The IWGSC:
Strategies & Activities to Sequence the Bread Wheat Genome

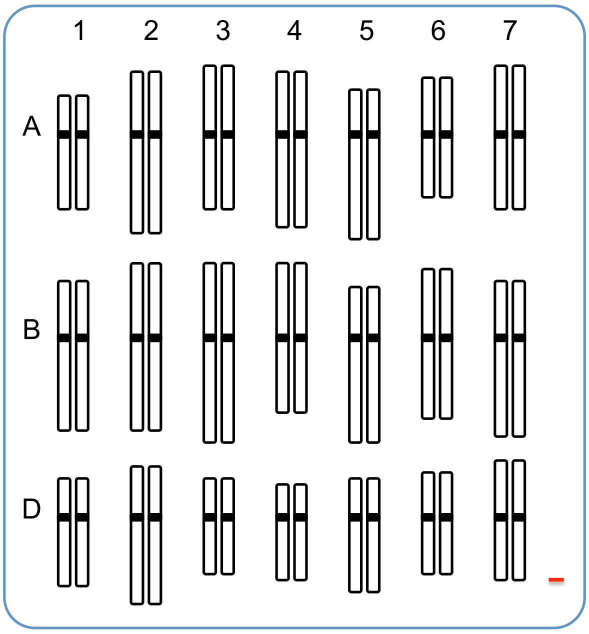
Frédéric CHOULET

INRA GDEC – Clermont-Ferrand, France

Streptomyces

8 Mb

❑ Bread Wheat Genome



❑ 17 Gb
❑ Hexaploid

Launched in 2005
on the initiative of
wheat growers

❑ The IWGSC

Executive director	K. Eversole J. Rogers
Co-chairs	C. Feuillet B. Keller R. Appels J. Dvorak B. Gill Y. Ogihara
Coordinating committee	65 members
General members	>800 members

22 sponsors

Arcadia Biosciences
ACPFG
Bayer CropScience
Biogemma
Céréales Vallée
CIMMYT
CSIRO
Dow AgroSciences
DuPont
Evogene
Florimond Desprez
GRDC
Graminor
HPI
INRA
Kansas Wheat
KWS
Limagrain
Monsanto
Punjab Ag Univ
RAGT
Syngenta

❑ The IWGSC

○ Why?

- Lay a foundation to accelerate wheat improvement

○ Vision

- High quality annotated genome sequence
- Ordered sequence linked to markers & phenotypes

❑ Wheat Improvement Is Complex

- Yield potential and stability
- Adaptation to climate change
- Durable resistance to pathogens
- Quality of grains

❑ An integrated toolbox

Genetics Resources
- Mapping pop.
- Association panels
- Genebanks
- Mutant pop.

Genomics Resources
- Physical maps
- Genome/Transcriptome
- HT genotyping platforms
- Bioinformatics

Phenotyping data
- in controlled conditions
- in the fields
- imaging technologies

Methodologies
- Modeling
- GM approaches
- Breeding methods

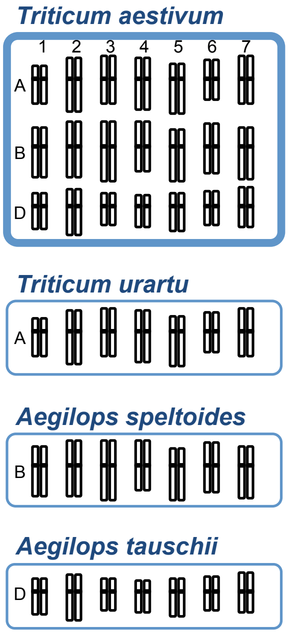
Agronomy
- agricultural practices, policies

Improved wheat varieties

❑ Goal and Strategy

- Produce a useful sequence:
  - Of what grows in the fields
  - Link to genetic maps and phenotypes

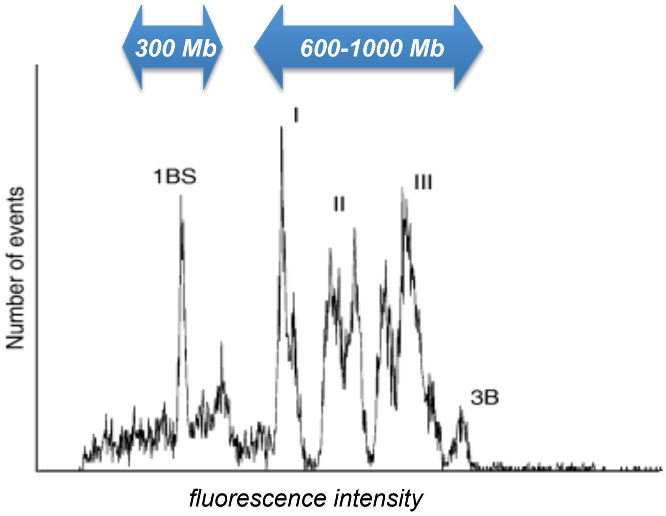
- Reduce the genome complexity
  - Flow sorting chromosomes
  - Physical map (BACs)
  - Sequence BACs

