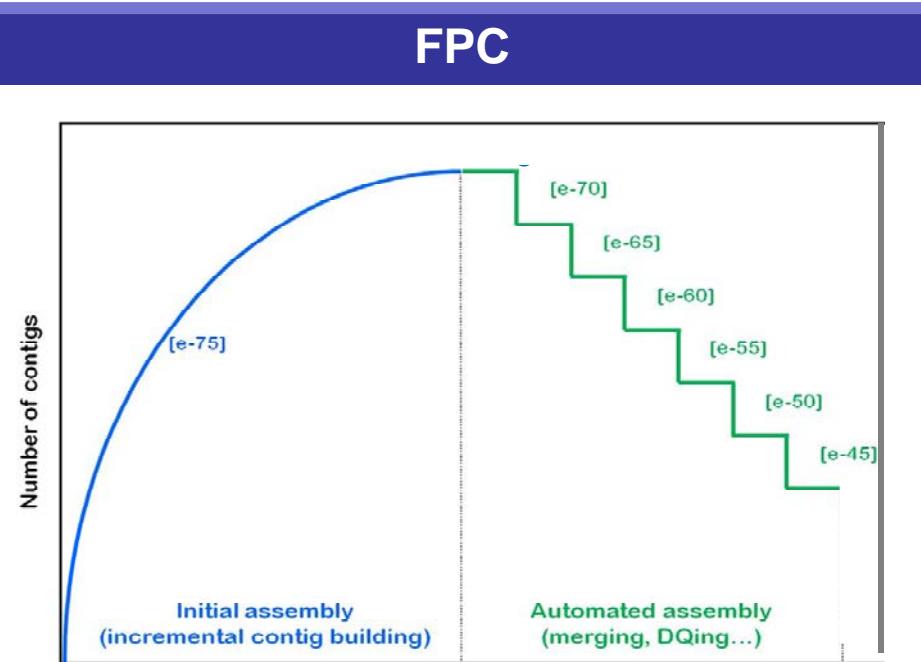
## Comparison to 12 sequenced contigs:

Coverage percentage in length	95.8%	94.9%
Number of chimerical contigs for 10 Mb	0.6	0.6
Percentage of mis-assembled BACs	9.5%	<b>2.7%</b>



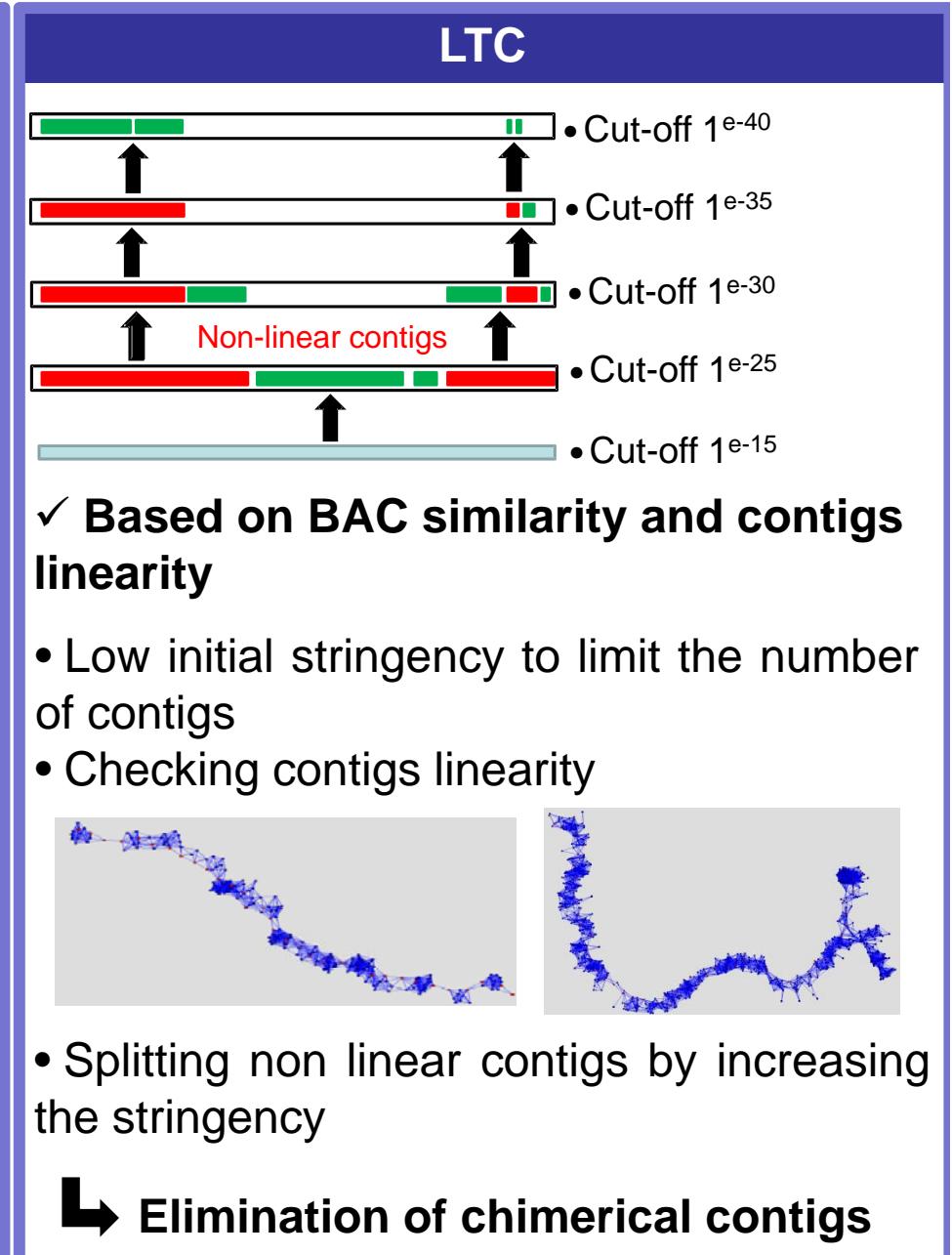
- ✓ Equivalent coverage in length
- ✓ Less mis-assembled BACs in WGP
- ✓ Equivalent number of chimerical contigs

