International
Wheat Genome
Sequencing
Consortium

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

Jane Rogers

IWGSC Deputy Executive Director

ACPFG Seminar
University of Adelaide
1st May 2014

The International Wheat Genome Sequencing Consortium

Launched in 2005 on the initiative of Kansas Growers

Executive director
K. Eversole

Co-chairs

R. Appels
J. Dvorak
C. Feuillet
B. Gill
B. Keller
Y. Ogihara

General members (> 800)

Coordinating Committee (65)

Sponsors (23)

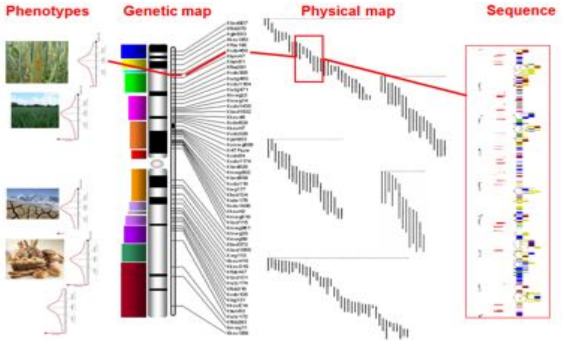
Arcadia
ACPFG
biogemma
Bayer CropScience
CSIRO
CIMMYT
Céréales Vallée
Dow
Dow AgroSciences
DuPont
The miracles of science
evogene
Florimond Desprez
GRDC
Grains Research & Development Corporation
Graminor
Heartland Plant Innovations
INRA
KWS
Kansas Wheat
Limagrain
Monsanto
RAGT
syngenta

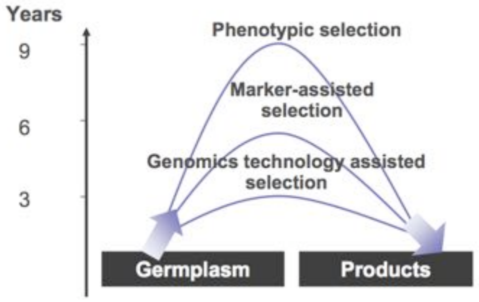
Why

- Lay a foundation to accelerate wheat improvement
- Increase profitability throughout the industry

Vision

- High quality annotated genome sequence, comparable to rice genome sequence
- Physical map-based, integrated and ordered sequence





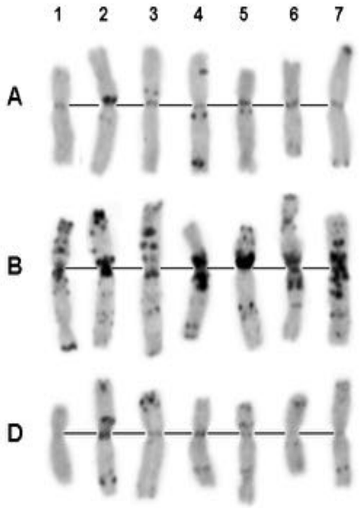
Wheat Improvement is Complex

- ➢ **Yield potential and yield stability**

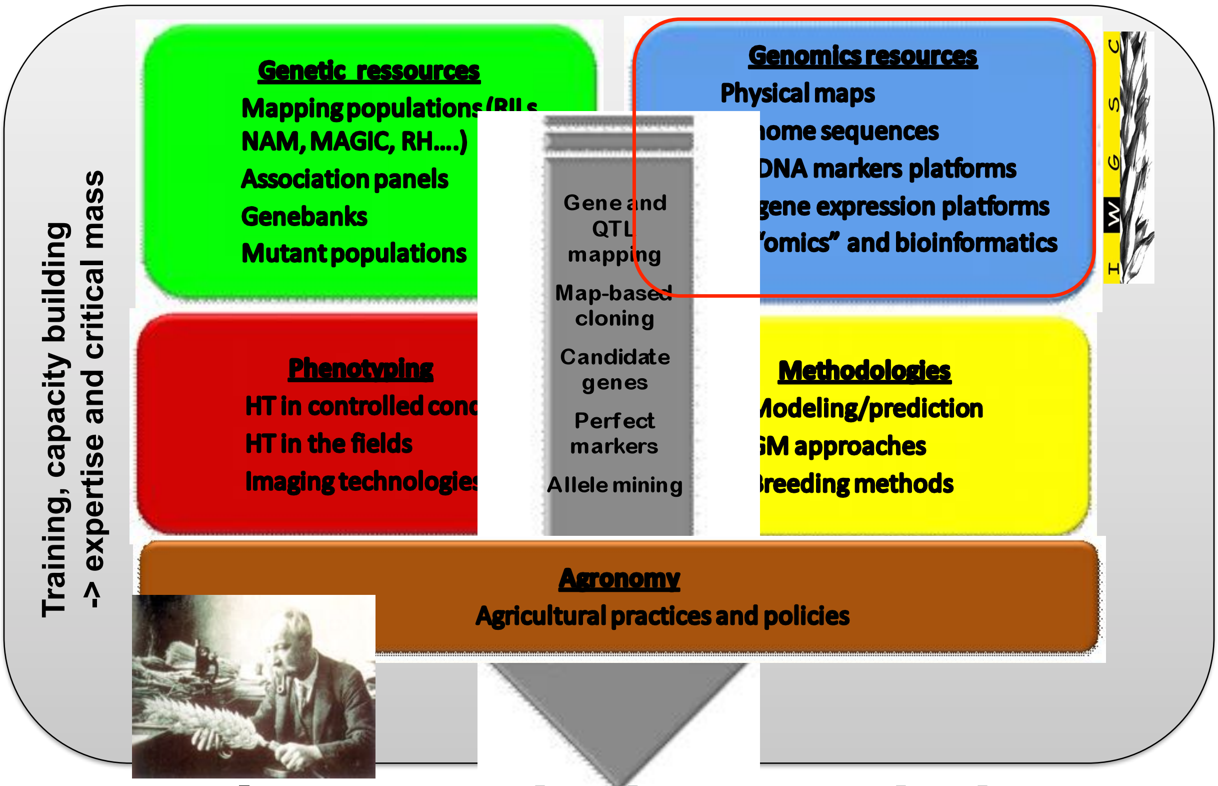
- ➢ **Adaptation to climate change**

- ➢ **Durable resistance to biotic stress**

- ➢ **Quality of grain and co-products**



Complexity Requires an Integrated Toolbox



Training, capacity building
-> expertise and critical mass

Genetic ressources
Mapping populations (RILs
NAM, MAGIC, RH….)
Association panels
Genebanks
Mutant populations

Genomics resources
Physical maps
nome sequences
DNA markers platforms
gene expression platforms
'omics" and bioinformatics

Gene and
QTL
mapping
Map-based
cloning
Candidate
genes
Perfect
markers
Allele mining

Phenotyping
HT in controlled cond
HT in the fields
Imaging technologie

Methodologies
Modeling/prediction
GM approaches
Breeding methods

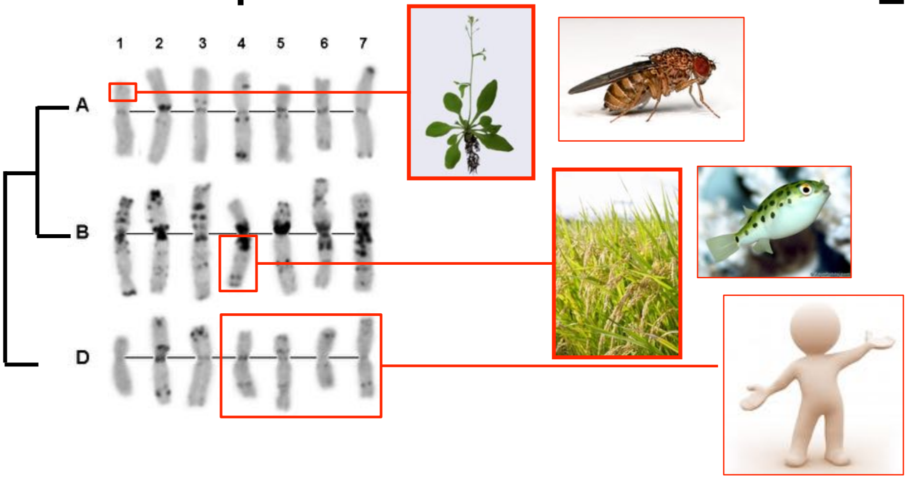
Agronomy
Agricultural practices and policies

