Excerpts from a Seminar at Bayer CropScience February 2015

Kellye Eversole

IWGSC Executive Director

Seminar
Ghent, Belgium
12 February 2015

International
Wheat Genome
Sequencing
Consortium

Bayer CropScience

Update on the International Wheat Genome Sequencing Consortium (IWGSC): Strategies and Resources

Kellye Eversole

IWGSC Executive Director

Seminar
Ghent, Belgium
12 February 2015

International Wheat Genome Sequencing Consortium

Bayer CropScience

IWGSC

international wheat genome sequencing consortium

10

2005 - 2015

IWGSC

2005

KANSAS WHEAT

Rediscover Wheat

K·STATE

INRA

2 sponsors
3 institutes

5 members
3 countries

The IWGSC, Inc.

22 sponsors
1,481 members and data users
More than 270 institutes
64 countries

2015

Board of Directors

Role: Overall strategy and organization

4 Permanent Directors: Appels, Eversole, Feuillet, & Keller

5-10 Non-Permanent Directors (including at least 1 Outside Director)

Leadership Team

Role: Daily management

Members: Appels, Eversole, Feuillet, Keller, & Rogers

Coordinating Committee

Role: Scientific strategy

Members: Sponsors & Leaders of IWGSC Projects & Initiatives

Members

Role: Participation & Input

Members: Open to Anyone

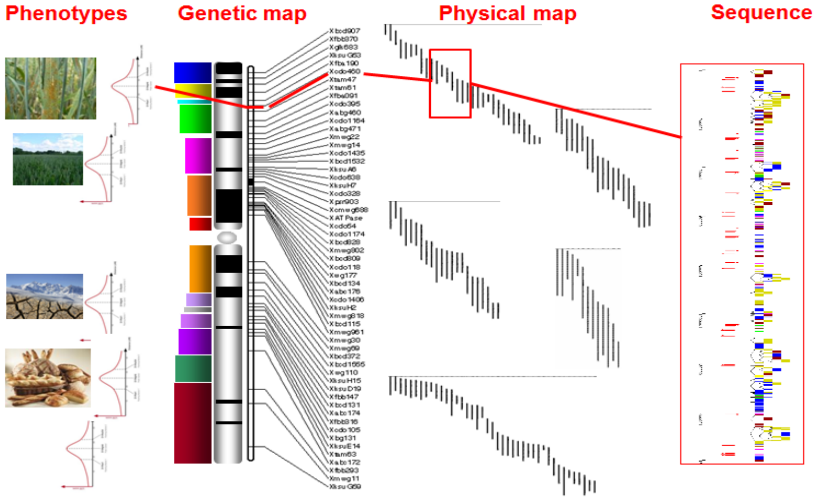
Vision

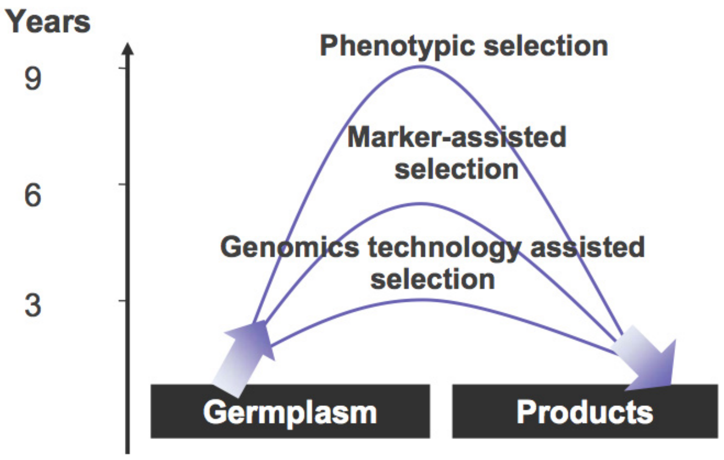
Goal

- 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





I W G S C

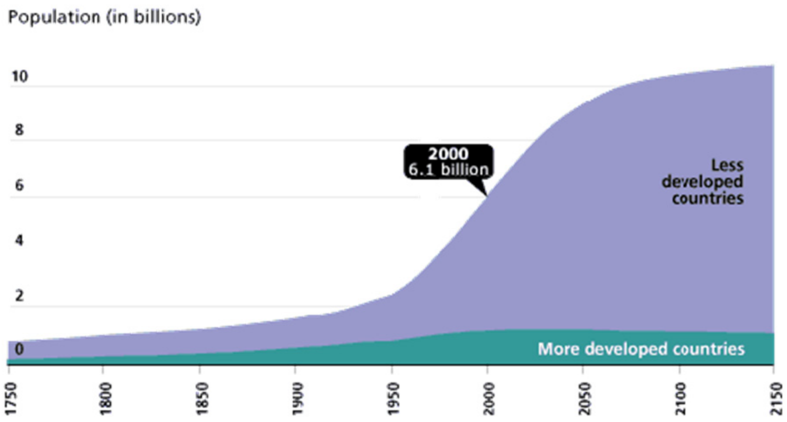
Challenges for the world's food crops

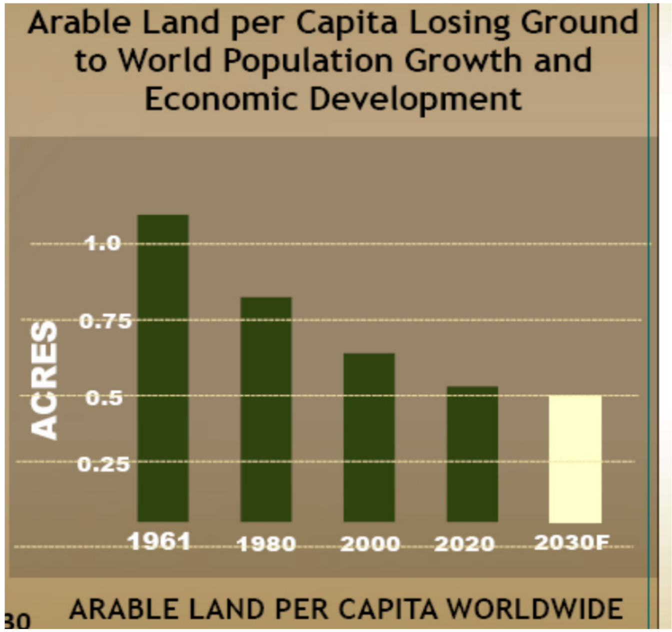
Sustainably provide sufficient quantity of high quality food

The Economist

Dealing with America's deficit
Obama's timid trip to Asia
Remote control for your car
Peter Drucker, still king of the gurus
The scientist who saw Nessie

How to feed the world





- **Decreasing annual yield growths and agricultural productivity**
- **More frequent periods of extreme weather patterns**
- **A general increase in global temperatures**

Crop improvement is complex

- ➢ **Yield potential and yield stability**
  - ✓ Photosynthesis efficiency
  - ✓ Harvest Index
  - ✓ Reduced inputs (fertilizers, pesticides..)

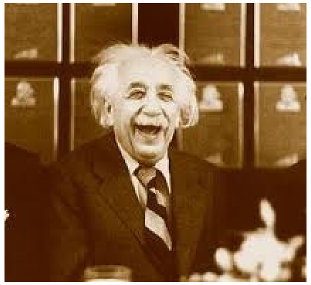
- ➢ **Adaptation to climate change**
  - ✓ Avoidance
  - ✓ Tolerance (Drought, heat, cold..)
  - ✓ Post stress recovery…

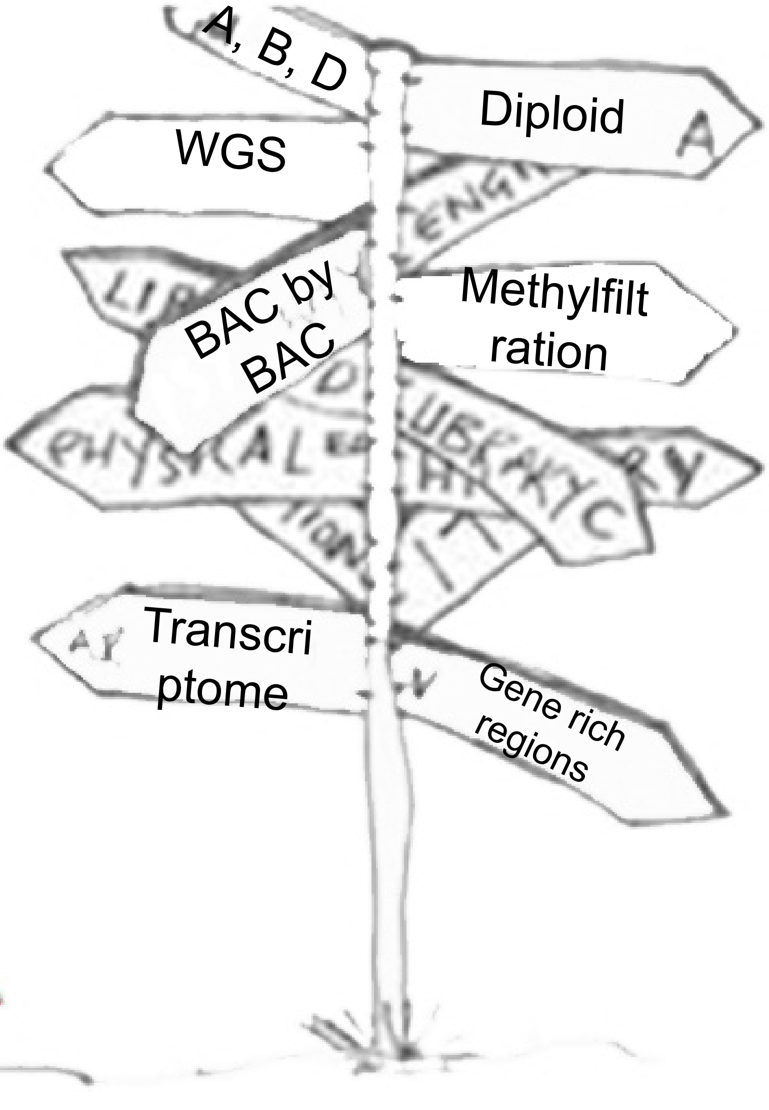
- ➢ **Durable resistance to biotic stress**
  - ✓ Usual suspects (virus, fungi)
  - ✓ New pests and diseases
  - ✓ Invasive species

- ➢ **Quality of grain and co-products**
  - ✓ Grain protein content
  - ✓ Starch, straw
  - ✓ Food safety - Allergenicity-mycotoxin contamination….