# FPC vs LTC



✓ Based only on BACs similarity

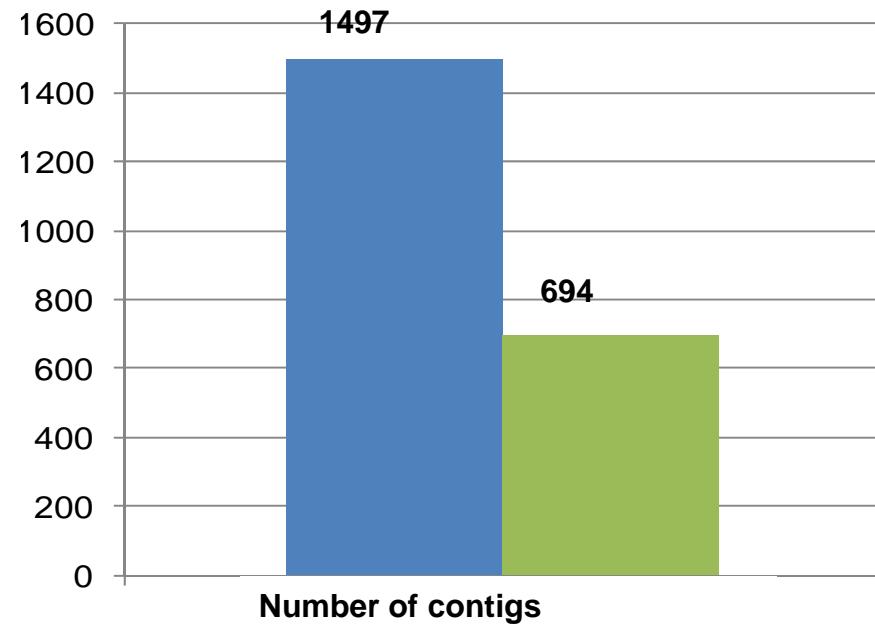
- High initial stringency to limit chimerical contigs
- Adding singletons to contigs extremity at each step
- Merging contigs at each step



# LTC vs FPC (1BL physical map)

- ✓ 1BL estimated size = 535 Mb
- ✓ 65,413 useful fingerprints (SNaPshot)