Improved wheat varieties

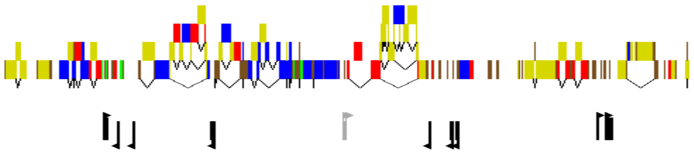
The bread wheat genome is……a challenge

✓**Allohexaploid**

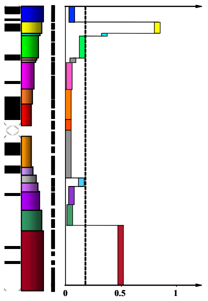
✓ **Large: 17** Gb



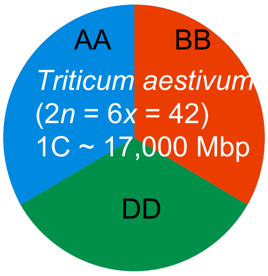
✓ **80-90% of TEs and repeats**

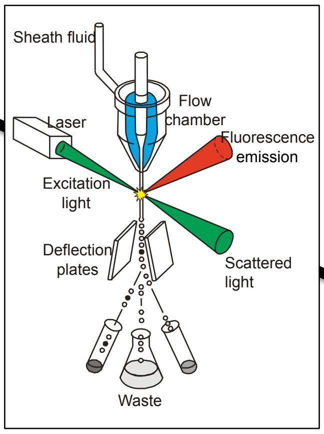


✓ **> 50% of non recombinogenic regions**

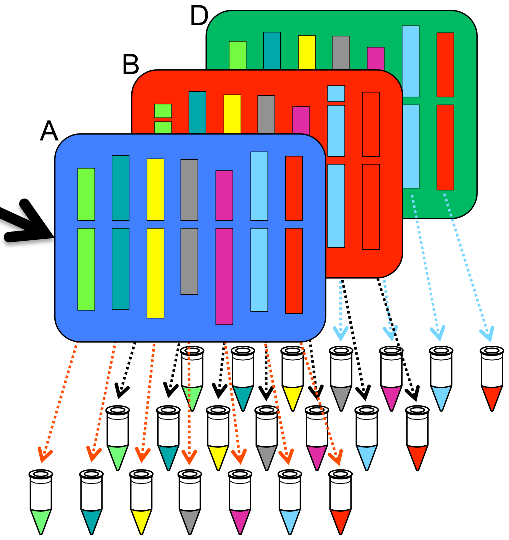


Managing the 17 Gb Hexaploid Genome





Dissection of the genome into single chromosomes (arms)



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

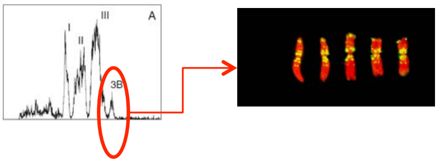
- **Chromosomes: 605 - 995 Mbp (3.6 – 5.9% of the genome)**
- **Chromosome arms: 225 - 585 Mbp (1.3 – 3.4% of the genome)**

Chromosome genomics

- **Chromosome specific BAC libraries and sequencing**

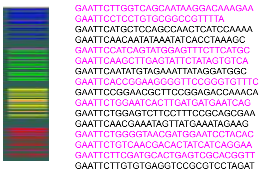
IEB

IWGSC Strategy for a Reference Wheat Genome

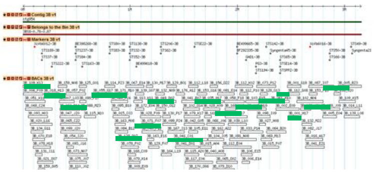


1. BAC library construction

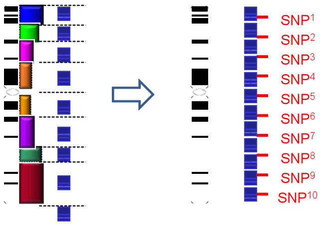
2. BAC fingerprinting (HICF/WGP)