How to produce a useful sequence?

IWGSC





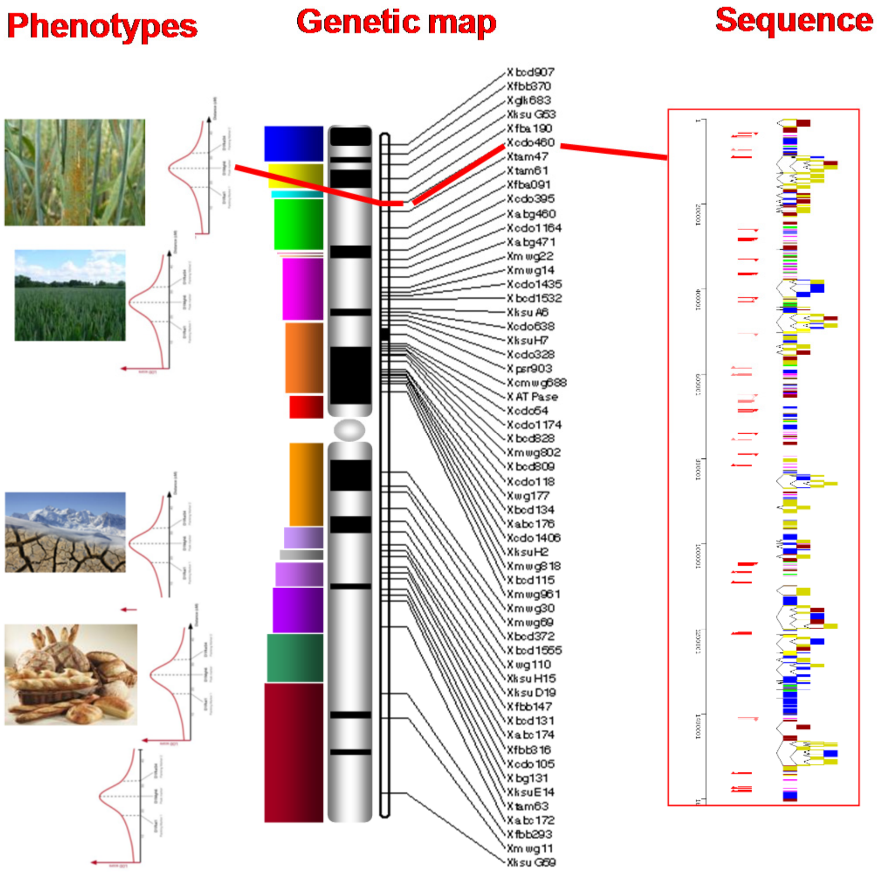
Key Considerations

- A genome sequence is only a tool
- What kind of sequence?
- For whom and for what purposes?
  - Comparative genomics
  - Markers
  - Gene cloning
  - Development of new varieties
  - Long-term breeding pipelines
- Goal - Optimize the cost/quality ratio to deliver a useful tool

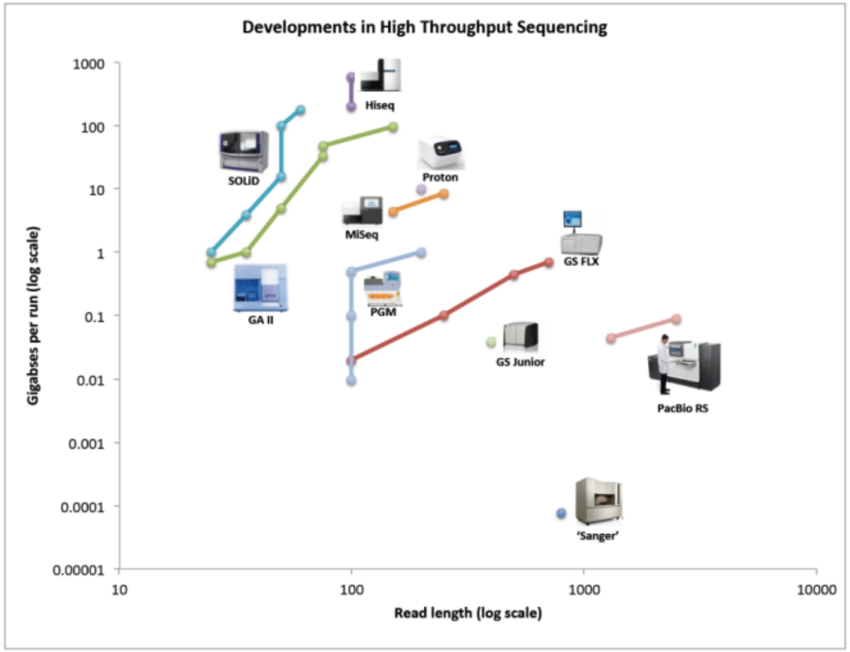
How to produce a useful sequence?

1-Sequence what grows in the fields! -> bread wheat

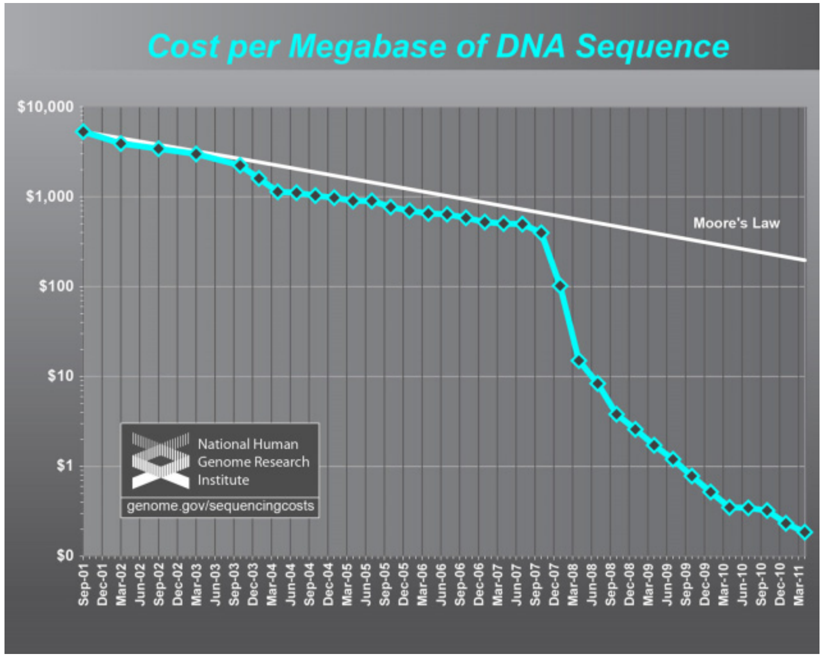
2- A sequence linked to the genetic maps and phenotypes



Faster and cheaper methods to tackle the wheat genome



http://flxlexblog.wordpress.com/2012/12/03/developments-in-next-generation-sequencing-a-visualisation/



Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI www.genome.gov/sequencingcosts.

Which Way for Wheat…

We offer three kinds of service:

GOOD - CHEAP - FAST

You can pick any two

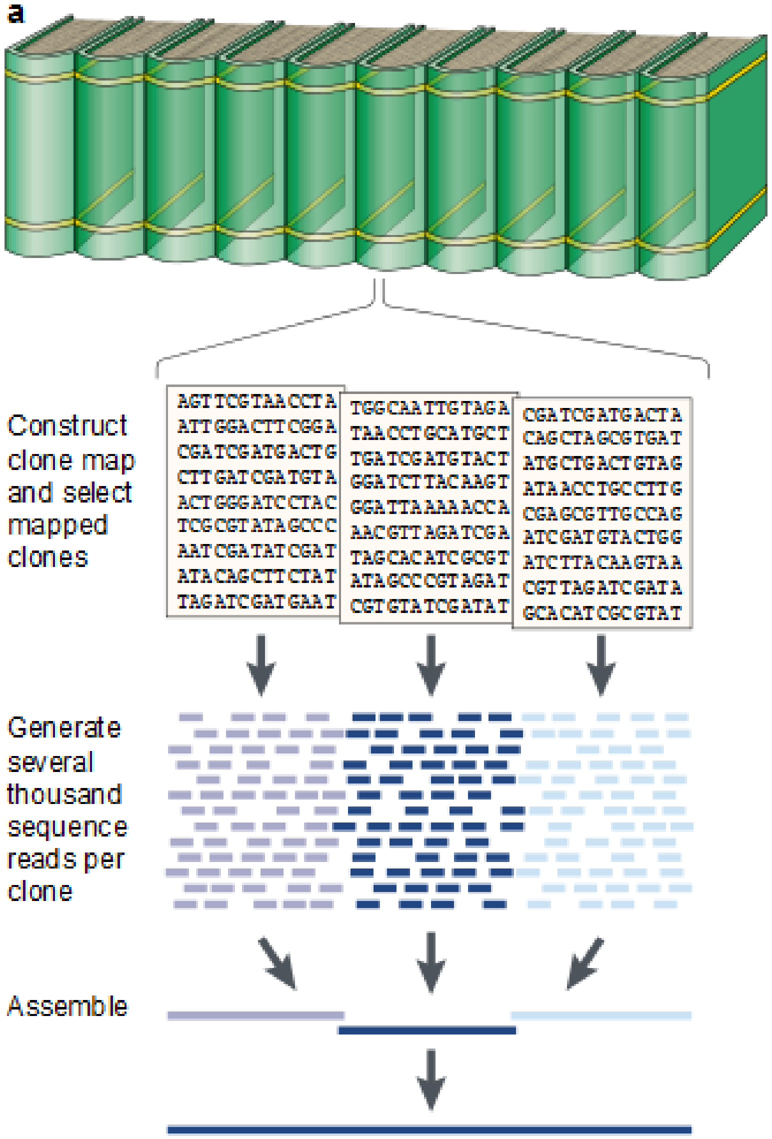
GOOD service CHEAP won't be FAST

GOOD service FAST won't be CHEAP

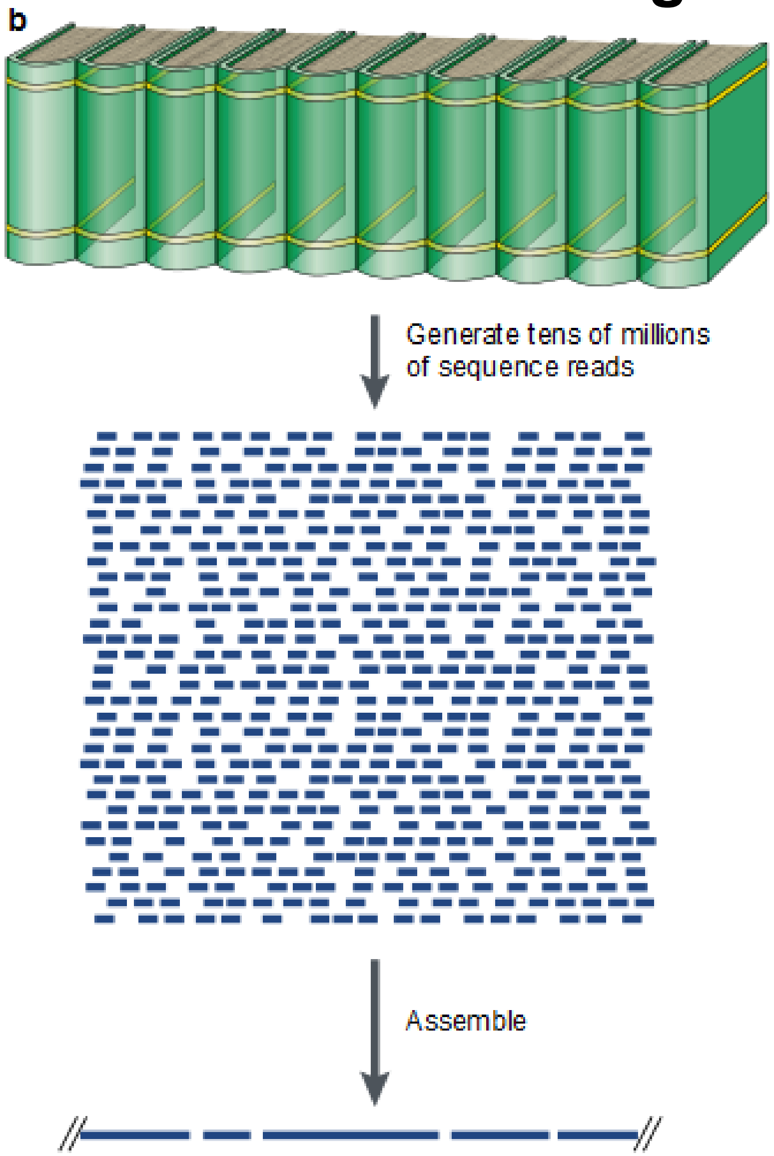
FAST service CHEAP won't be GOOD

Two possible approaches

BAC by BAC

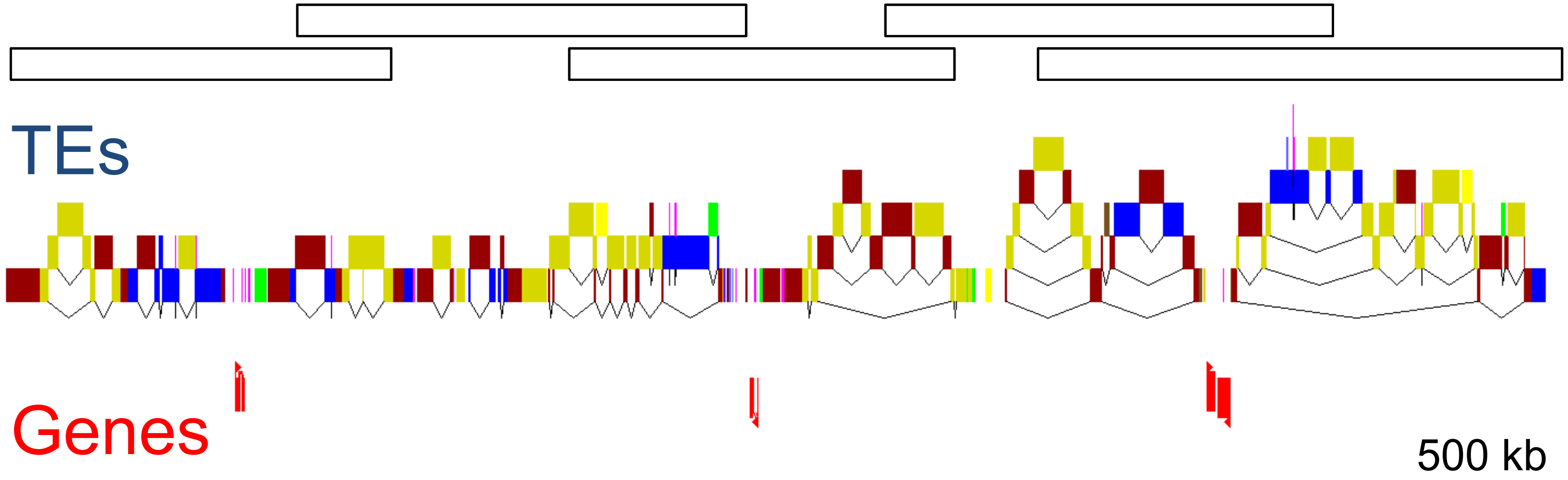


Whole Genome Shotgun

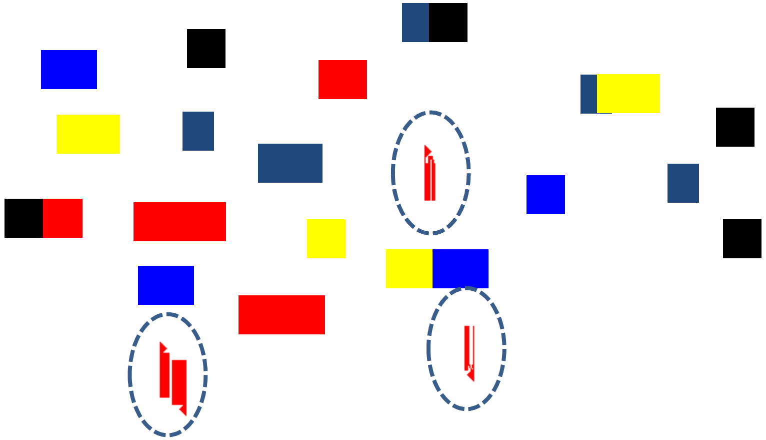


BAC by BAC vs WGS

BAC by BAC (physical map) √



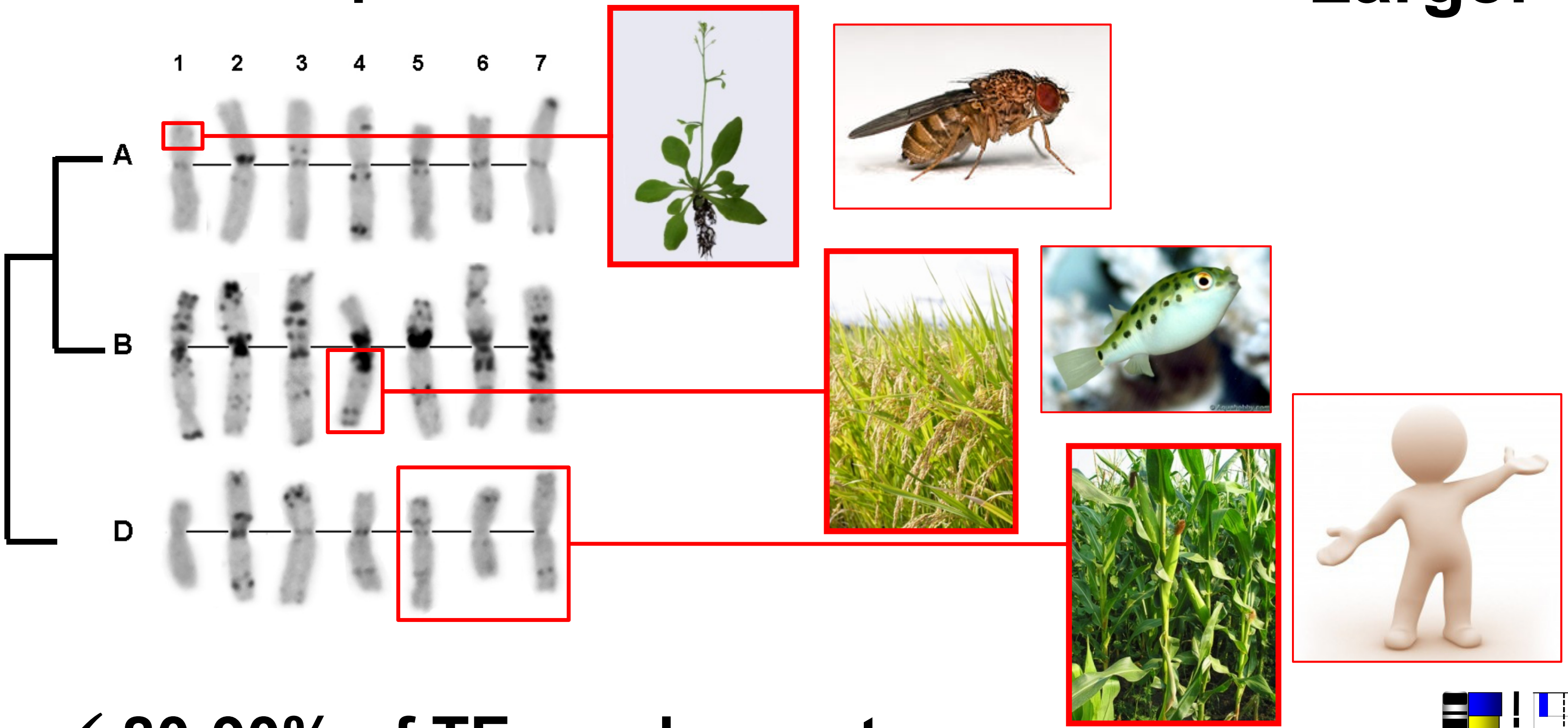
Whole genome shotgun



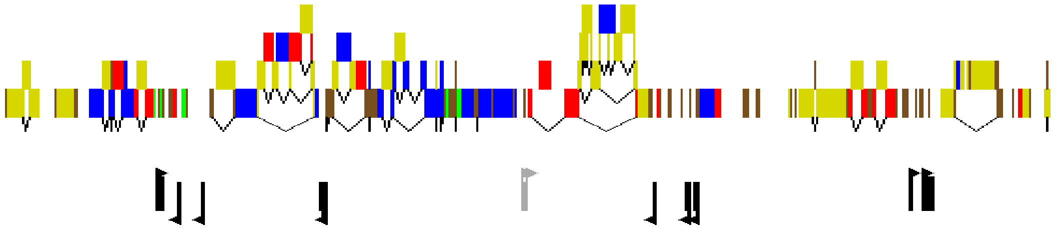
The bread wheat genome is……a challenge

✓Allohexaploid