FPC  
LTC

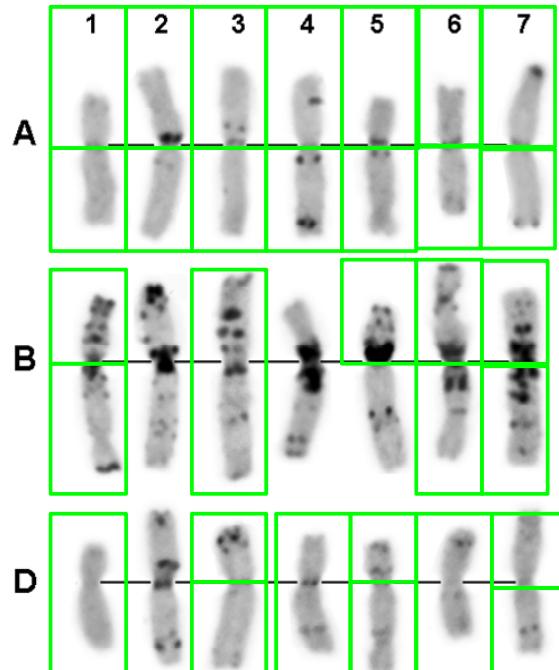


IWGSC workshop  
Tuesday 17th 1-6pm



- ✓ LTC significantly improves physical mapping in wheat

# IWGSC physical maps: January 2012



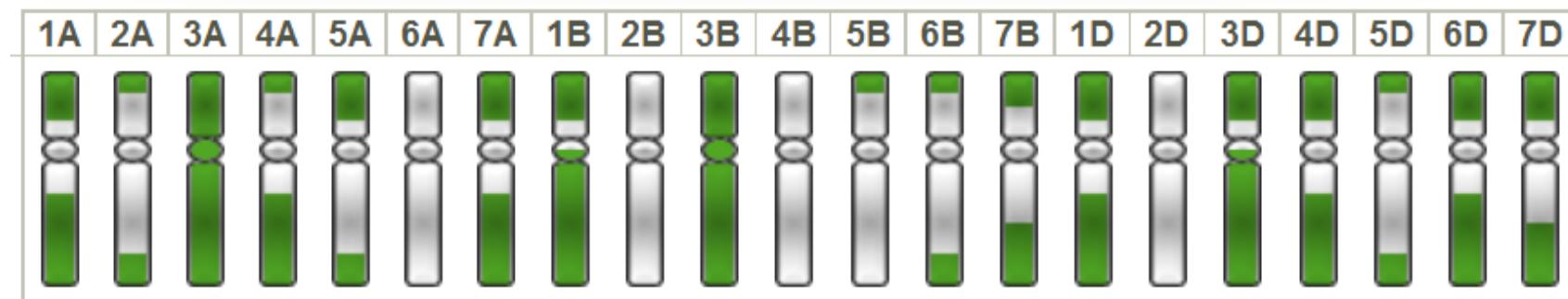
IEB



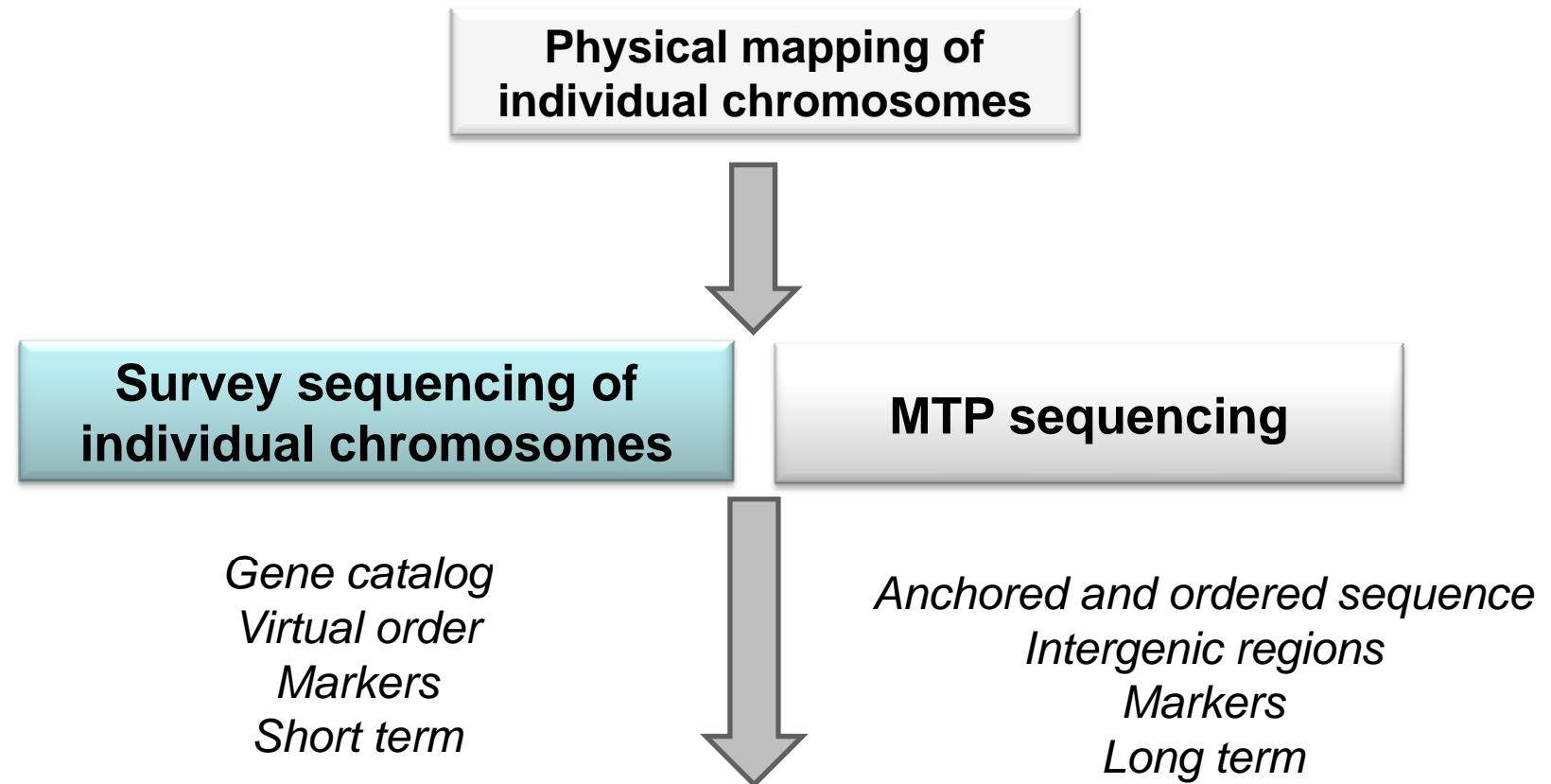
TOTAL : 2,253,312 BAC clones

~ 16 x , 118 kb

Completion by Summer 2012

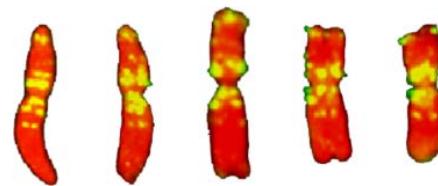