3. Contig assembly by FPC/LTC



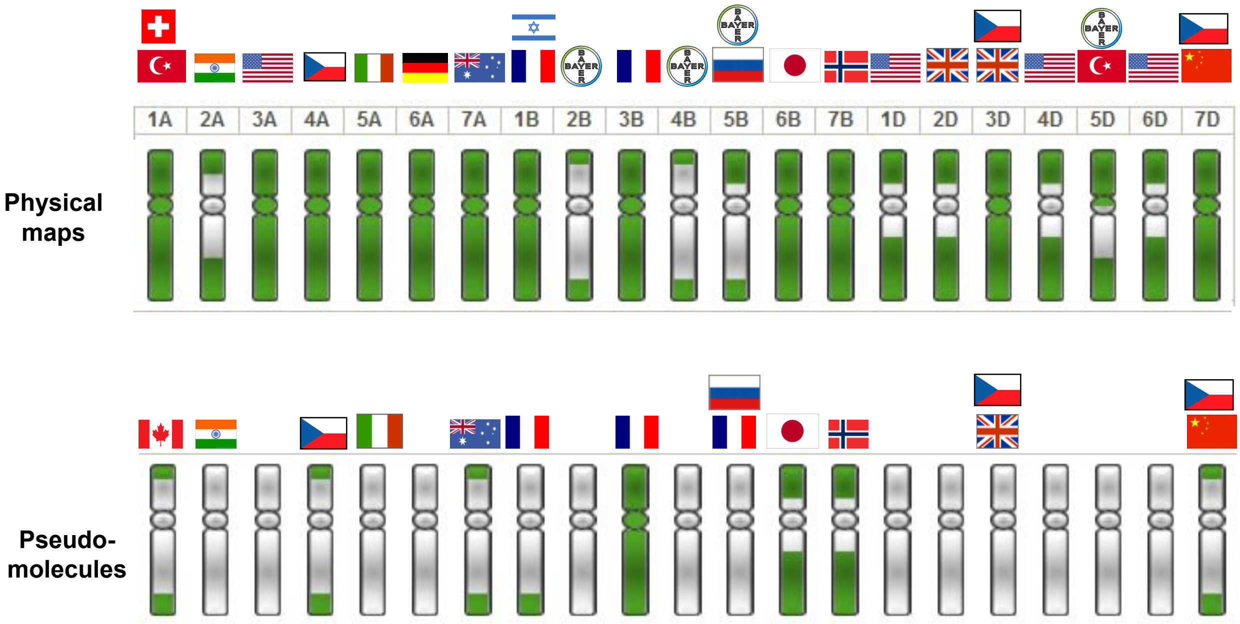
4. MTP sequencing /Scaffold assembly



5. Pseudomolecule construction (meiotic/LD/RH mapping)

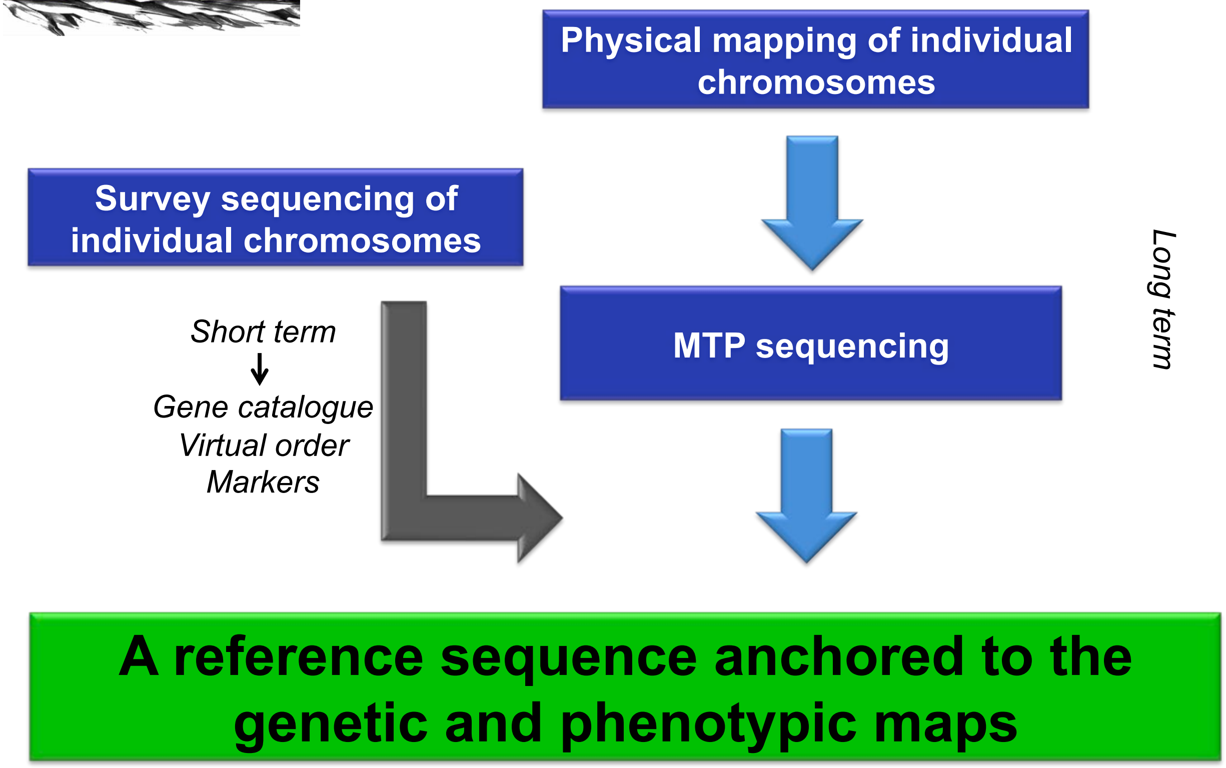
6. Automated and curated annotation

Current status of individual projects



Roadmap to the Wheat Genome Sequence

IWGSC



Physical mapping of individual chromosomes

Survey sequencing of individual chromosomes

MTP sequencing

Short term
↓
Gene catalogue
Virtual order
Markers

Long term

A reference sequence anchored to the genetic and phenotypic maps

IWGSC Chromosome Survey Sequencing Initiative

TGAC The Genome Analysis Centre™

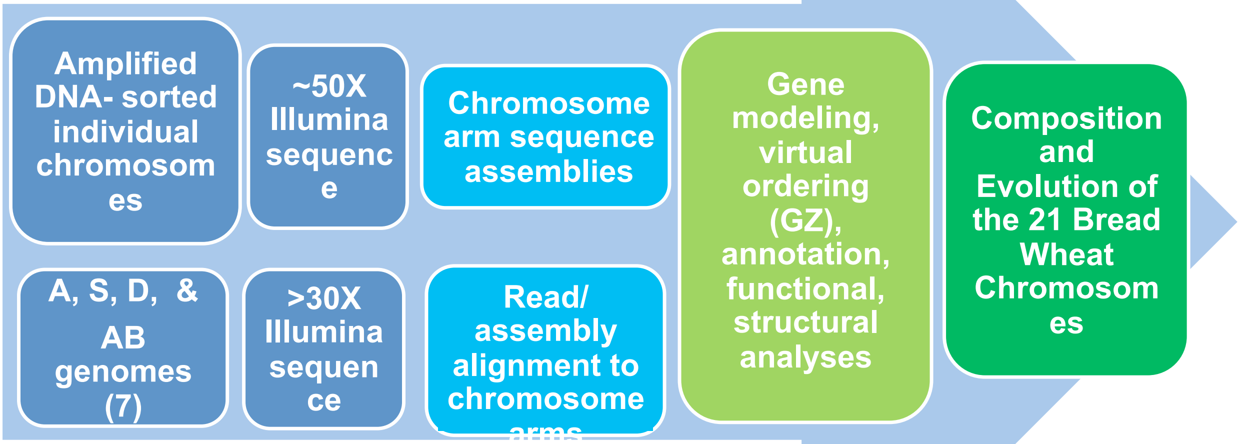
Institute of Experimental Botany of the AS CR, v. v. i.

INRA

HelmholtzZentrum münchen
German Research Center for Environmental Health

biogemma

Graminor



Amplified DNA- sorted individual chromosomes

A, S, D, & AB genomes (7)

~50X Illumina sequence

>30X Illumina sequence

Chromosome arm sequence assemblies

Read/ assembly alignment to chromosome arms

Gene modeling, virtual ordering (GZ), annotation, functional, structural analyses

Composition and Evolution of the 21 Bread Wheat Chromosomes

WGRC Wheat Genetic & Genomic Resources Center

ANR

University of Saskatchewan

USDA NIFA

ICARDA

BBSRC bioscience for the future

The Research Council of Norway

Department of Biotechnology
Ministry of Science & Technology,
Government of India

Australian Government
Australian Research Council

Bioplatforms Australia

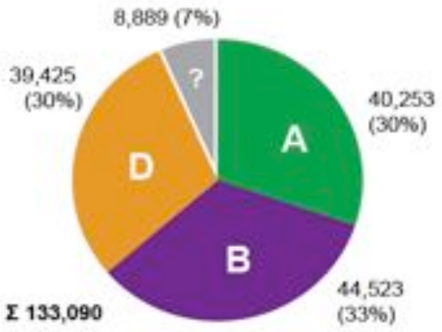
NIAS

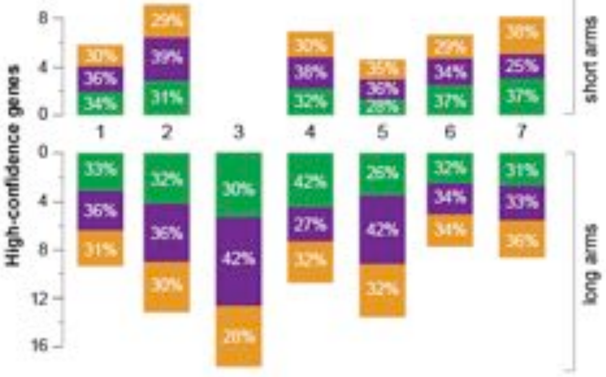
Together, we are Kansas Wheat

Chromosome Survey Sequence Data

- **Sequence assemblies for 40 chromosome arms + chromosome 3B**
  - ❖ Total length **10.2 Gb**
  - ❖ assembled sequence per chromosome **128Mb (1DS) – 639Mb (3B)**
  - ❖ contigs / chromosome arm range from **88k (6DS) – 508k (2DL)**
  - ❖ N50 contig length = 2.4 kb (1.5kb - 4.3kb) after filtering contigs < 200bp
  - ❖ N50 contig length after repeat masking = **6.1kb** (1.7kb - 8.9kb)

- **Gene annotation** (RNASeq, FLcDNA, grass genomes): **124,201 allocated to chromosomes**



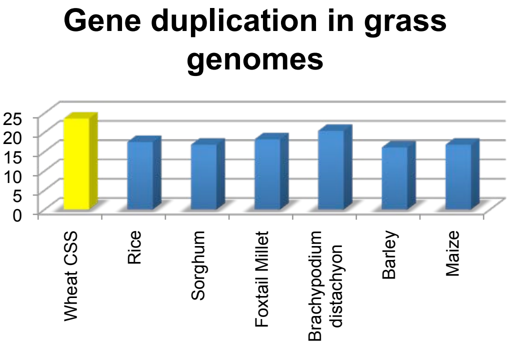


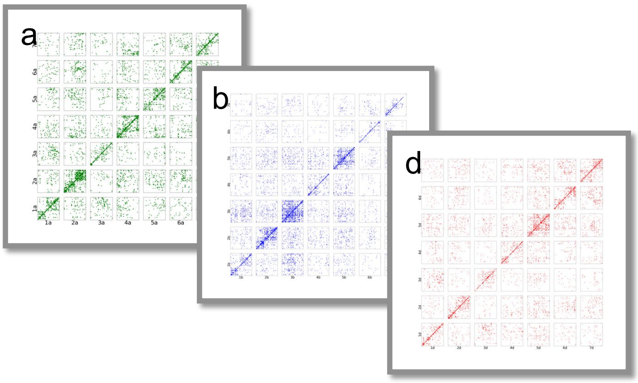
IWGSC Chromosome Survey Sequence

- Almost full wheat gene complement identified and allocated to chromosome arms
- ~ 70% genes are alternatively spliced, ca 3.5 transcripts / gene
- Using Klaus Mayer’s GenomeZipper synteny-based approach, on average 53% of genes virtually ordered along chromosomes
- Over 3.5 M markers mapped to contigs:

  1.3M wheat markers + 2.3M SNPs including SSR, EST, DArT, SNP (90k) markers...
- 13.2 million SNPs from POPSeq aligned to contigs

Gene duplications are abundant in wheat



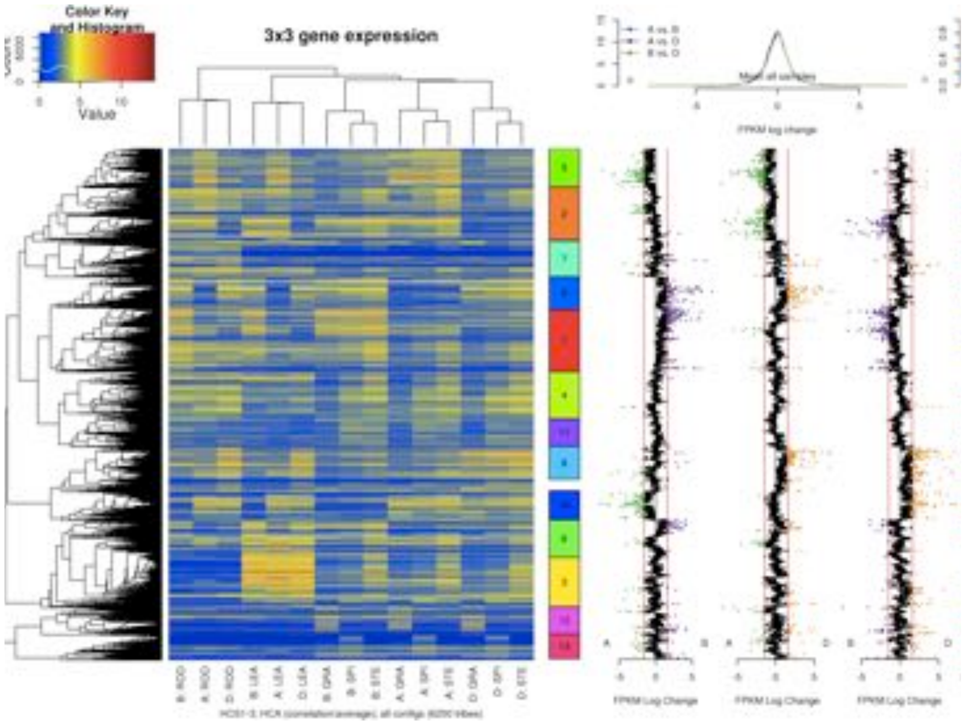


Duplications in wheat underestimated due to sequence fragmentation

73% of duplicates are located on the same chromosome arm

Differential gene expression in individual tissues

RNASeq: 5 tissues, 3 developmental stages – L. Pingault, E.Paux



Origins of Bread Wheat Genomes

Triticum monococcum
- Illumina
- Assembly (MIPSv3):

Triticum urartu
- Illumina
- Assembly (TGACv1)

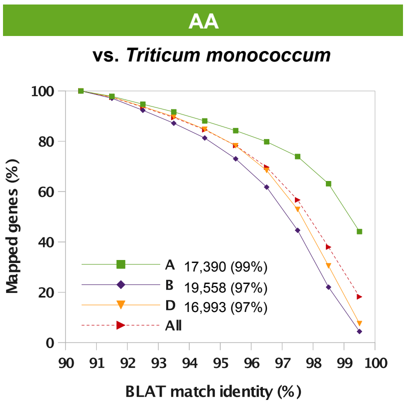
Aegilop. speltoides
- Illumina
- Assembly (TGACv1)

Aegilops sharonensis
- Illumina
- Assembly (B. Steuernagel)

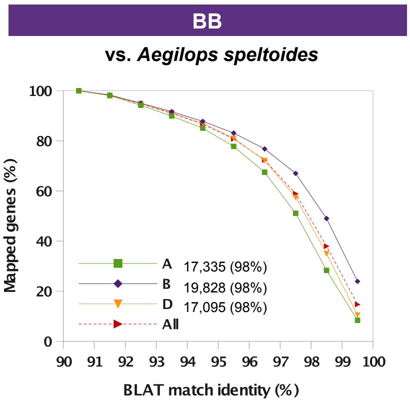
Triticum durum (CN)
Triticum durum (IT)
- Illumina
- Assemblies (TGAC v1)

Aegilops tauschii
- Assembly (BGI): 4.1Gb (N50: 50kb)

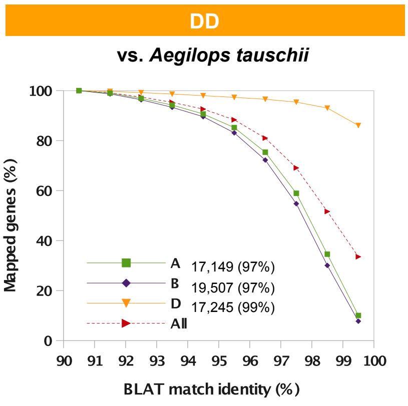
Relationship of Bread Wheat to diploid species


AA

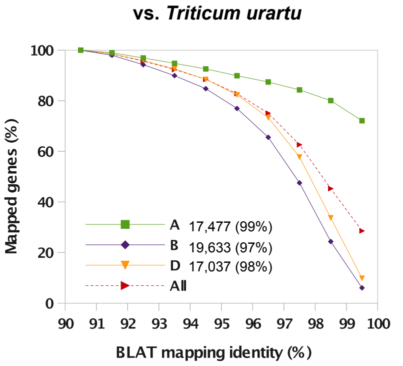
vs. *Triticum monococcum*

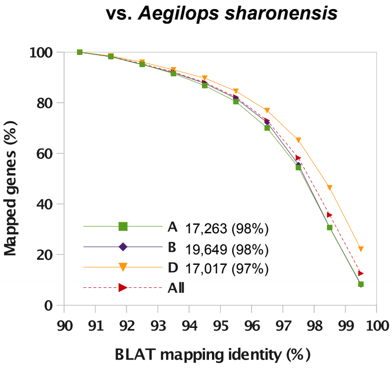

BB

vs. *Aegilops speltoides*


DD

vs. *Aegilops tauschii*


vs. *Triticum urartu*

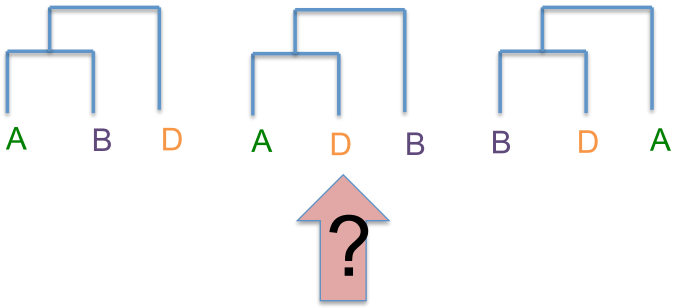

vs. *Aegilops sharonensis*

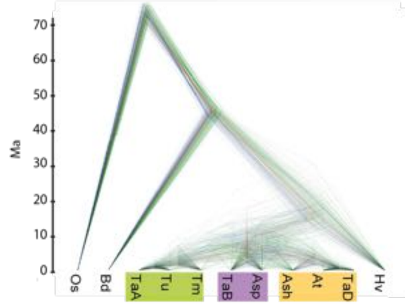
- BLAT: HCS 1 CDS vs. Assemblies
- Best match
  i. max coverage
  ii. max match identity