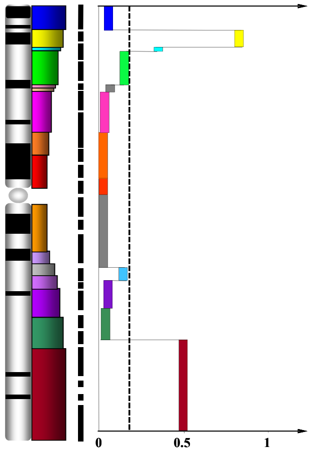
✓ Large: 17 Gb



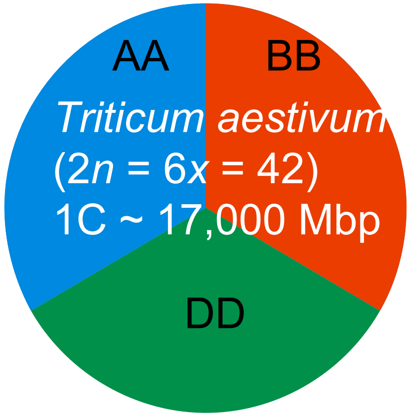
✓ 80-90% of TEs and repeats

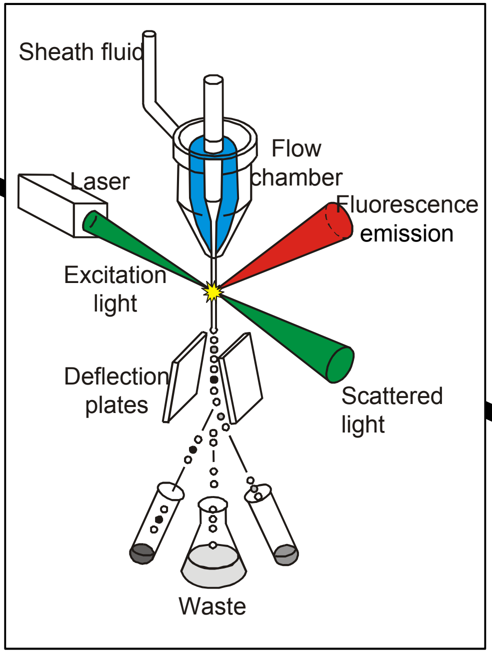


✓ > 50% of non recombinogenic regions



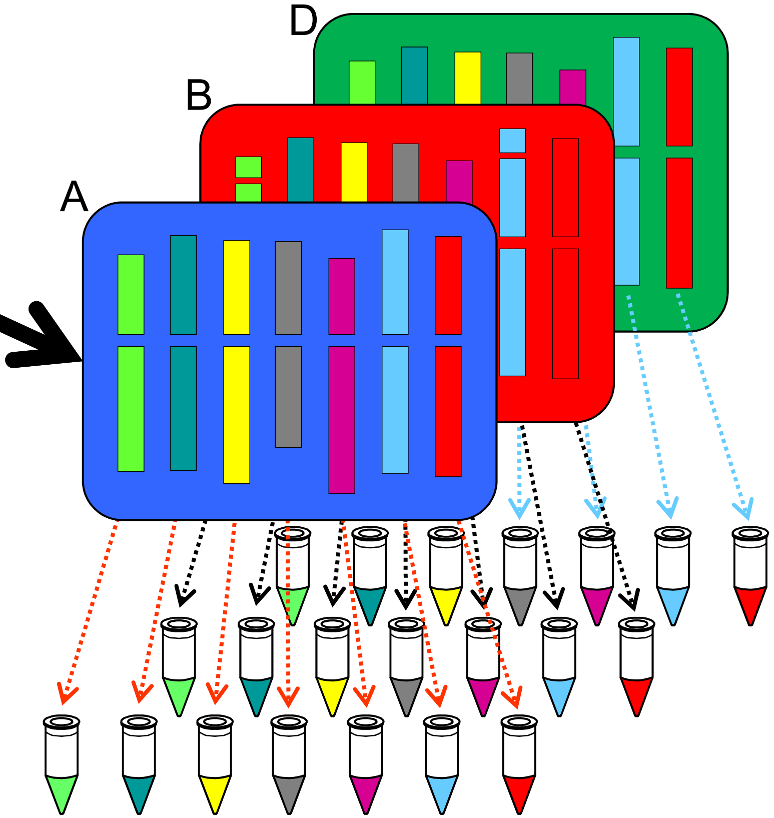
Managing the 17 Gb, Hexaploid Genome





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

Dissection of the genome into single chromosomes (arms)



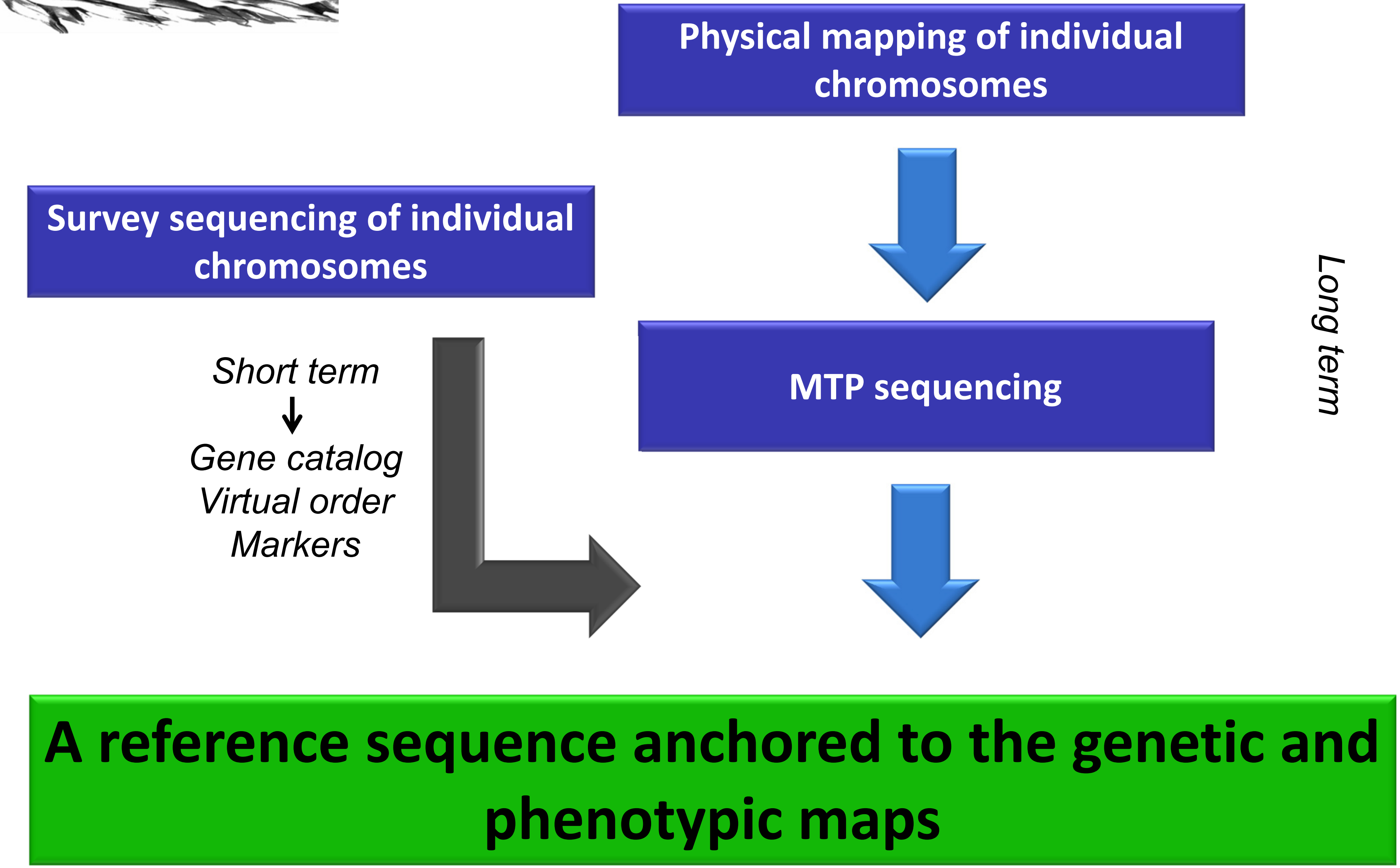
- **Chromosomes: 605 - 995 Mbp (3.6 – 5.9% of the genome)**
- **Chromosome arms: 225 - 585 Mbp (1.3 – 3.4% of the genome)**