# Combined strategies to establish a wheat reference genome sequence



**A reference sequence anchored to the genetic and phenotypic maps**

# Chromosome Survey Sequencing



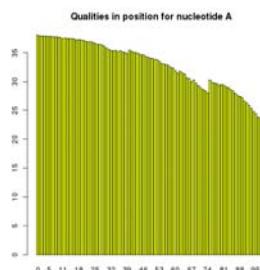
Amplified DNA/sorted chromosomes



IEB



Illumina reads (2\*108 bp/) PE 0.5 kb  
Min 50 x

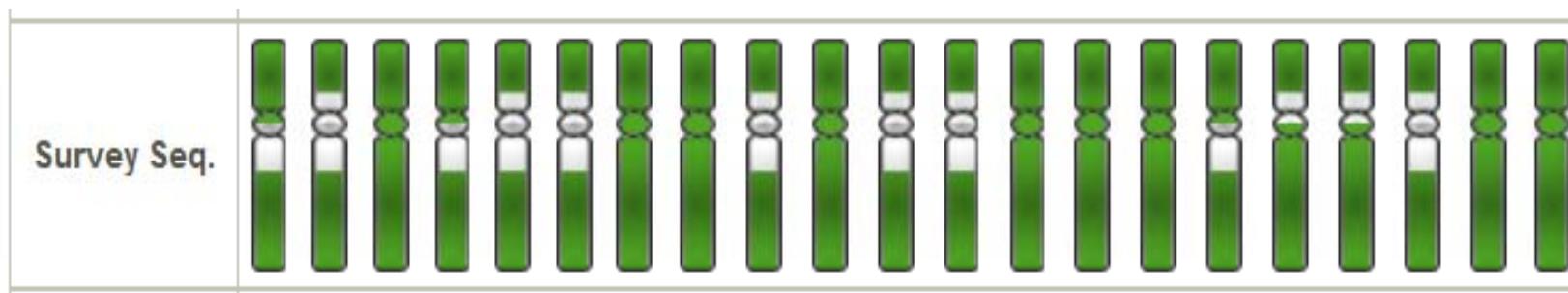


Assembly (ABySS)