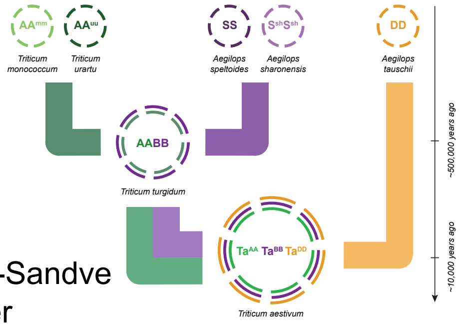
17,635 HCS1 A
20,144 HCS1 B
17,470 HCS1 D

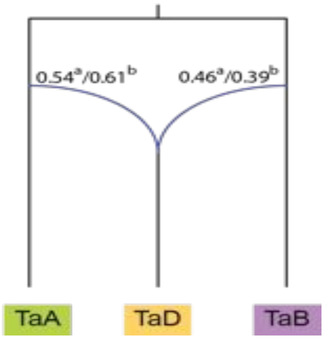
55,249 HCS1

Origins of Bread Wheat genomes





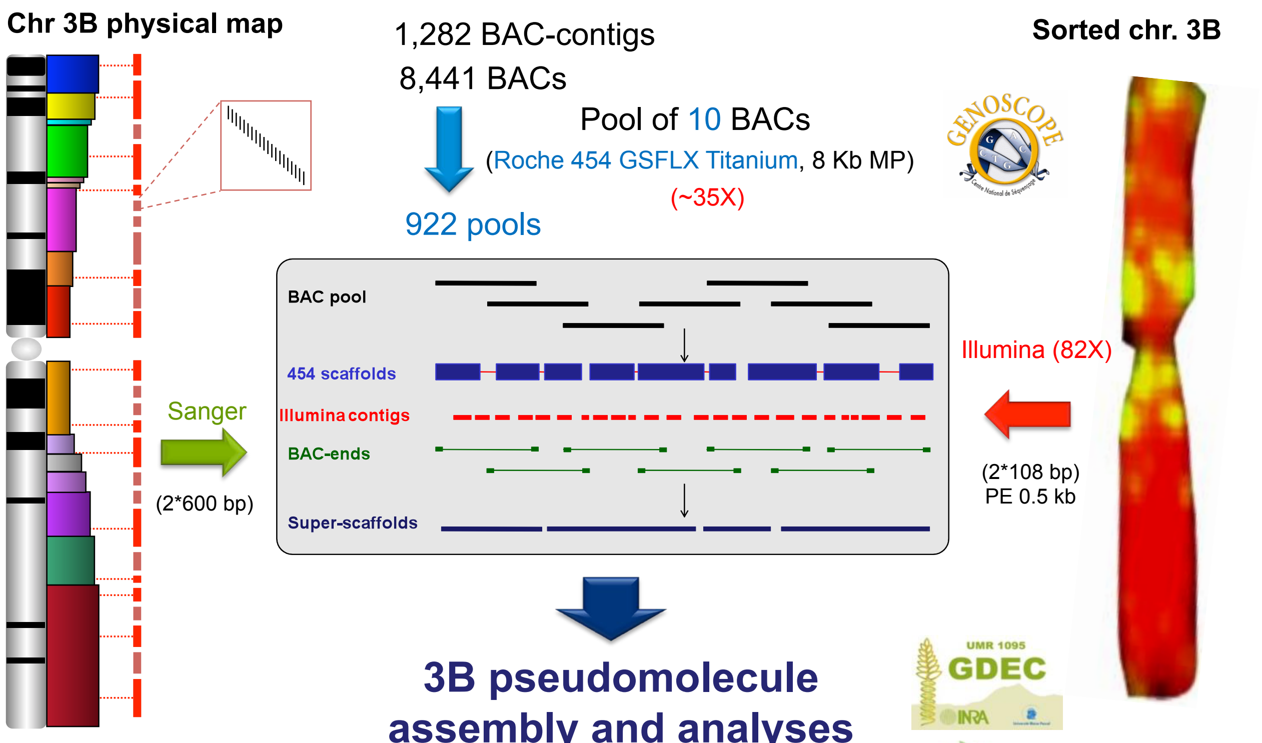




Simen Rød-Sandve
Klaus Mayer

Topological incrongruence points to hybridisation between A and B progenitors
D lineage is a hybrid of A and B lineages

3B SEQuencing Project



Chr 3B physical map

1,282 BAC-contigs

8,441 BACs

Pool of 10 BACs

(Roche 454 GSFLX Titanium, 8 Kb MP)

(~35X)

922 pools

Sorted chr. 3B

Sanger

(2*600 bp)

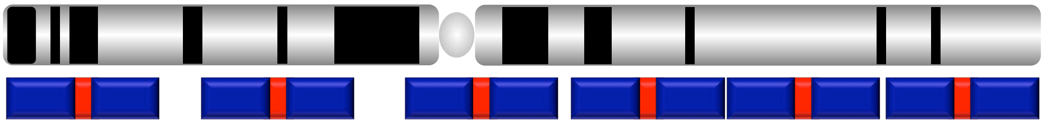
Illumina (82X)

(2*108 bp)
PE 0.5 kb

3B pseudomolecule assembly and analyses

Catherine Feuillet, Fred Choulet et al

The 3B pseudomolecule



No. Scaffolds= 1358 (Max 4.2 Mb)

Total sequence = 774.4 Mb

N50 = 949 Kb

% N = 6 %

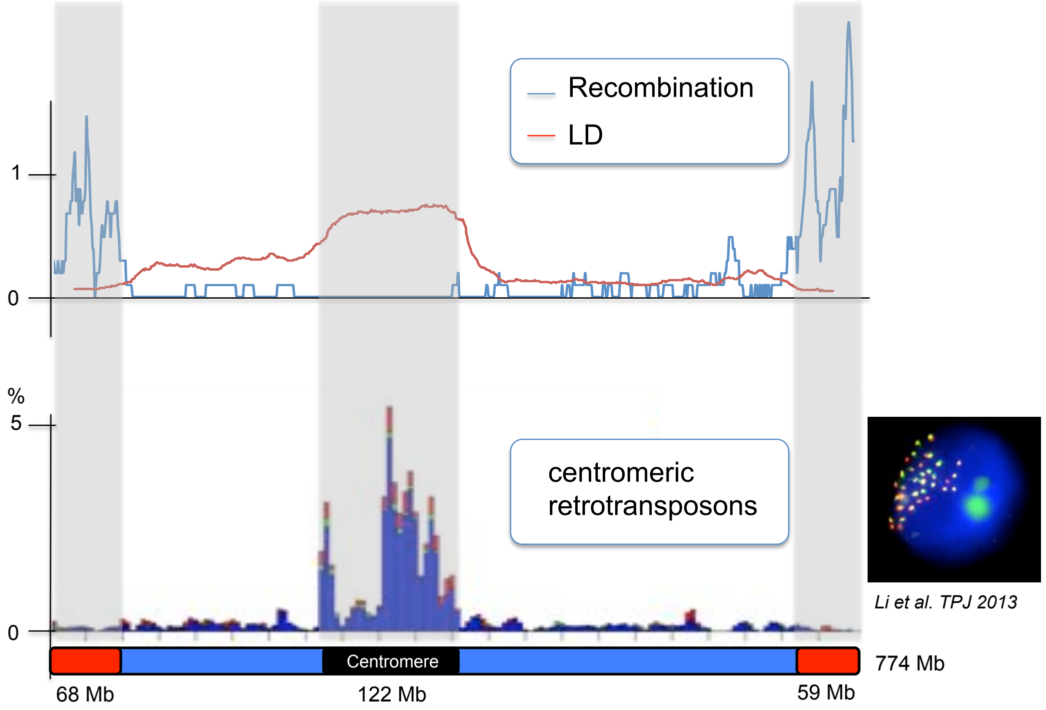
RNA-seq on samples from 5 organs x 3 stages