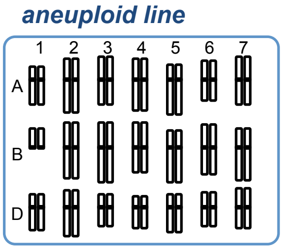


❑ Flow sorting

Laboratory of
Molecular Cytogenetics and Cytometry
Institute of Experimental Botany AS CR

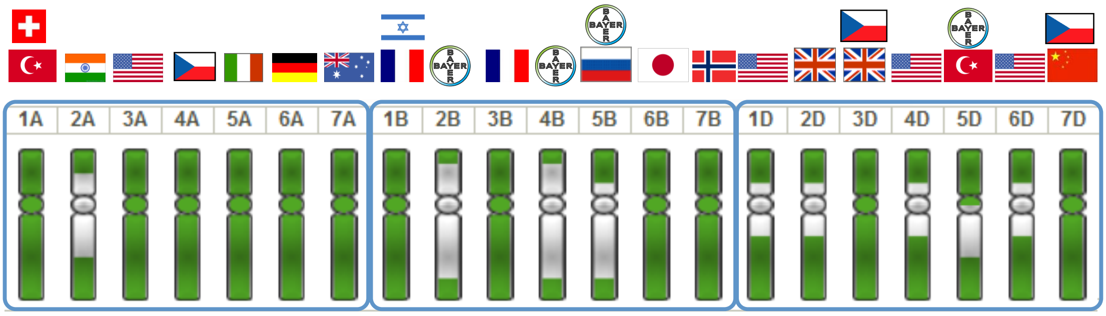
J. Dolezel





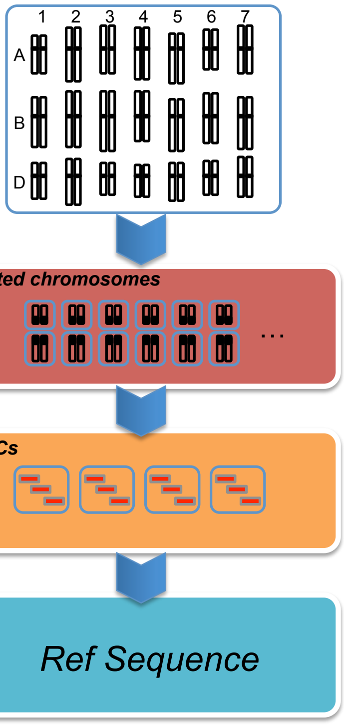
➔ 2014 – all BAC libraries available @ CNRGV