K-mer 71

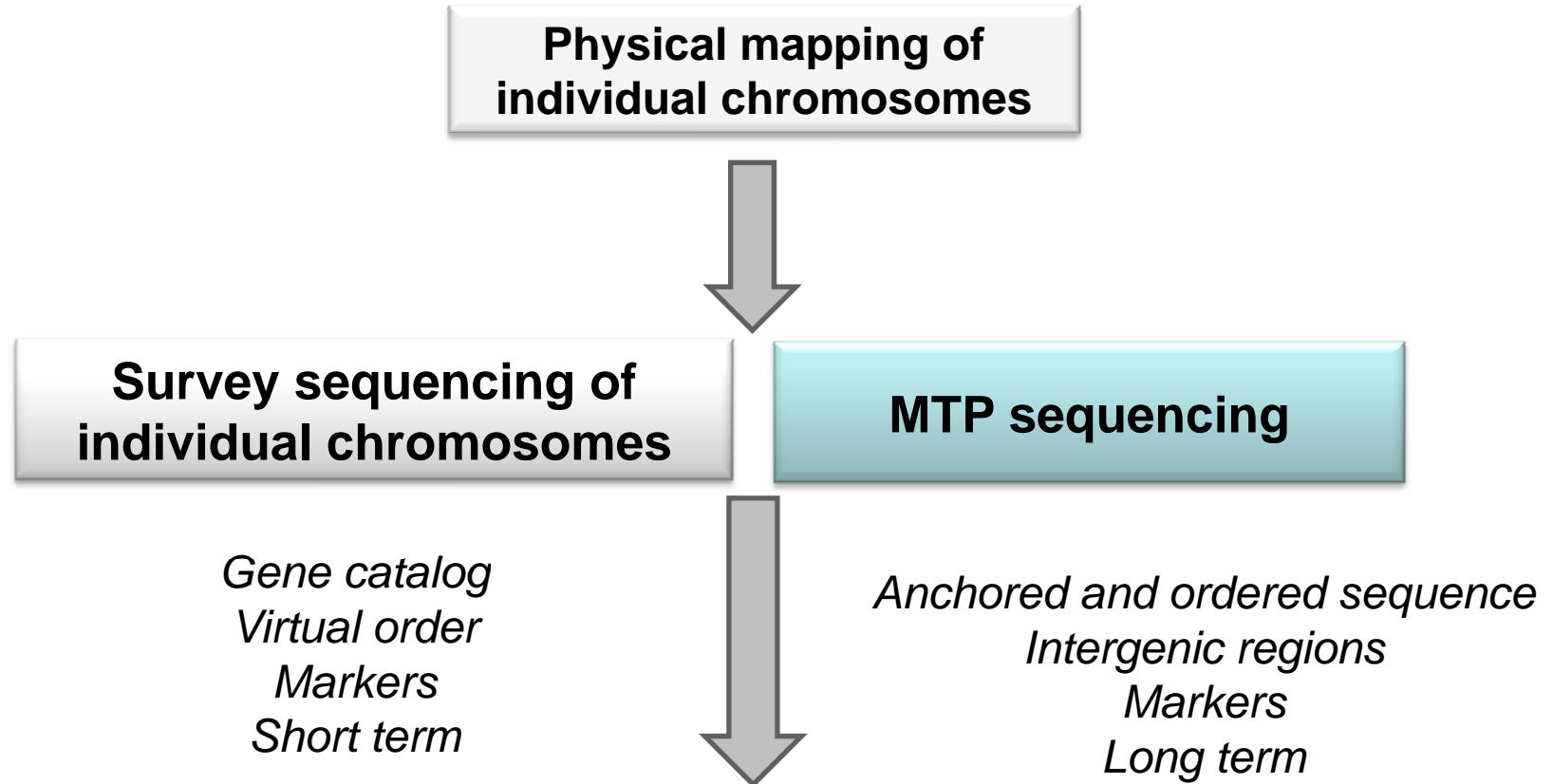
Contigs > 200bp N50 = 2.1 kb

TGAC   
The Genome Analysis Centre™