TriAnnot + ClariTE

Leroy et al, 2012, FPGG

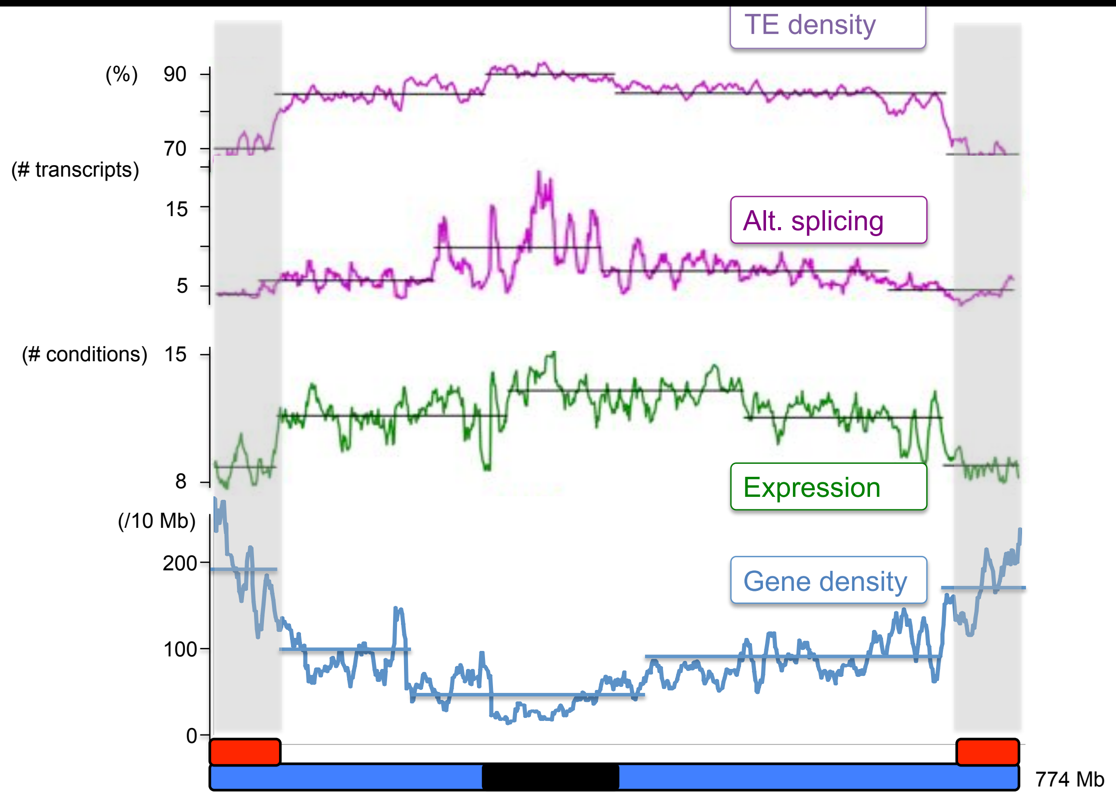
No. genes = 7,264 (~17Mb; < 2%): 5326 full genes, 1938 pseudogenes
No. NTR = 3,693
No. nc RNA = 791
TEs = 234,606 (89%)

Chromosome structure



Li et al. TPJ 2013

Chromosome structure

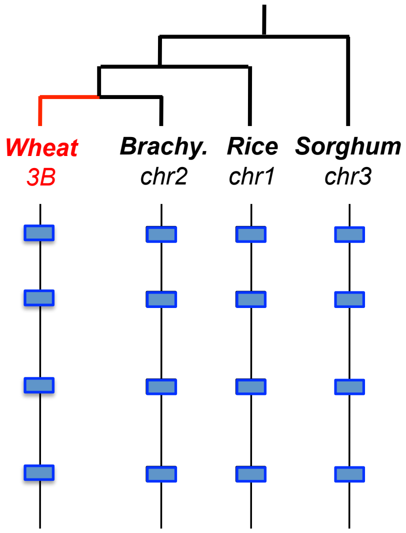


Chromosome evolution

- Gene loss?

Conserved genes Bd-Os-Sb:

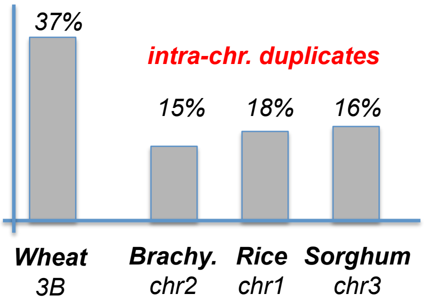
- 91% found on 3B

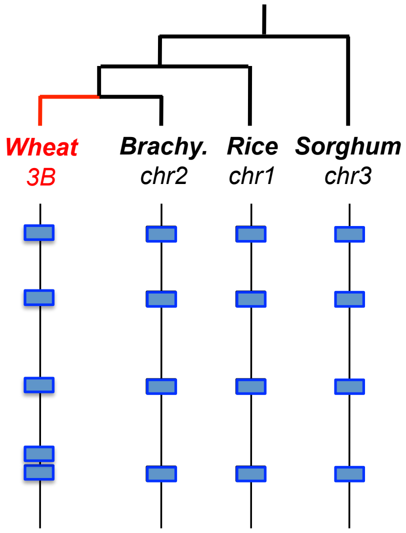


➔ Limited gene loss following polyploidization

Chromosome evolution

- Intra-chromosomal duplications

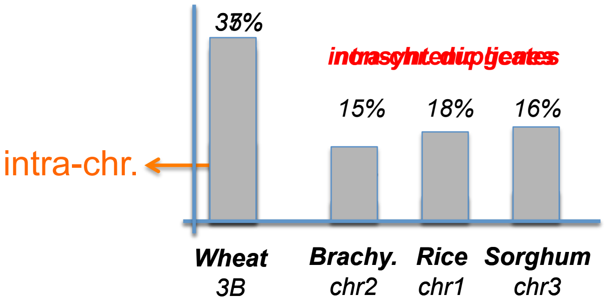


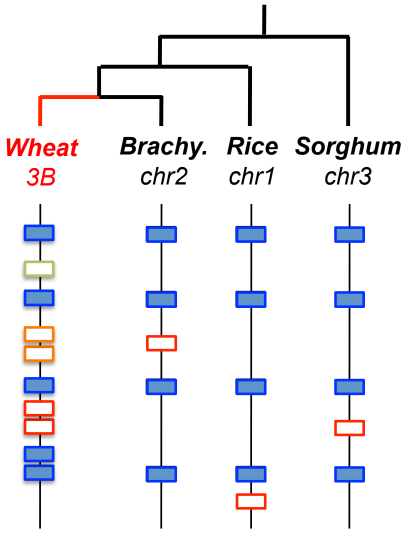


➔ Higher rate of lineage specific intra-chr. gene duplication

Chromosome evolution

- ○ Intra-chromosomal duplications
- ○ Inter-chromosomal duplications





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

➔ Higher rate of lineage specific inter-chr. gene duplication

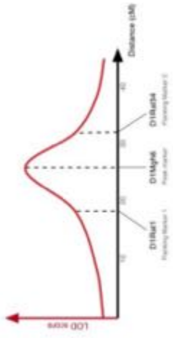
Map-based cloning projects on 3B

40 genes associated with QTL mapped on 3B

→4722 markers on 3B consensus map, 3102 in 964 SC (679 Mb =82 % of the sequence)

→13 map-based cloning projects

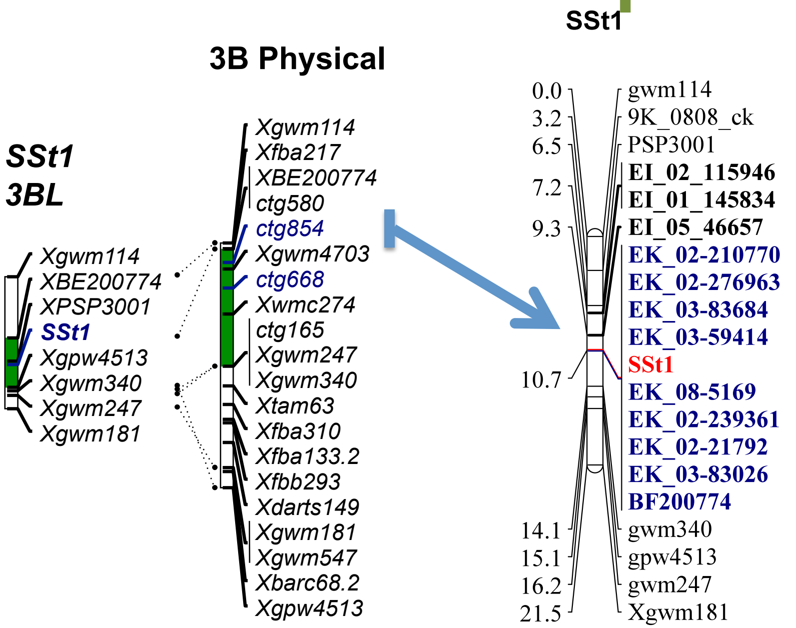
- ➢ Disease resistance genes (Sr, Lr, Yr, Stb…)
- ➢ Solid stem (saw fly)
- ➢ Yield
- ➢ Drought tolerance
- ➢ Boron transporter
- ➢ Flowering time
- ➢ NUE
- ➢ Chromosome pairing…



UNIVERSITY OF SASKATCHEWAN

Curtis Pozniak

A Breeder's Perspective



1 year

-1 EST marker, not closer
- ascertainment bias

3 weeks

-12 cosegregating markers (SNPs)
- no phenotyping in the fields
- No ascertainment bias

Wheat@URGI

Xwmc430

Examples: Xwmc430 QTL Tae. 1272327

WHAT'S NEW ?