Chromosome genomics

- **Chromosome specific BAC libraries and sequencing**

IEB

Roadmap to the Wheat Genome Sequence



Physical mapping of individual chromosomes

Survey sequencing of individual chromosomes

MTP sequencing

Long term

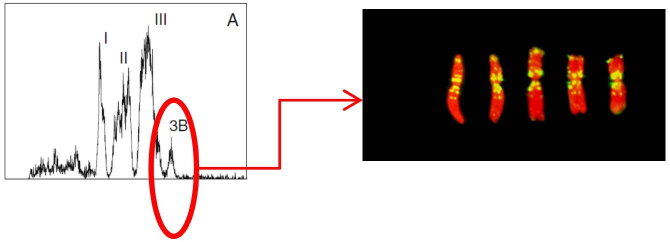
Short term

↓

Gene catalog
Virtual order
Markers

A reference sequence anchored to the genetic and phenotypic maps

IWGSC Strategy for a Reference Wheat Genome Sequence

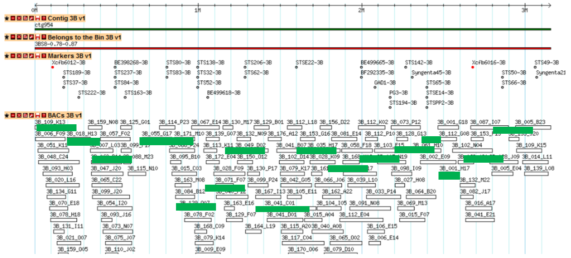


1. **BAC library construction**

2. **BAC fingerprinting (HICF/WGP)**

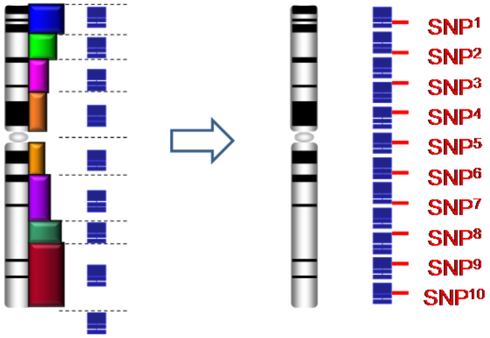
```
GAATTCTTGGTCAGCAATAAGGACAAAGAA
GAATTCCTCCTGTGCGGCCGTTTTA
GAATTCATGCTCCAGCCAACTCATCCAAAA
GAATTCAACAATATAAATATCACCTAAAGC
GAATTCCATCAGTATGGAGTTTCTTCATGC
GAATTCAAGCTTGAGTATTCTATAGTGTCA
GAATTCAATATGTAGAAATTATAGGATGGC
GAATTCACCGGAAGGGGTTCCGGGTGTTTC
GAATTCCGGAACGCTTCCGGAGACCAAACA
GAATTCTGGAATCACTTGATGATGAATCAG
GAATTCTGGAGTCTTCCTTTCCGCAGCGAA
GAATTCAACGAAATAGTTATGAAATAGAAG
GAATTCTGGGGTAACGATGGAATCCTACAC
GAATTCTGTCAACGACACTATCATCAGGAA
GAATTCTTCGATGCACTGAGTCGCACGGTT
GAATTCTTGTGTGAGGTCCGCGTCCTAGAT
```

3. **Contig assembly by FPC/LTC**



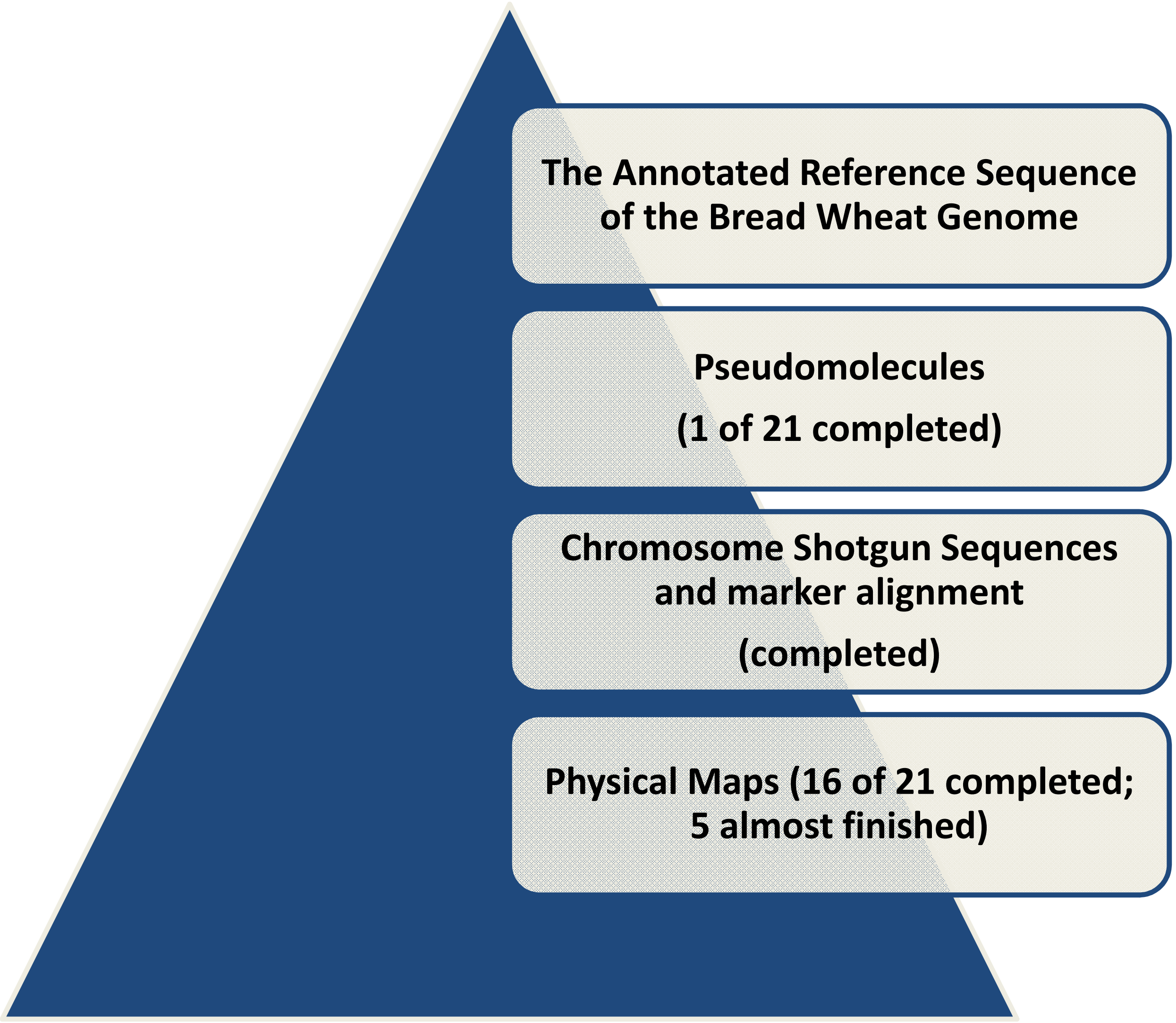
4. **MTP sequencing /Scaffold assembly**

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



6. **Automated and curated annotation**

IWGSC Projects



The Annotated Reference Sequence of the Bread Wheat Genome

Pseudomolecules

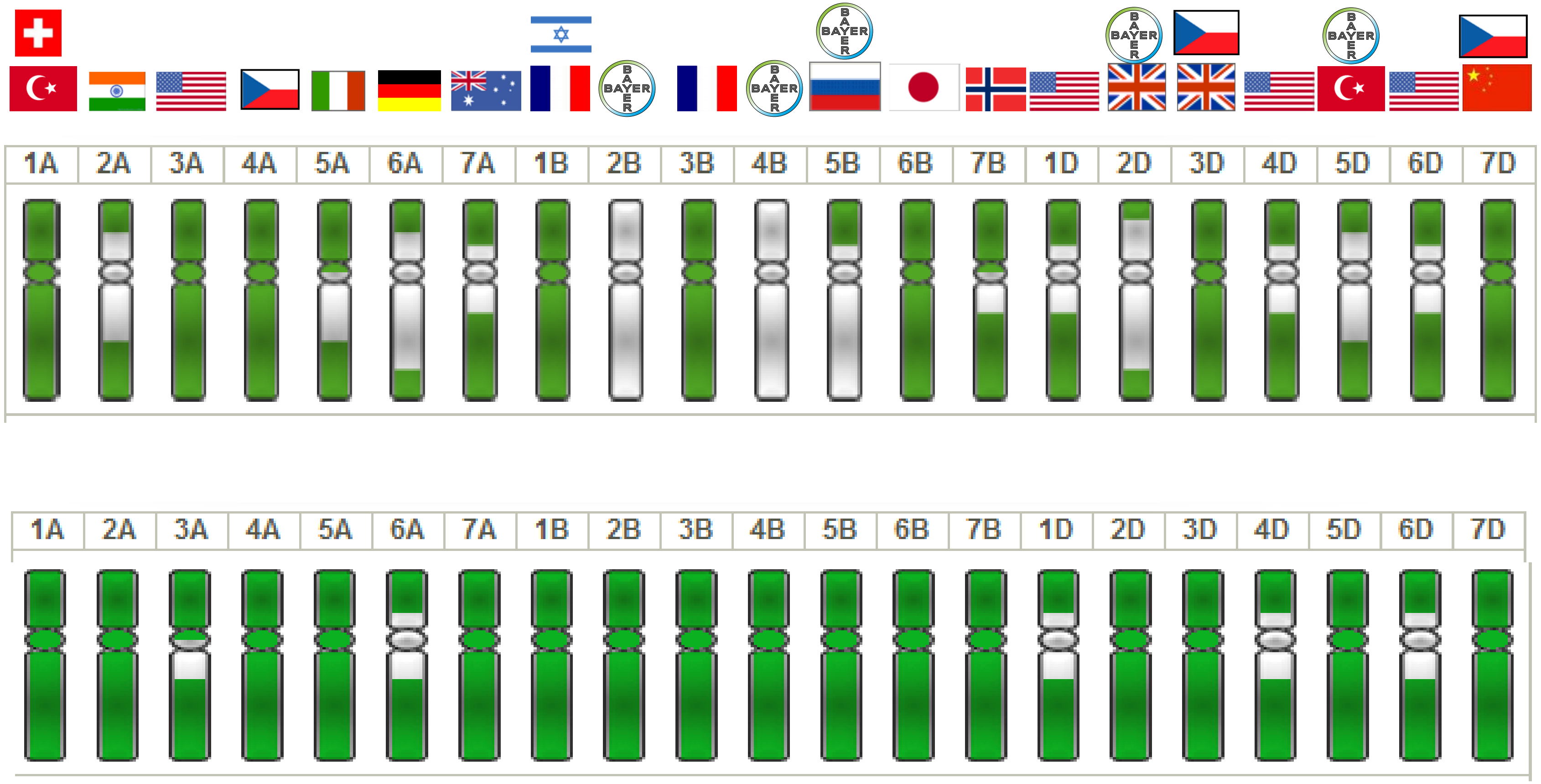
(1 of 21 completed)