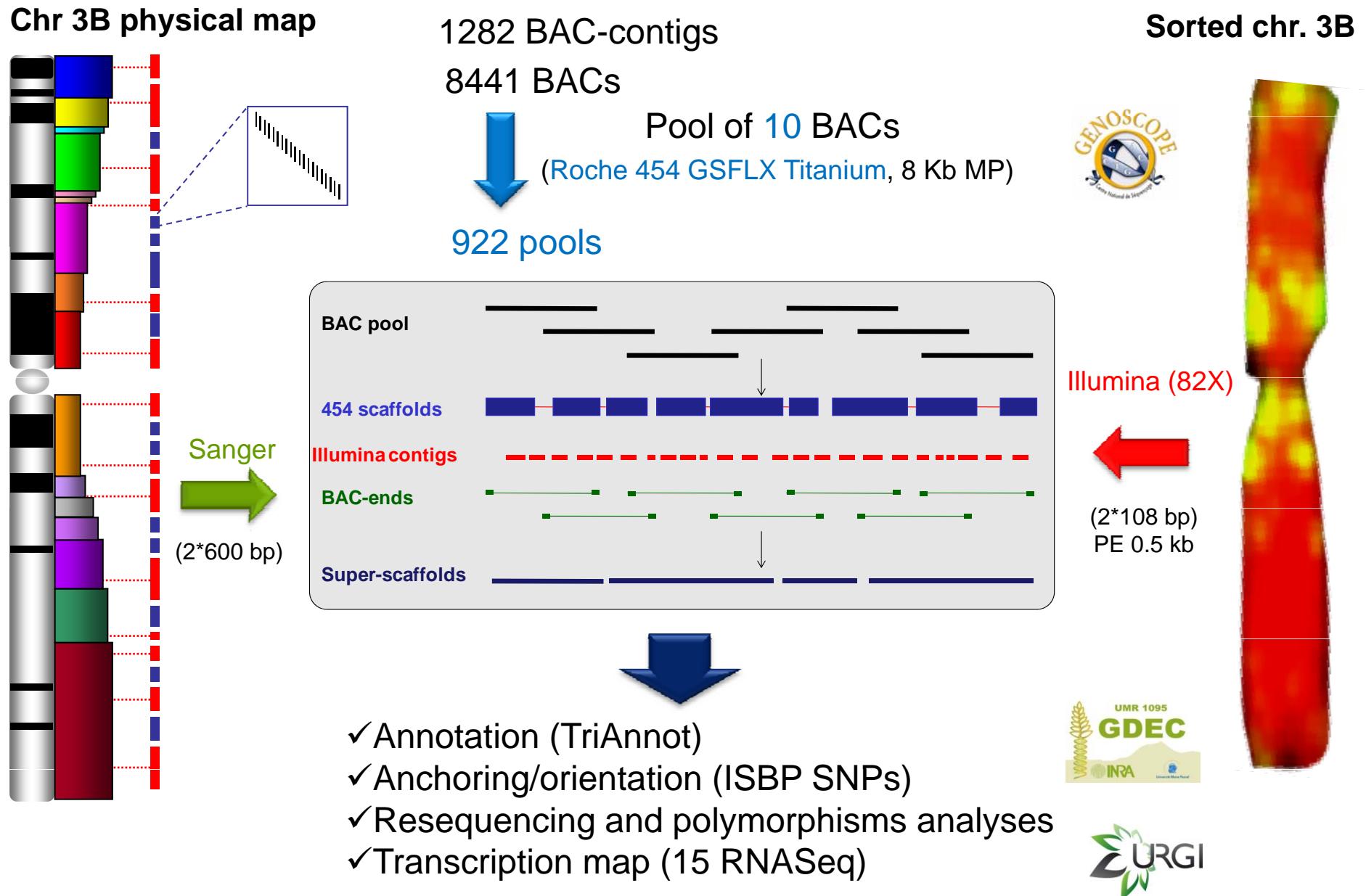
<http://urgi.versailles.inra.fr/Species/Wheat/Sequence-Repository>