24 Mar 2014
3B, survey seq POPSEQ & GenomeZipper data available
• 3B data (Genomic DNA, CDSs, annotation of features and a ...

18 Mar 2014
Public BLAST available
Public BLAST to query survey sequence and 454 assemblies is

04 Mar 2014
New RNA-Seq data
New oriented RNA-Seq library is publicly available.
...

14 Feb 2014
GnpIS version 13.3 is now available
GnpIS version 13.3 is now available. Major new functionalities and improvements are:
○ New ...

14 Feb 2014
Diploid wheat species assemblies are publicly available
Diploid wheat species WGS assemblies are publicly available for download.
...

09 Jan 2014
Survey sequence and Gene models are publicly available
Survey sequence assemblies and Gene models download are publicly available
...

EVENTS & PUBLICATIONS

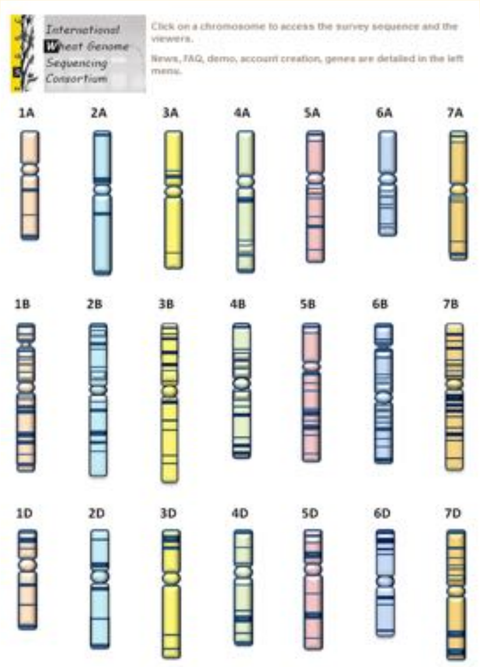
16 Dec 2013
Talks at PAG 2014
Computer demo 1: Saturday 11 at 1.50 PM Wheat 3B Annotation ...

07 Sep 2013
IWGSC Sequence Repository : new data and browsers ...
COM (communication)
International Wheat Genetics Symposium (IWGS) 12th 2013, Yokohama, Japan

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

Michael Alaux

Sequence Repository



International Wheat Genome Sequencing Consortium

Click on a chromosome to access the survey sequence and the viewers.

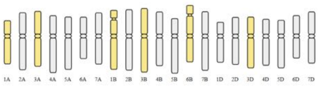
News, FAQ, demo, account creation, genes are detailed in the left menu.

Click on a chromosome to access the survey sequence chromosome arm assembly for:

BLAST search (Public BLAST is available)

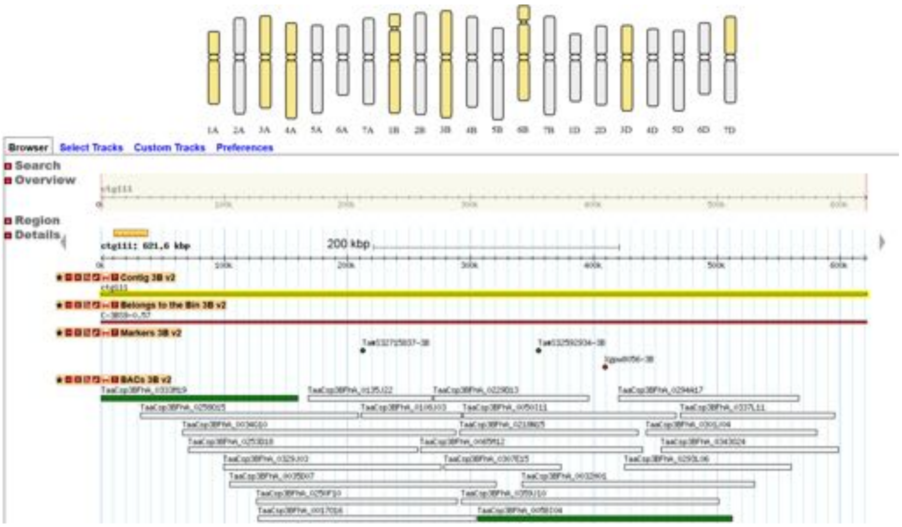
Download (Public)

Viewers:

- ❑ Physical maps 
- ❑ Annotated reference sequence 3B (Public)

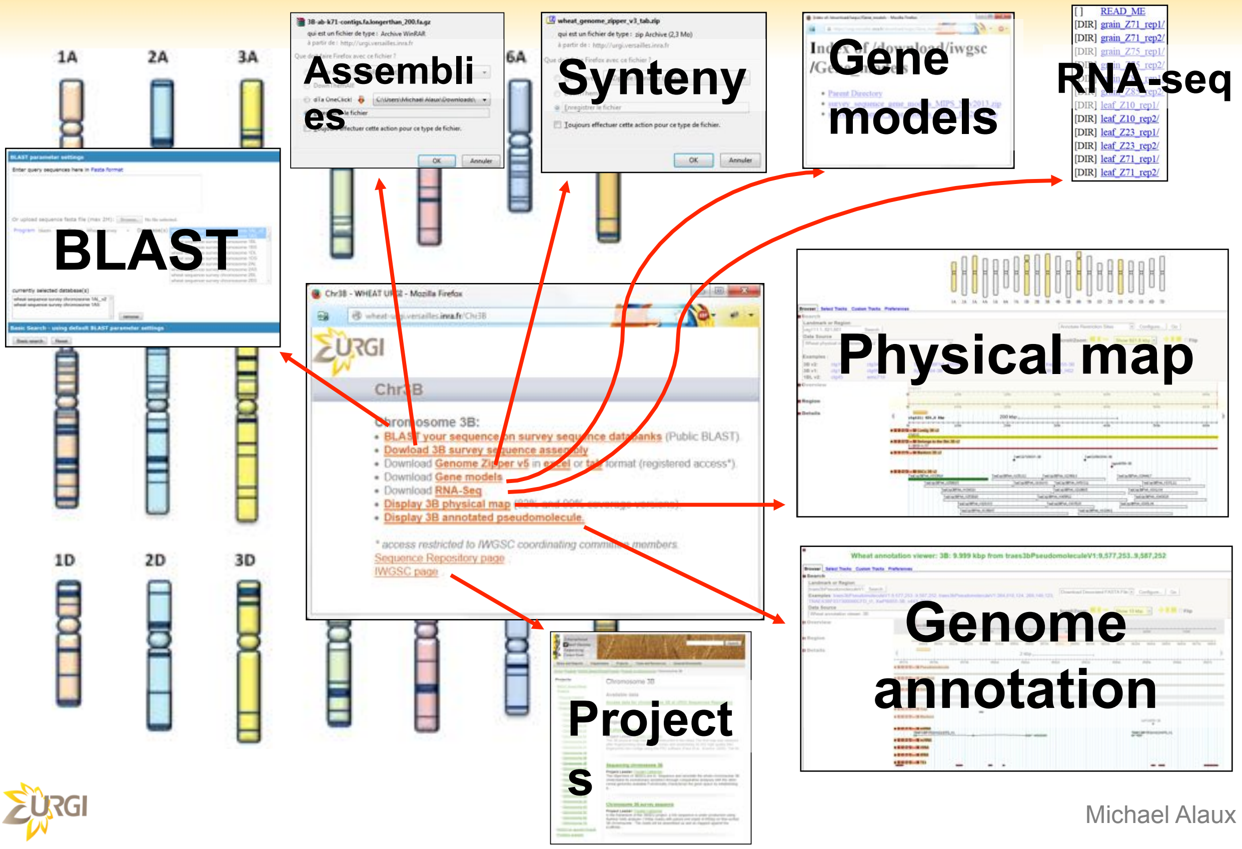
http://wheat-urgi.versailles.inra.fr/Seq-Repository