❑ Current status – Physical maps



www.wheatgenome.org

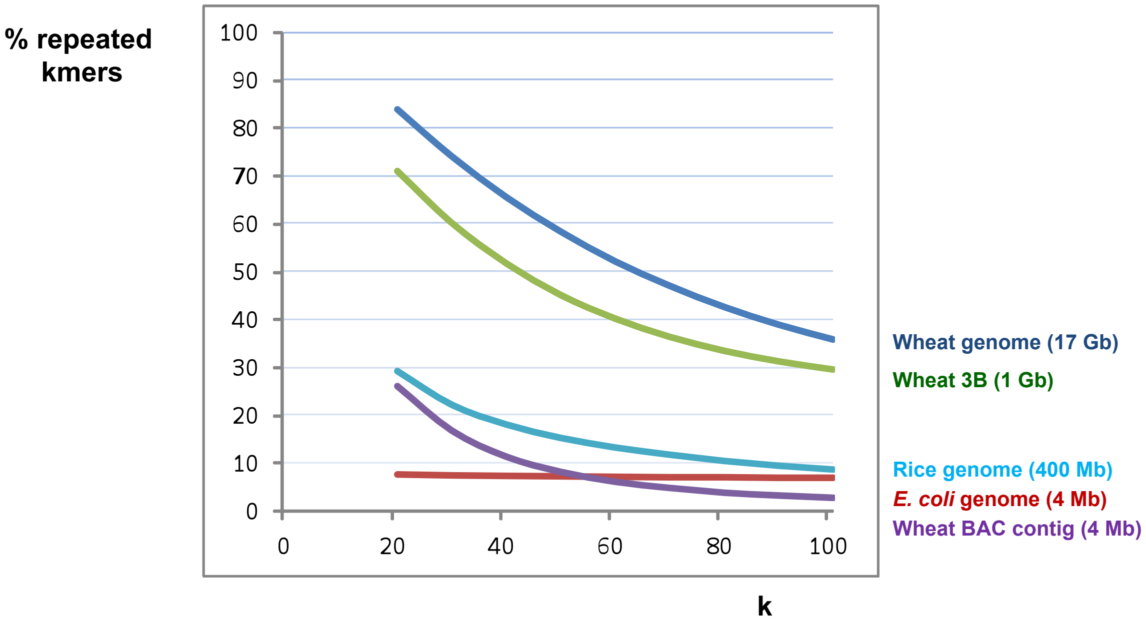
❑ IWGSC strategy

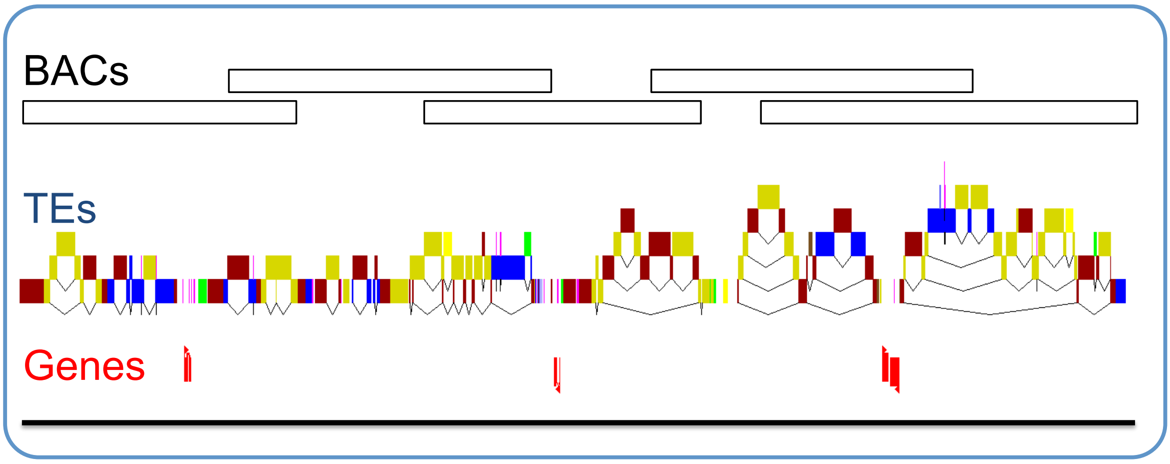


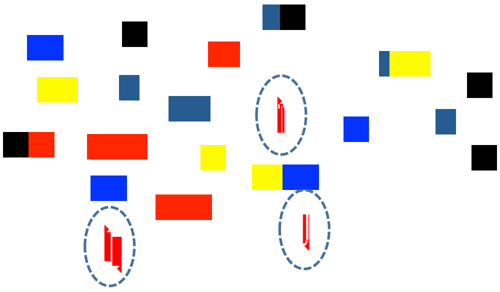
sorted chromosomes

BACs

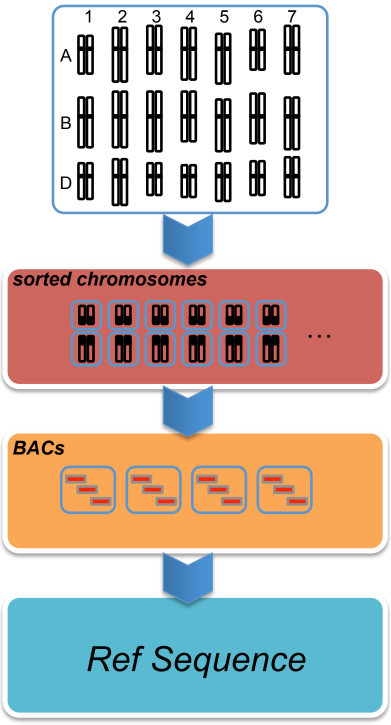
Ref Sequence



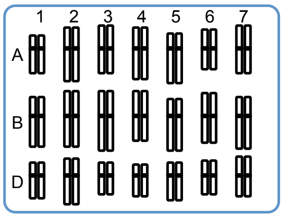




Whole genome shotgun assembly



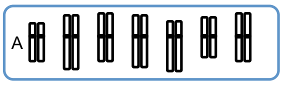
❑ WGS



Roche/454 – depth 5x

WGS *Brenchley et al. Nature 2012*

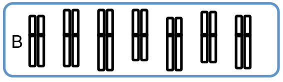
Triticum urartu



Illumina (PE+MP lib.)

