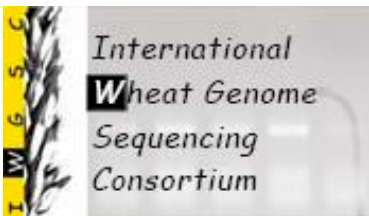


Odyssey of the IWGSC Reference Genome Sequence: 12 years 1 month 28 days 11 hours 10 minutes and 14 seconds.

Kellye Eversole
IWGSC Executive Director

Plant Genomics and Gene Editing Congress

Amsterdam, The Netherlands
16 March 2017



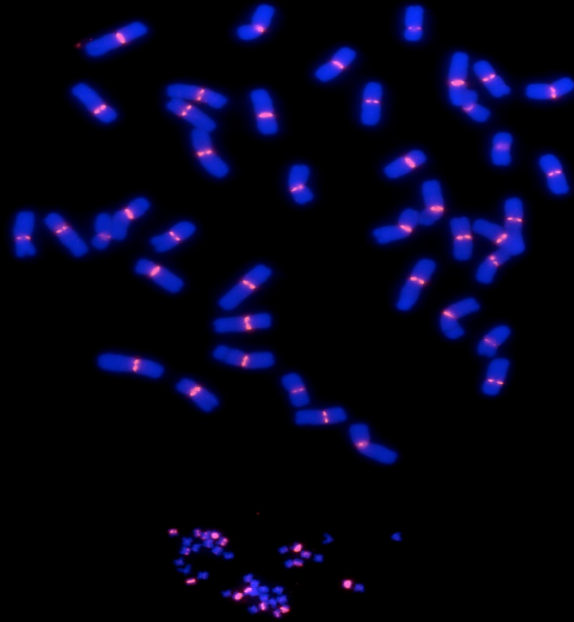
The odyssey begins... 2005



2005 - Genome sequencing – the ‘wheat’ challenge

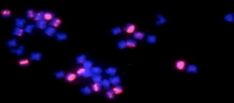
Wheat

Triticum aestivum
(16 Gb)



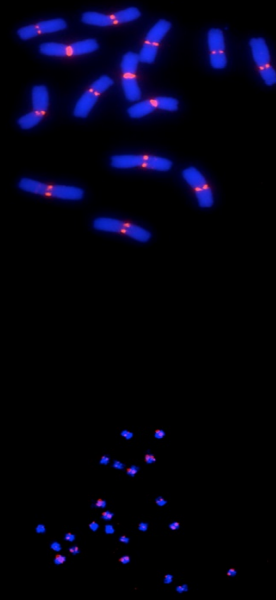
Soybean

Glycine max
(1.1 Gb)



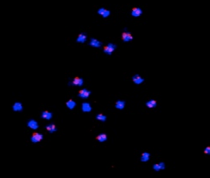
Barley

Hordeum vulgare
(5 Gb)



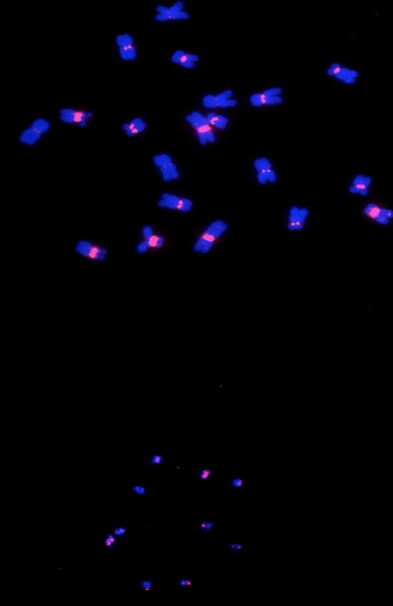
Rice

Oryza sativa
(0.45 Gb)



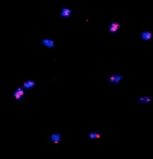
Corn

Zea mays
(2.5 Gb)

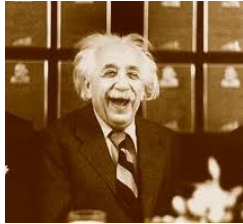


Arabidopsis

Arabidopsis thaliana
(0.15 Gb)