Sequence Repository: Physical maps



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

Sequence Repository: ex. 3B



Assemblies

Synteny

Gene models

RNA-seq

BLAST

Physical map

Genome annotation

Projects

Chr3B - WHEAT URGI - Mozilla Firefox

Chr3B

Chromosome 3B:
- BLAST your sequence on survey sequence databanks (Public BLAST)
- Dowload 3B survey sequence assembly
- Download Genome Zipper v5 in excel or tab format (registered access*).
- Download Gene models
- Download RNA-Seq
- Display 3B physical map (82% and 90% coverage versions).
- Display 3B annotated pseudomolecule.

* access restricted to IWGSC coordinating committee members.

Sequence Repository page

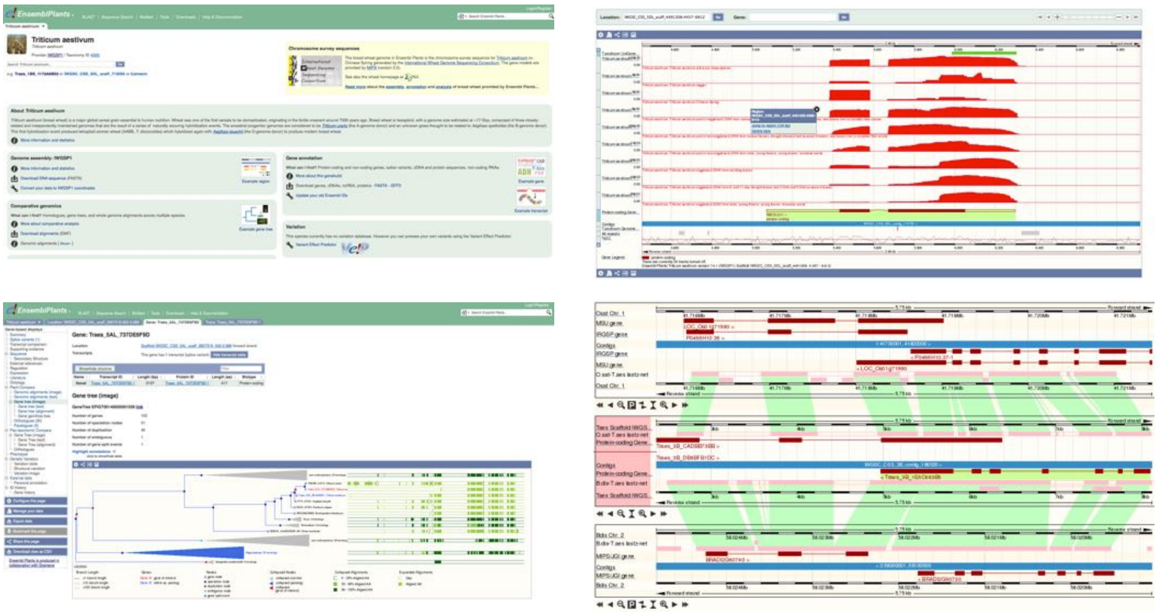
IWGSC page

URGI

Michael Alaux

EnsemblPlants

http://plants.ensembl.org/Triticum_aestivum



Interested in wheat data standards?

RDA
RESEARCH DATA ALLIANCE
Research Data Sharing without barriers

Wheat Initiative

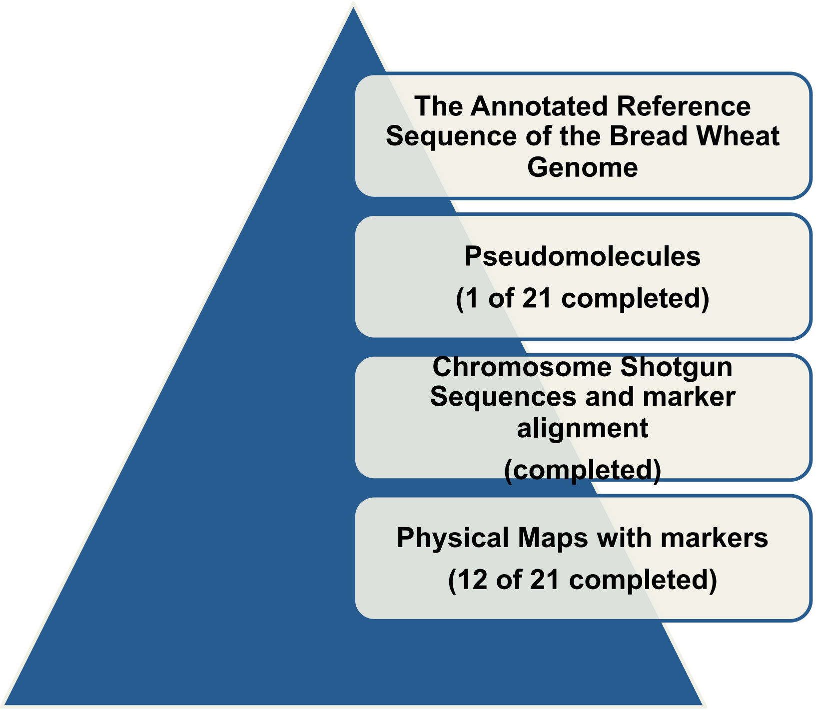
Data standards in the Wheat research community

The Wheat Data Interoperability working group has been created within the context of the Research Data Alliance (RDA) to support the Wheat Information System (WheatIS) Expert Working Group of the Wheat Initiative.

They are inviting views from the community on the formats and types of data that are required and the standards that should be applied to ensure efficient linking and interoperability of different wheat databases. Fill in the survey before 6 May at:

https://t.co/8jMObdQXCS

IWGSC Projects



The Annotated Reference Sequence of the Bread Wheat Genome

Pseudomolecules
(1 of 21 completed)

Chromosome Shotgun Sequences and marker alignment
(completed)

Physical Maps with markers
(12 of 21 completed)

Thank you for your attention!

www.wheatgenome.org

@wheatgenome

Acknowledgments –Chromosome 3B

GDEC

Génétique Diversité Ecophysiologie des Céréales

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

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

Acknowledgements – Chromosome Survey Sequence

Jaroslav Dolezel & the IEB team

HelmholtzZentrum münchen
German Research Center for Environmental Health

Klaus Mayer & team
Matthias Pfeiffer
Manuel Spannagl
Mihaela Martis

UNIVERSITY OF SASKATCHEWAN

Curtis Pozniak

URGI

Michael Alaux

IWGSC

Kellye Eversole

TGAC The Genome Analysis Centre™

Mario Caccamo
Sarah Ayling
Jon Wright
Ricardo Ramirez-Gonzalez

UMR 1095 GDEC INRA

Catherine Feuillet & the 3BSeq Group

Hirokazu Handa
David Edwards
Burkhard Steuernagel
Jesse Poland
Eduard Akhunov
Luigi Cattivelli
Simen Sandve
Odd-Arne Olsen
Andrew Sharpe
& Many More....

IWGSC Sponsors

Arcadia, ACPFG, biogemma, Bayer CropScience, CSIRO, DOW, Dow AgroSciences, CIMMYT, Céréales Vallée, DUPONT The miracles of science, evogene, Florimond Desprez, GRDC Grains Research & Development Corporation, Graminor, Heartland Plant Innovations, INRA, Lochow, KWS, Kansas Wheat, Limagrain, Monsanto, RAGT, syngenta, IWGSC