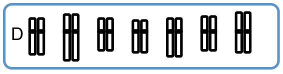
WGS *Ling et al. Nature 2013*

Aegilops speltoides



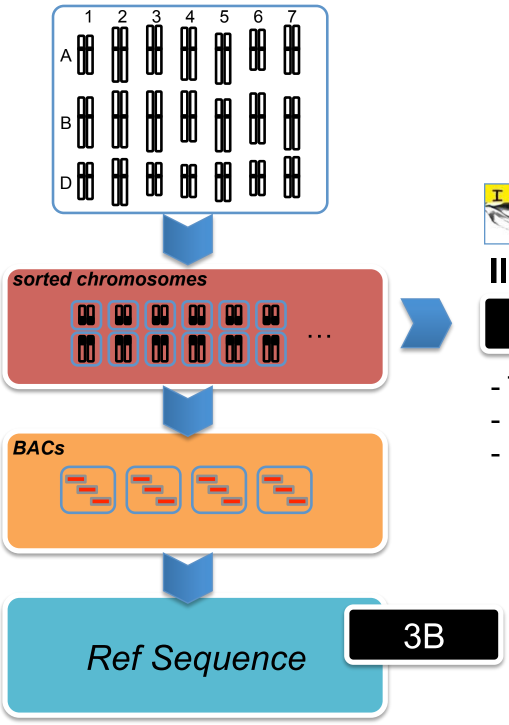
Scaffold N50: 50 kb

Aegilops tauschii



WGS *Jia et al. Nature 2013*

❑ Survey Sequences



sorted chromosomes

BACs

Ref Sequence

3B

IWGSC

Illumina 80x (PE 2x100bp)

Chr. Survey Sequences

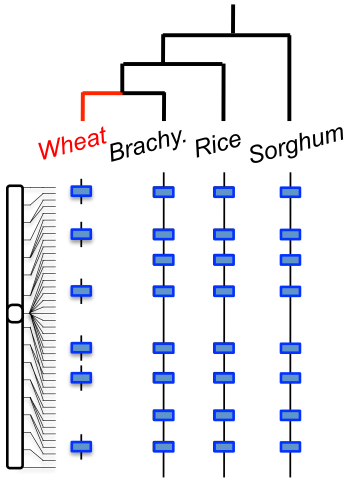
- TGAC, Norwich
- IEB, Olomouc
- MIPS IBIS, Munich

❑ Survey Sequences

- ○ 1 **assembly** per chr. arm
  - 10 Gb
  - 10 million contigs
  - contig N50: 2.4 kb

- ○ **Annotation**:
  124,000 genes (HC)
  - AA: 40,000
  - BB: 45,000
  - DD: 39,000

- ○ **"GenomeZipper"** *(Mayer et al, MIPS)*



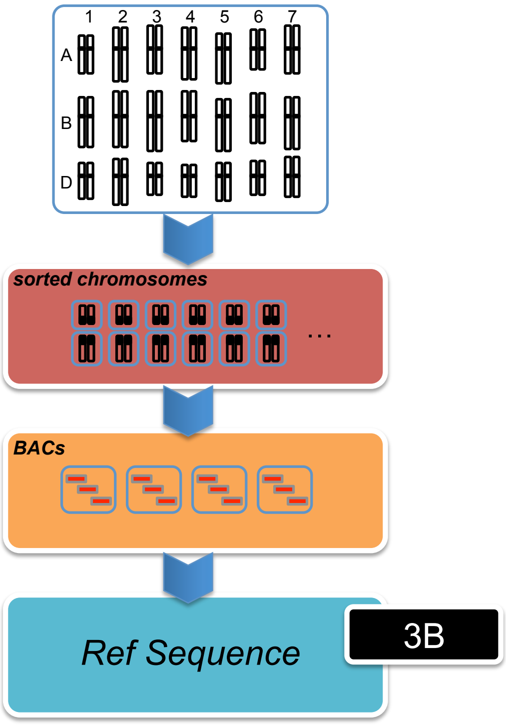
66,000 genes "zipped"

+9,000 genes anchored with POPSeq approach *(Stein et al, IPK)*

❑ Survey Sequences

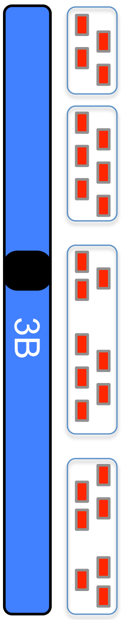
- ○ **Results**
  - Limited gene loss after polyploidization
  - High level of "small scale duplications"
  - No subgenome dominance
  - Phylogeny A-B-D
  - Core and pan genome of wheat
  - 3.5 M markers mapped
  - 13 M SNPs from POPSeq
  - Homoeolog-specific gene expression