Chromosome Shotgun Sequences and marker alignment

(completed)

Physical Maps (16 of 21 completed; 5 almost finished)

Wheat Chromosome Physical Maps



November 2013

January 2015

Physical maps (of varying quality) have been constructed for all chromosomes.
5 maps are being improved.

Status reports from IWGSC website

IWGSC Chromosome Shotgun Sequencing Initiative

I W G S C

TGAC The Genome Analysis Centre™

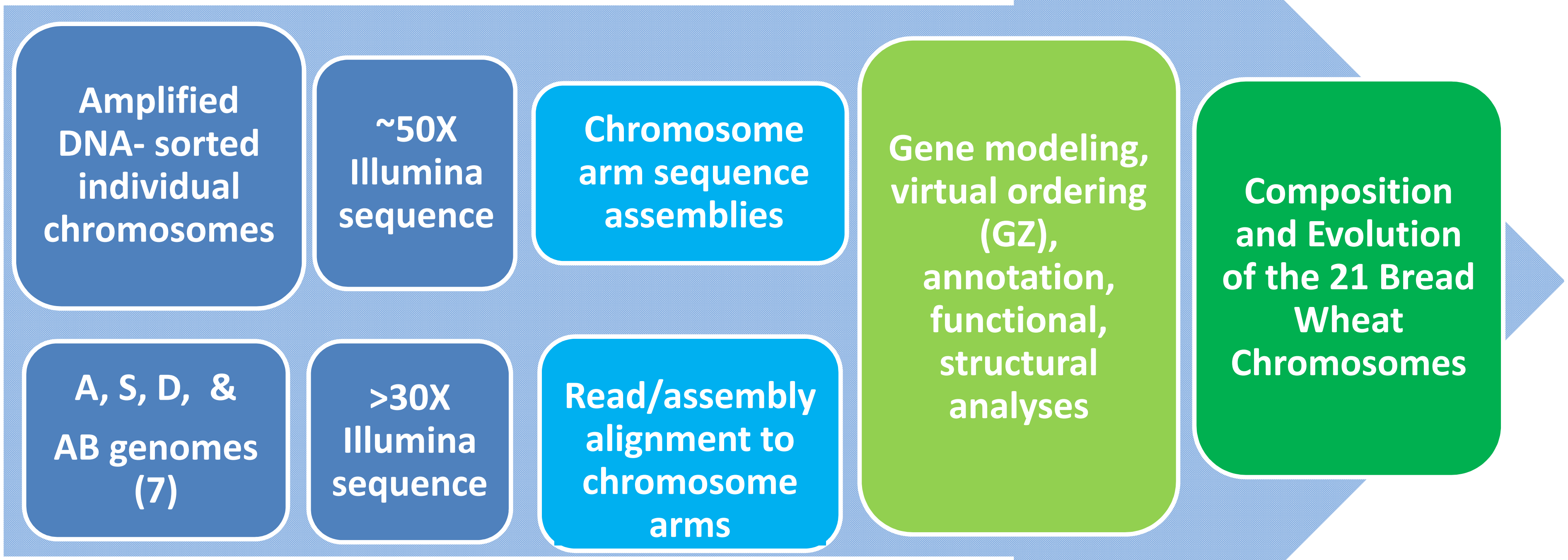
Institute of Experimental Botany of the AS CR, v. v. i.

INRA

biogemma

Graminor

HelmholtzZentrum münchen
German Research Center for Environmental Health



Amplified DNA- sorted individual chromosomes

~50X Illumina sequence

Chromosome arm sequence assemblies

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

>30X Illumina sequence

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

Agence Nationale de la Recherche ANR

University of Saskatchewan

USDA NIFA
United States Department of Agriculture
National Institute of Food and Agriculture

Australian Government
Australian Research Council

Bioplatforms Australia

ICARDA

BBSRC bioscience for the future

The Research Council of Norway

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

NIAS

Together, we are
KANSAS WHEAT

The elusive dream of regenerating the heart *p. 252*

Summer selections for your reading pleasure *p. 258*

Mesoscale eddies help drive ocean circulation *p. 322*

Science

$10
18 JULY 2014
sciencemag.org

AAAS

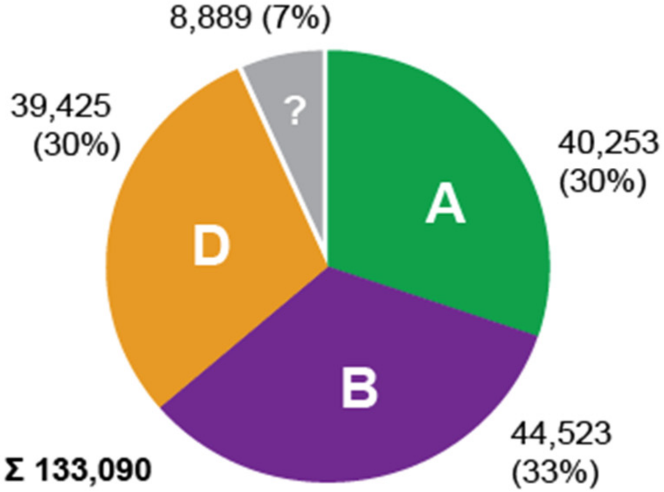
Slicing the wheat genome

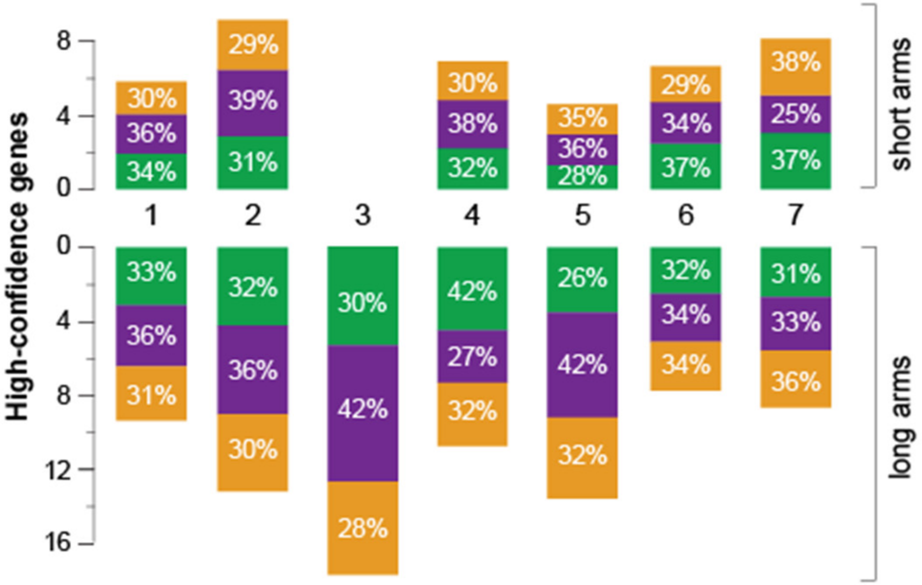
Sequencing illuminates bread wheat evolution and breeding potential *p. 285*

IWGSC Chromosome Survey Sequence Data

- **Sequence assemblies for 40 chromosome arms + chromosome 3B**
  - ❖ Total length **10.2 Gb**
  - ❖ **128Mb (1DS) – 639Mb (3B)** assembled sequence per chromosome
  - ❖ N50 contig length after repeat masking = **6.1kb** (1.7kb-8.9kb)

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

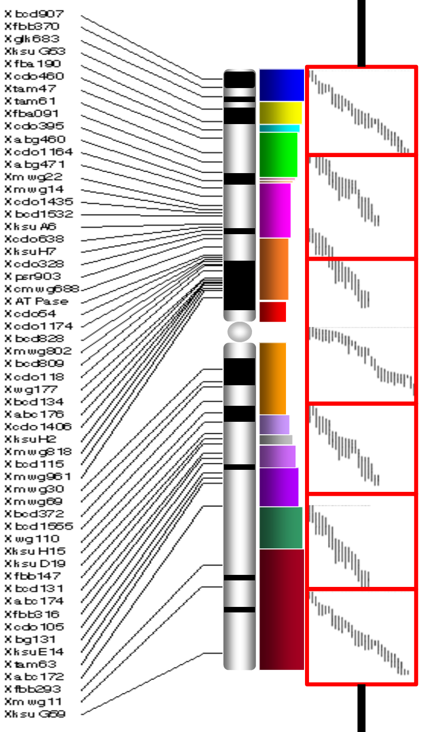




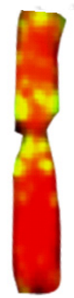
IWGSC Chromosome Survey Sequencing - Summary