# Combined strategies to establish a wheat reference genome sequence

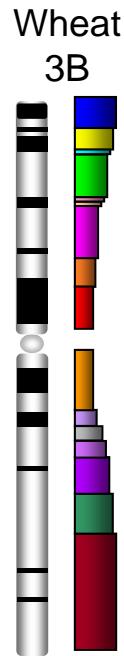


A reference sequence anchored to the genetic and phenotypic maps

# 3B SEQuencing Project (1Gb)



# 3B sequence automated annotation



**8 734 genes**

16 136 scaffolds  
16 136 fasta files

IWGSC workshop  
Tuesday 17th 1-6pm

15 544 predicted genes

Cluster 712 cores  
< 1 day

TAAC.pm

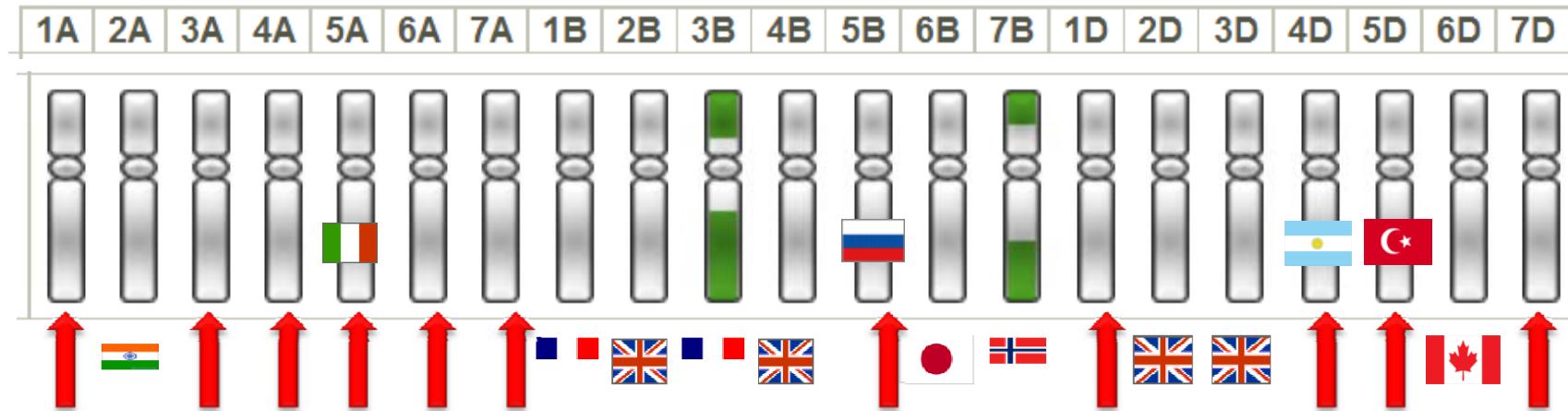
- **Filtering**  
No homology : 7282  
Similarity to Transposase: 44
- **Automated validation**  
Start/stop, splicing sites : 2 815 (34%)

**9 351 genes selected**

BlastClust.pl

- Identity : 99 %
- Overlap : 90 %

# IWGSC MTP sequencing



The screenshot shows the homepage of the International Wheat Genome Sequencing Consortium (IWGSC) website. The header features the IWGSC logo and the text "International Wheat Genome Sequencing Consortium". The top navigation bar includes links for "Login", "Register", "Contacts", "Site map", and "Sponsors". Below the header is a search bar with a "Search" button. The main content area features a large image of a wheat field. A navigation menu at the bottom includes "News and Reports", "Organization", "Projects" (which is highlighted in green), "Tools and Resources", and "General Documents".

[www.wheatgenome.org](http://www.wheatgenome.org)