❑ 3BSEQ



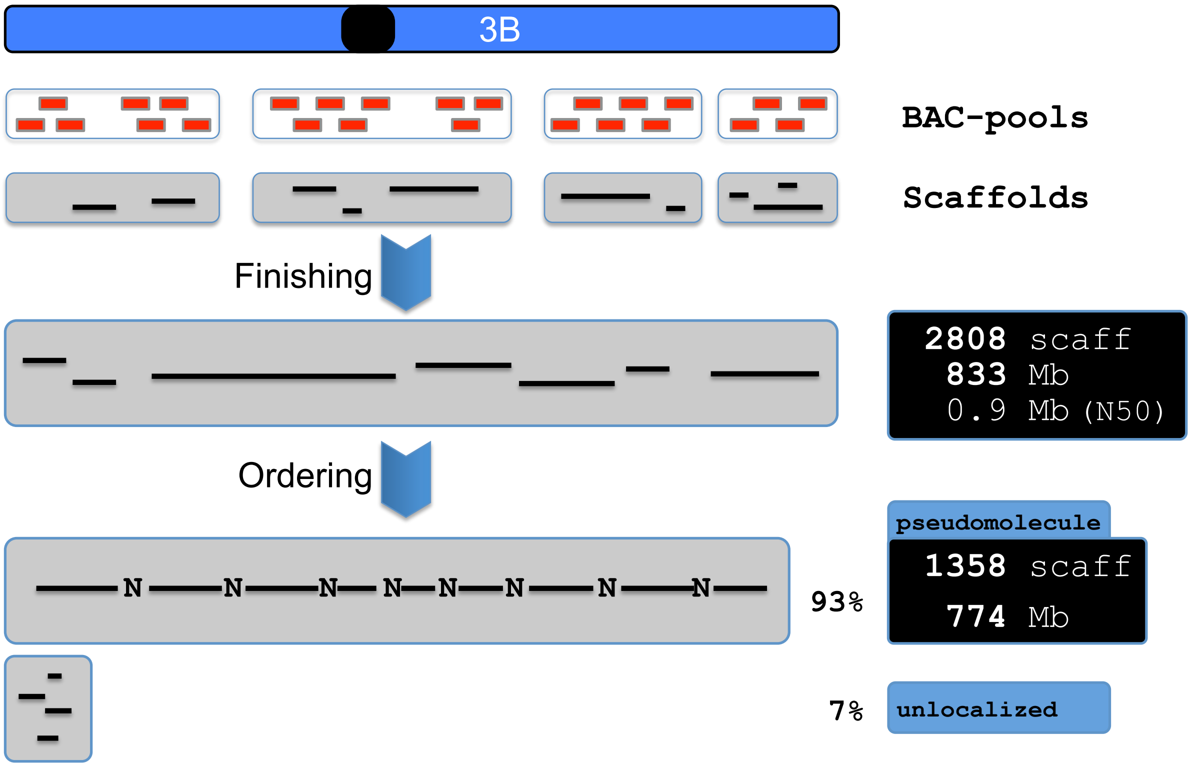
- INRA GDEC *(C. Feuillet)*
- CEA-IG-Genoscope *(P. Wincker)*
- INRA URGI *(H. Quesneville)*
- INRA CNRGV *(H. Berges)*

❑ 3B MTP-Sequencing



• #BACs	8452
• #BAC pools	922
• #Roche 8 kb MP lib.	922
• #GS-FLX runs	150
• bp coverage (Roche/454)	36x
• BAC-ends (Sanger)	42,551
• 3B Survey Seq (Illumina)	82x

❑ Building a pseudomolecule



2808 scaff
833 Mb
0.9 Mb (N50)