- Almost full wheat gene complement identified and allocated to chromosome arms
- On average, 53% of genes virtually ordered along chromosomes
- High level of inter- and intrachromosomal duplication
- Over 3.5 M markers mapped to contigs (1.3M wheat markers + 2.3M SNPs) - SSR, EST, DArT, SNP (90k) markers...
- 13.2 million SNPs from POPSeq aligned to contigs

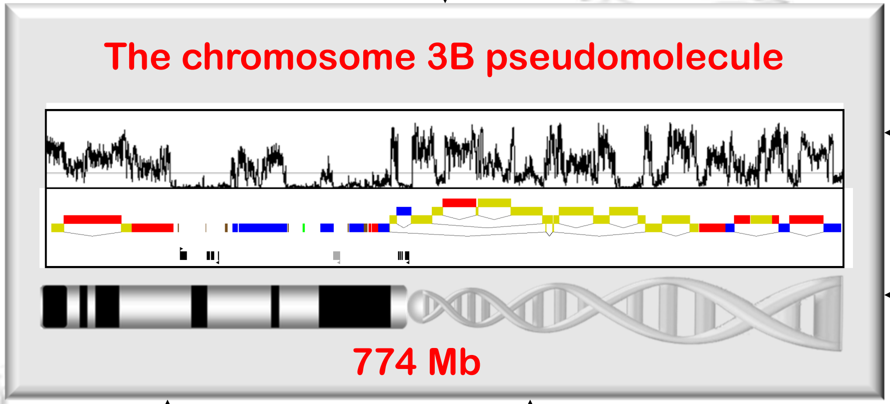
Sequencing the hexaploid wheat chromosome 3B



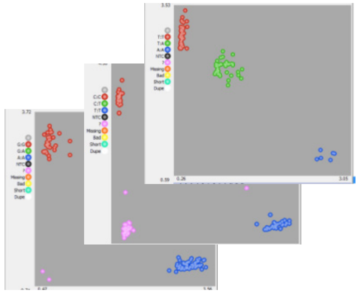
Roche GSFLX sequencing
of BAC pools



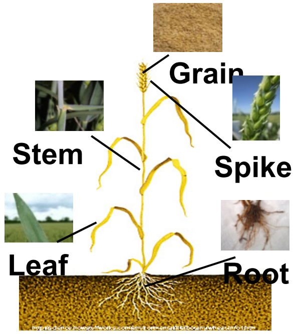
Illumina sequencing of
sorted chromosome



Sanger
sequencing
of BESs



Genotyping data
1,100 lines x 3,000 SNPs

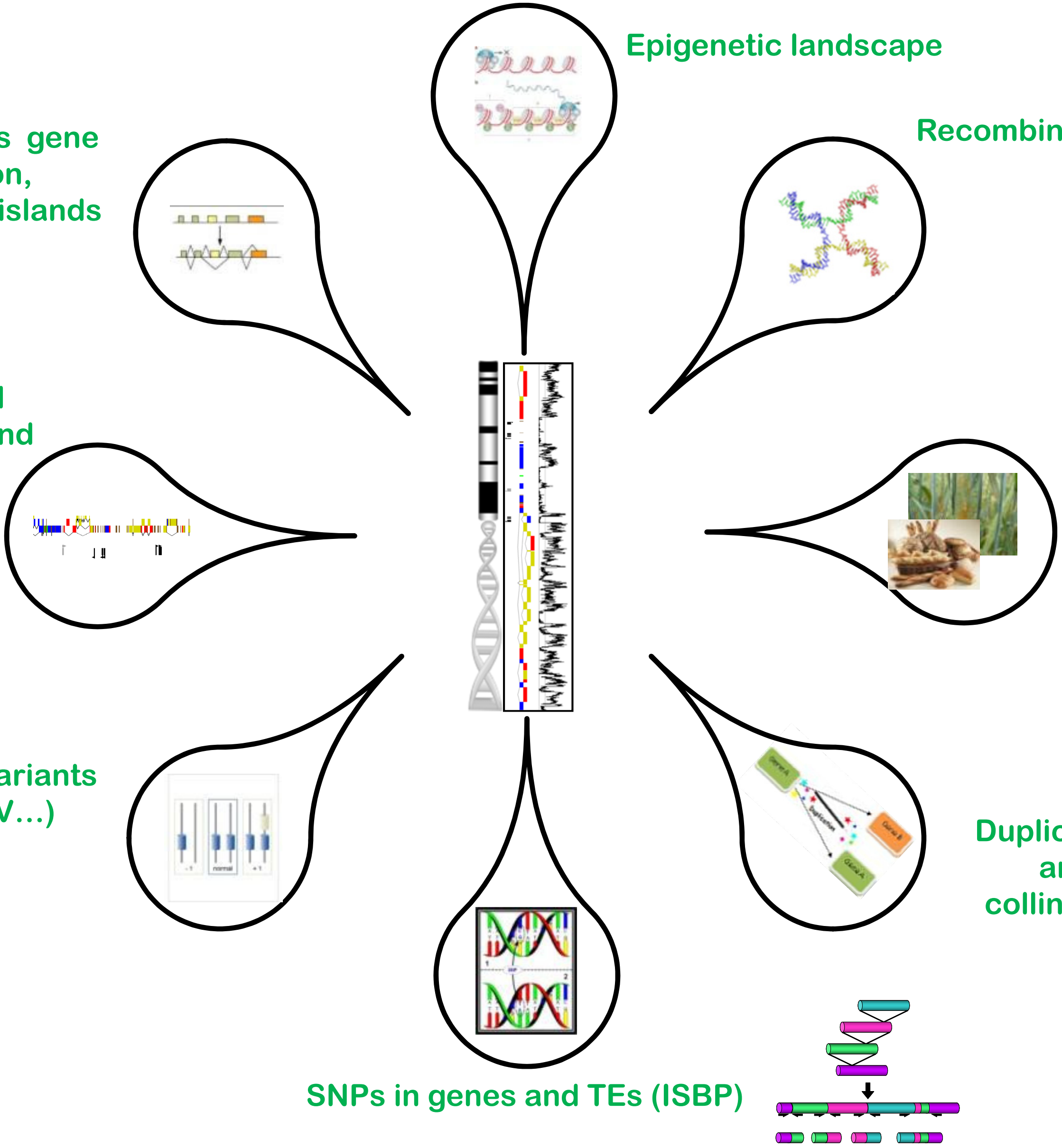


RNAseq on 15
conditions

TriAnnot

Feuillet, Choulet, & Paux

The 3B reference sequence: an enabling tool to discover, understand, and apply



Epigenetic landscape

homoeologous gene
expression,
coexpression islands

Recombination, LD analysis

TE organisation and
impact on structure and
expression

Map-based
cloning

Structural variants
(CNV, PAV…)

Duplications
and
collinearity

SNPs in genes and TEs (ISBP)

C Feuillet, 2013

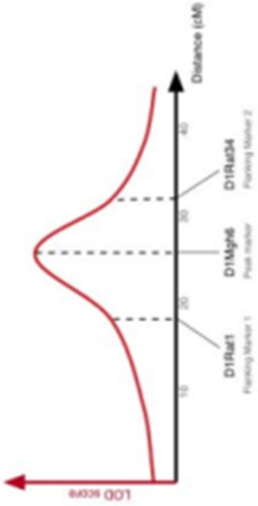
3B Reference is facilitating map-based cloning

40 genes and QTL mapped on 3B

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

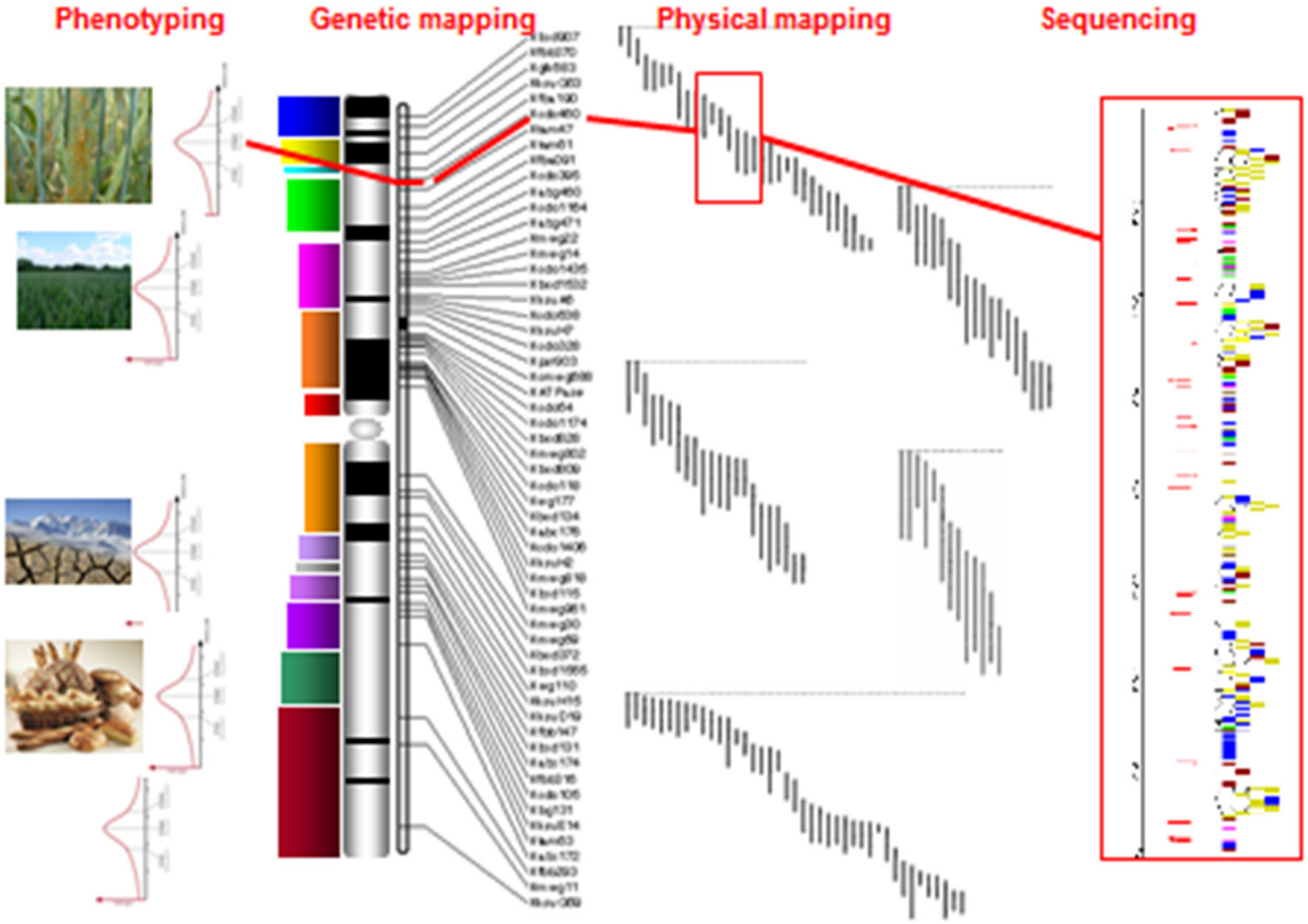
→13 map-based cloning projects

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



C Feuillet, 2013

A useful genome sequence: Wheat 3B



- Unlimited markers
- Efficient gene cloning
- Pre-breeding
- Precision breeding
- Shortens time to market
- Improves seed value capture