93%

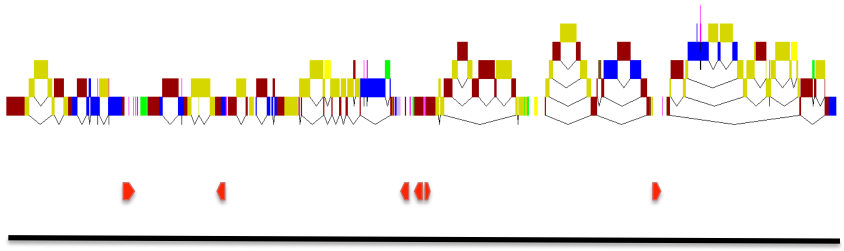
pseudomolecule

1358 scaff
774 Mb

7%

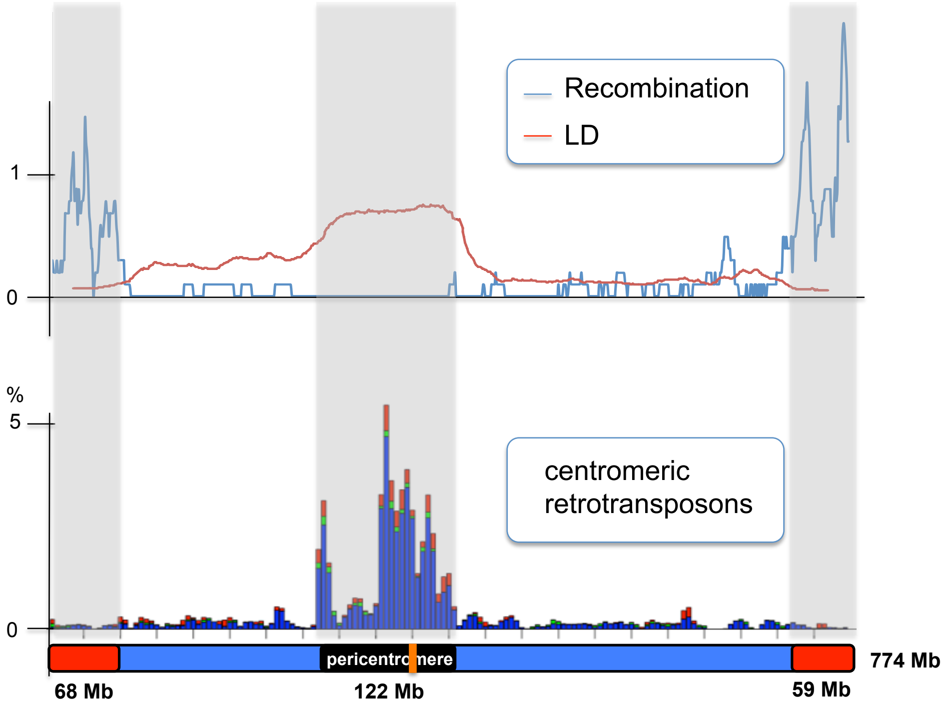
unlocalized

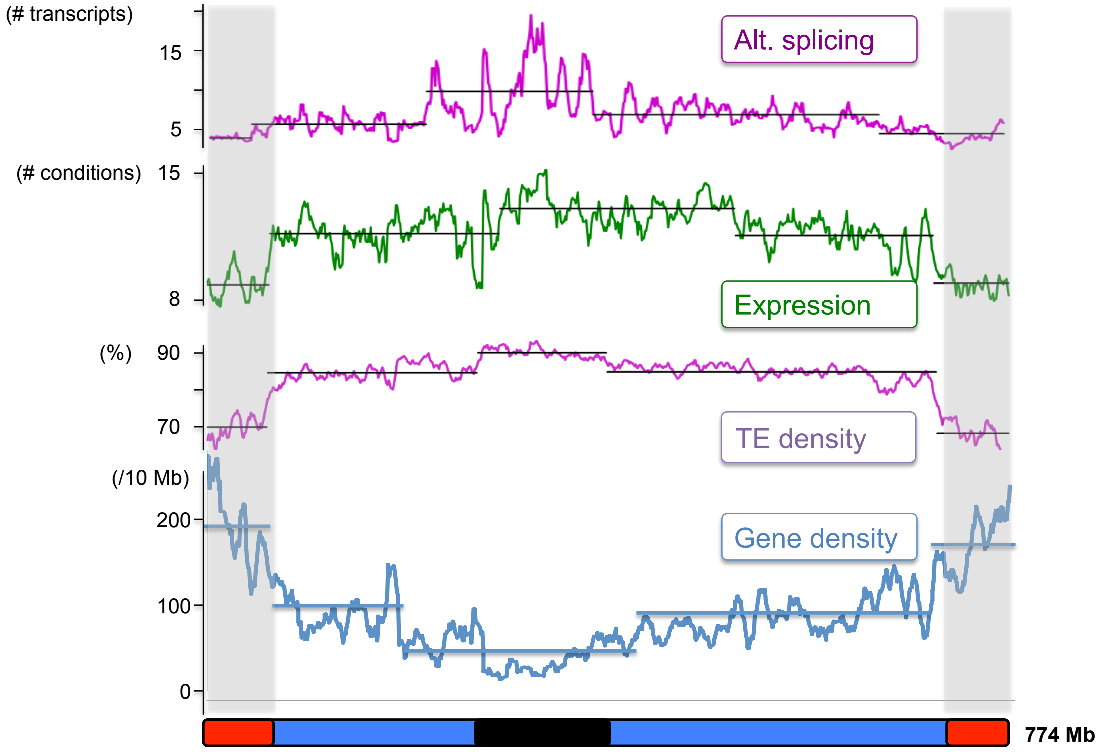
❑ Annotation



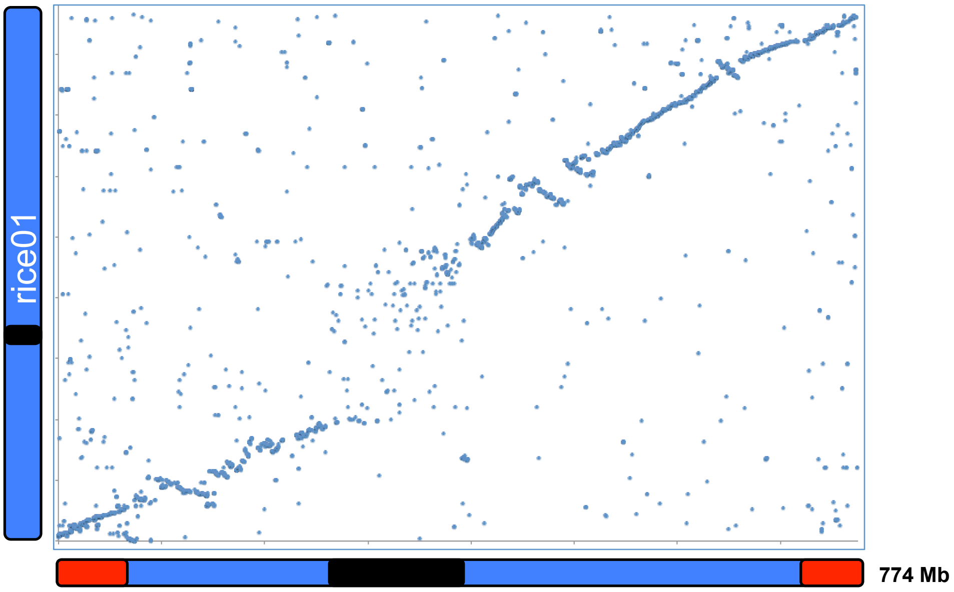
	pseudomolecule
• protein coding genes	7,264
(pseudogenes: 27%)	
• transposable elements	234,606 (85%)

❑ Structure, Recombination



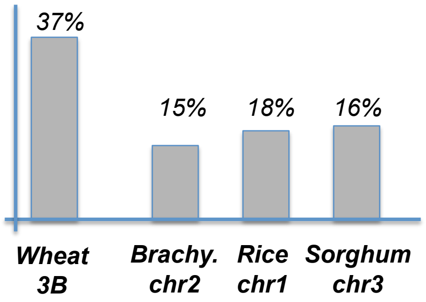


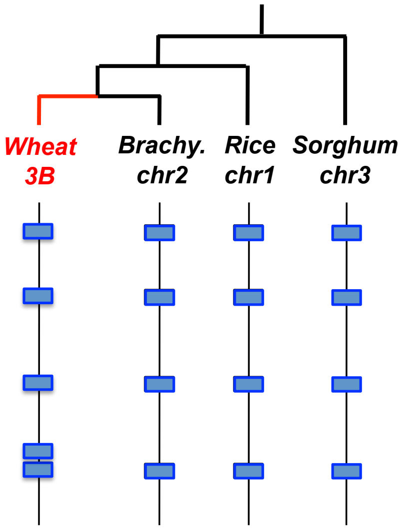
❑ Synteny



❑ Lots of small-scale duplications

- ○ *intra-chr. duplicates*

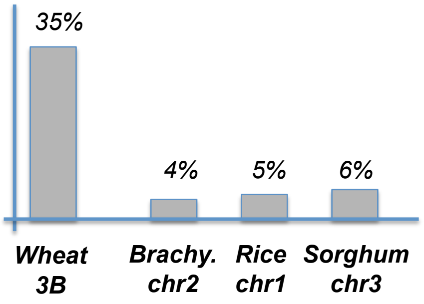


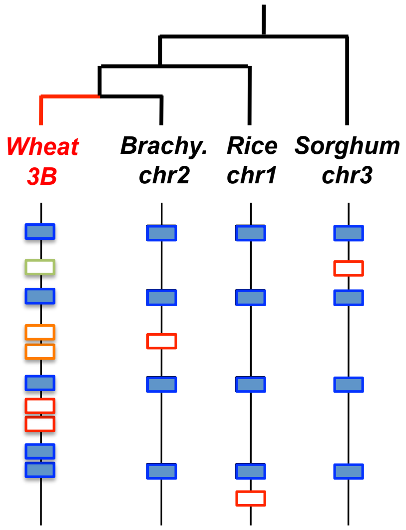


# syntenic genes	3899	3400-3600

❑ Lots of small-scale duplications

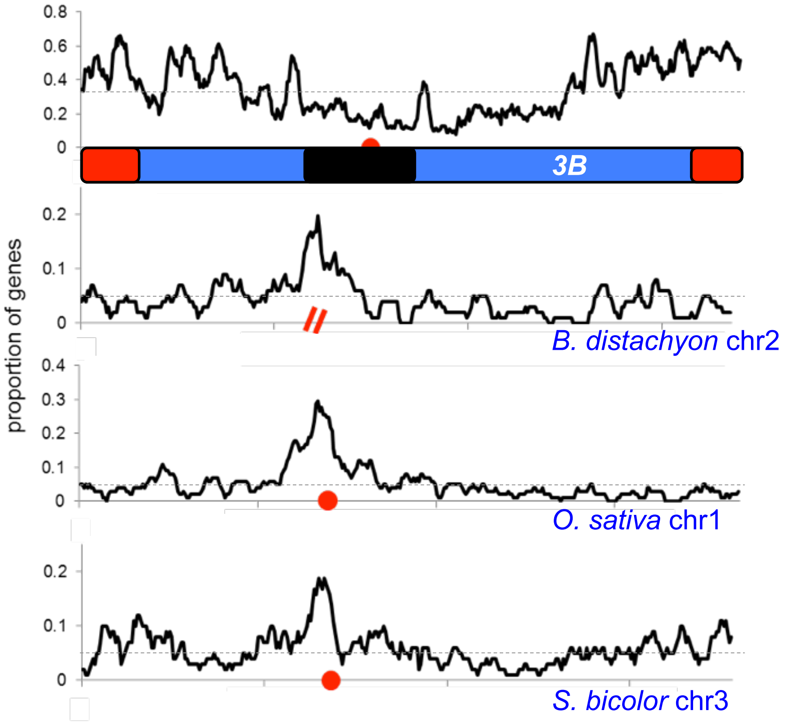
- ***intra-chr. duplicates***
- ***inter-chr. duplicates***