IWGSC Resources

- All resources accessible at URGI, Versailles

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

- IWGSC CSS Raw sequence reads in the SRA
- IWGSC CSS assemblies available at URGI and EBI
- IWGSC CSS assemblies & gene models integrated into EnsemblPlants

  http://plants.ensembl.org/Triticum_aestivum

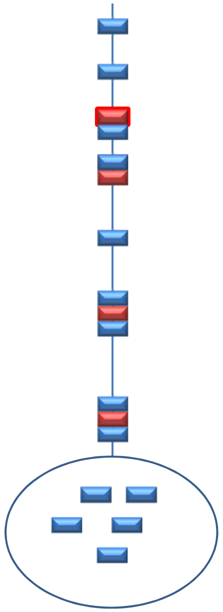
Perspectives on Sequencing Wheat ….

- 1994 to 1998:
  - It will cost $1billion to sequence maize
  - Arabidopsis can answer all questions, much more cheaply (Only $110 million to sequence)
  - We only need ESTs for economically important crops
  - If you have full length cDNAs, you will have everything you want
- 2001 to 2003:
  - Okay, so Arabidopsis is not a cereal, rice is the answer
  - Since it is not possible, we will focus money on coming up with a new solution
  - $500 million later… Eureka! We can focus on sequencing only the genes
- 2003 to 2008:
  - $30 million for sequencing maize
- Today …
  - Breeders and biologists alike agree having the genes are not enough
  - But, it's too late, you have your sequence

- 1994 to 2004:
  - Wheat ESTs and a physical map of the D-genome progenitor is all you need
- 2004 to 2008:
  - It is not possible to sequence wheat & not possible to sort chromosomes
  - It will cost $1billion to sequence wheat
  - We only need ESTs, a handful of cDNAs, and the D-genome sequence
- 2008 to 2010
  - Okay, so you can do chromosome based physical maps…. But
  - Brachypodium sequence is the answer!
  - Whole genome based on other crops is only way to go, but we need to keep spending money on in-silico approaches to figure out how to do a whole genome approach
- 2010 to today
  - You only need the genes
  - We can sequence the progenitors because we know all of them
  - 5X shotgun is good enough
  - Survey sequence is enough

Moving Towards the IWGSC Target

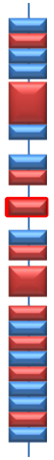
Whole Chromosome Shotgun



CS chromosome survey sequences (2014)

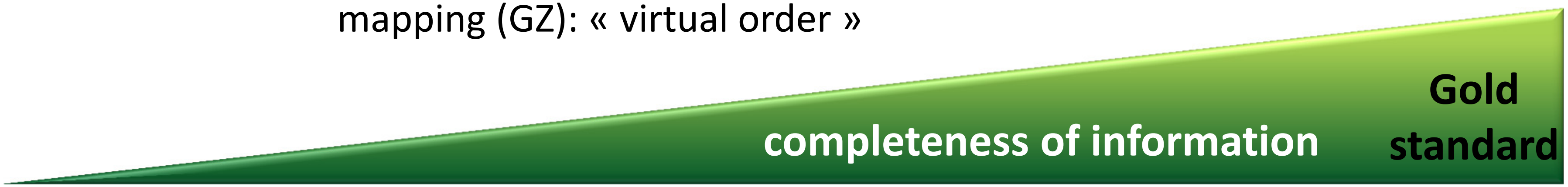
Max 53% using synteny and genetic mapping (GZ): « virtual order »

Chromosome MTP

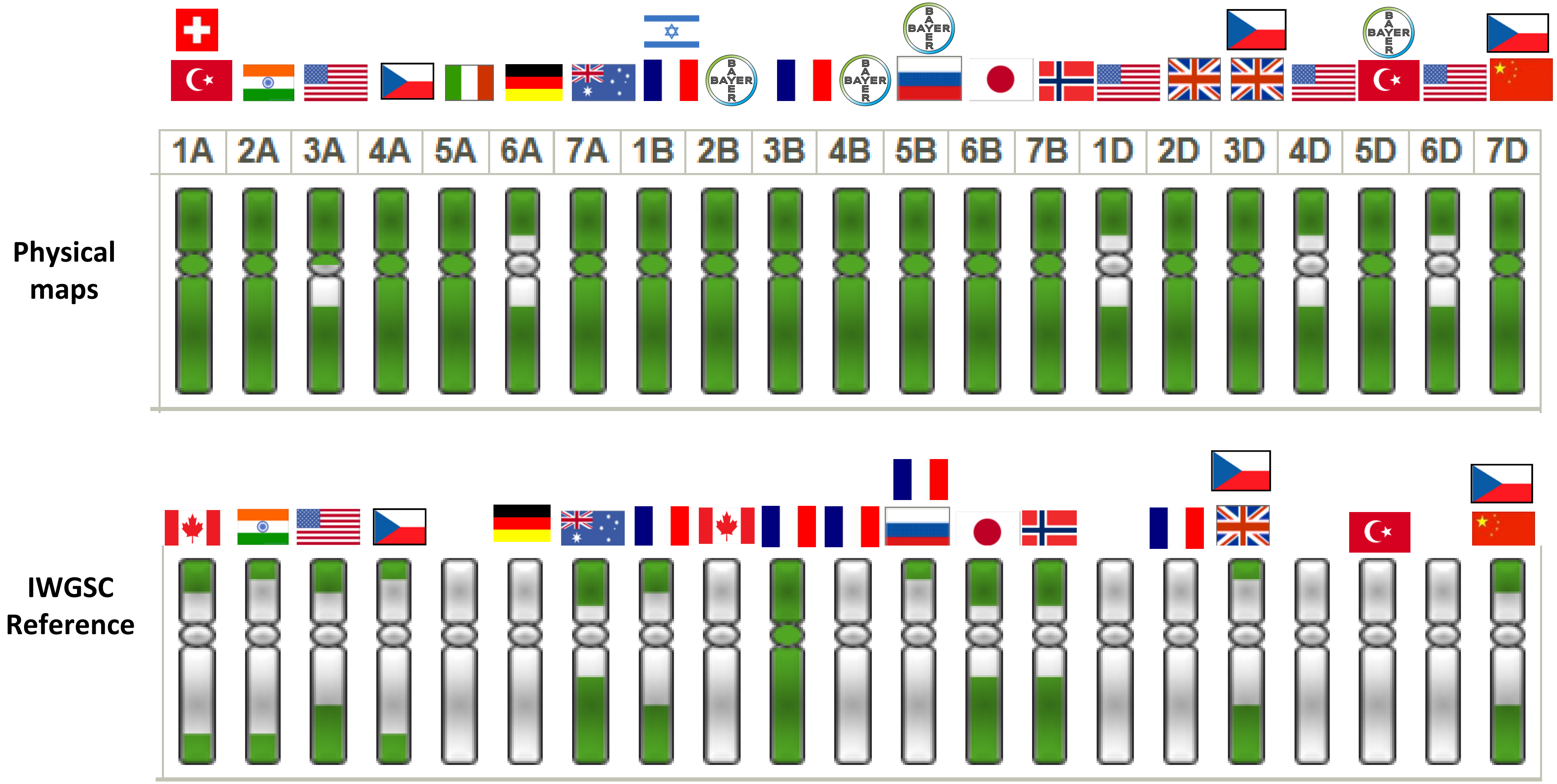


CS 3BSEQ (2014)

94% real order



Progress towards completion of Bread Wheat Projects



IWGSC

Our Sponsors

Arcadia
BIOSCIENCES

ACPFG

biogemma

BAYER
Bayer CropScience

CSIRO

Dow
Dow AgroSciences

CIMMYT
International Maize and Wheat Improvement Center

Céréales
Vallée

DUPONT
The miracles of science

GRDC
Grains Research & Development Corporation
Your GRDC working with you

evogene

FLORIMOND DESPREZ

Graminor

HEARTLAND PLANT INNOVATIONS

INRA
Institut National de la Recherche Agronomique

KWS
Seeding the future
since 1856

KANSAS WHEAT
Rediscover Wheat

Limagrain
from earth to life

MONSANTO

RAGT

syngenta

IWGSC

Acknowledgements

- IWGSC Leadership Team:
  - Rudi Appels, Murdoch University
  - Kellye Eversole, IWGSC
  - Catherine Feuillet, (INRA)-Bayer CropScience
  - Beat Keller, University of Zurich
  - Jane Rogers, IWGSC

- 63 members of the Coordinating Committee

- Physical mapping and sequencing project leaders in 19 countries, their team members, and collaborators

For More Information:
www.wheatgenome.org

Kellye Eversole, Exec. Director
(eversole@eversoleassociates.com) +1.202.352.4210

Jane Rogers, Deputy Exec. Director
(janerogersh@gmail.com)

IWGSC