# syntenic genes	3899	3400-3600
# nonsyntenic genes	2065	150-210

❑ Nonsyntenic gene distribution



❑ Structural and functional Partitioning

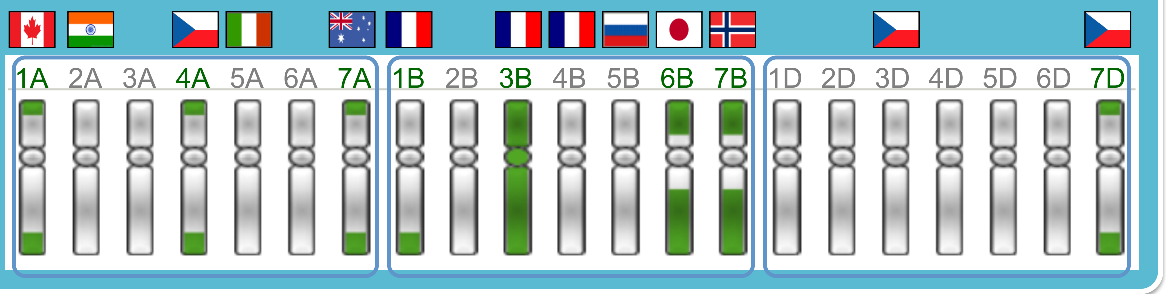
❑ Map-based cloning on 3B

➔ 13 gene cloning projects

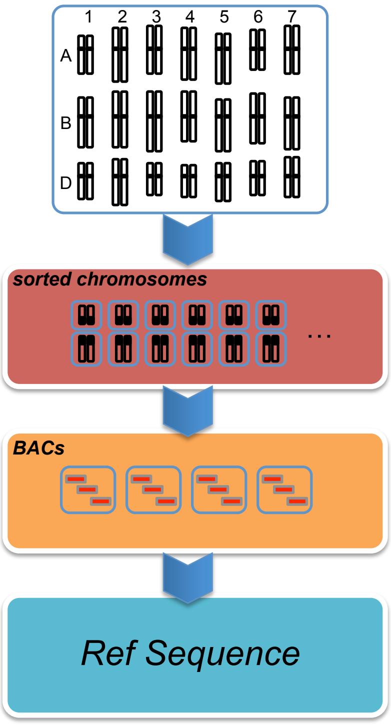
- Disease resistance genes (Sr, Lr, Yr, Stb…)
- Solid stem (saw fly)
- Yield
- Drought tolerance
- Boron transporter
- Flowering time
- Nitrogen Use Efficiency
- Chromosome pairing…

❑ Pseudomolecules, current status

○ *Current Status*



❑ IWGSC resource access



sorted chromosomes

BACs

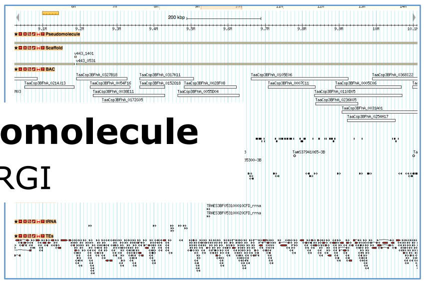
Ref Sequence

○ **Survey Seq**

http://wheat-urgi.versailles.inra.fr/Seq-Repository/

http://plants.ensembl.org/Triticum_aestivum

○ **3B pseudomolecule**
Gbrowse @ URGI



Acknowledgments – 3B team

GDEC

Génétique Diversité Ecophysiologie des Céréales

Catherine Feuillet
Etienne Paux
Pierre Sourdille
François Balfourier
Philippe Leroy
Nicolas Guilhot
Jacques Le Gouis
Sébastien Theil
Lise Pingault
Josquin Daron
Natasha Glover
Nelly Cubizolles
Elodie Rey

Genoscope

A. Alberti
V. Barbe
J. Poulain
S. Durand
S. Mangenot
JM. Aury
A. Couloux
P. Wincker

BIA

C. Gaspin

VIB

K. Vandepoele

MIPS

K. Mayer et al.

URGI

M. Alaux
L. Couderc
V. Jamilloux
H. Quenesville

CNRGV

H. Berges
A. Bellec

IEB

J. Dolezel
J. Safar

TGAC

J. Rogers, M. Caccamo et al.

SAB

P. Schnable
S. Rounsley
D. Ware
J. Rogers
K. Eversole