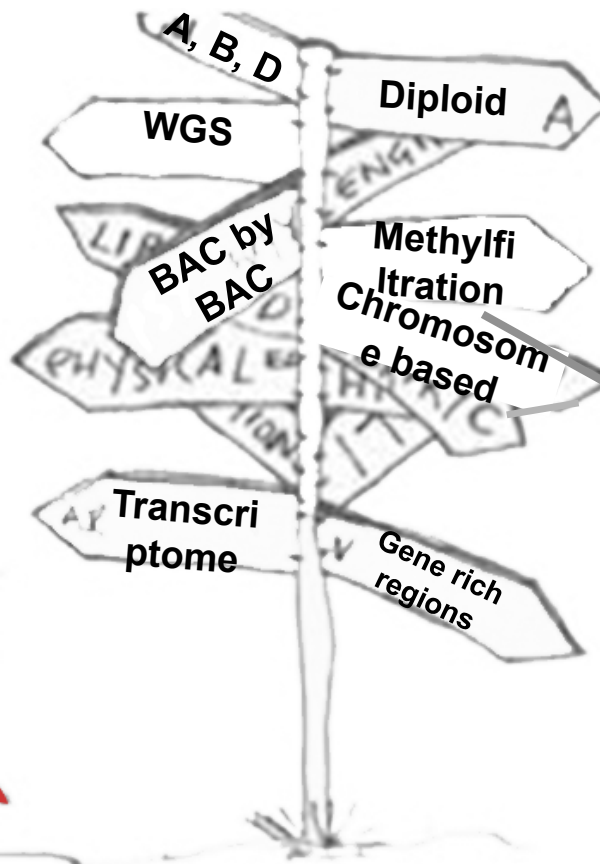
How to produce a useful sequence?



I W G S C



?????? \$
\$\$\$\$



Technology Neutral

For what do we want it?

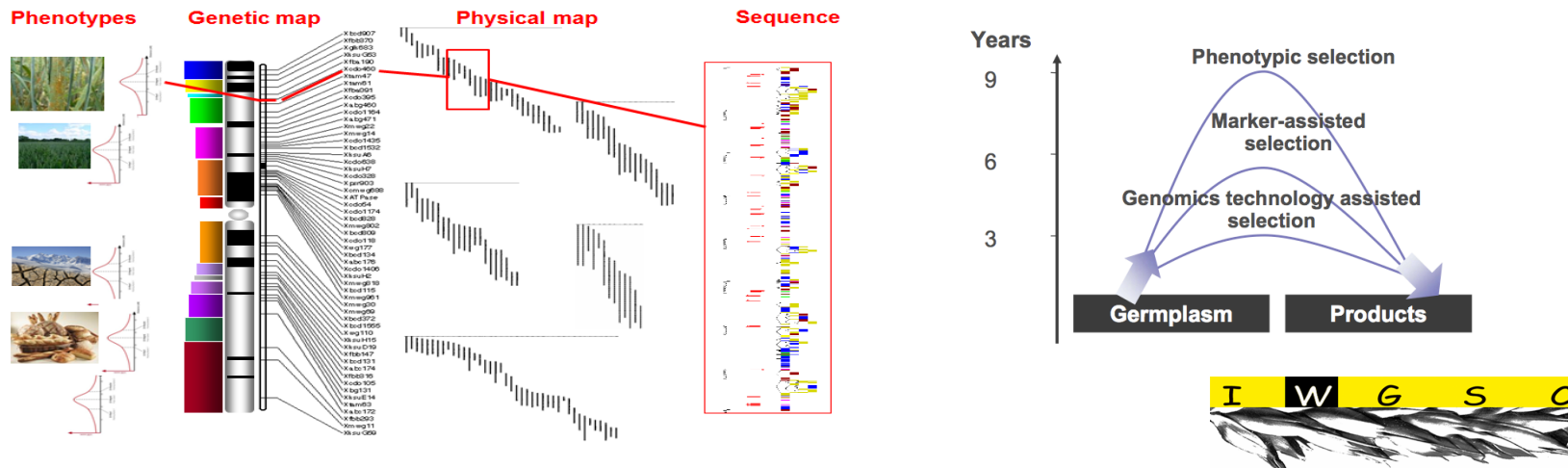
Vision

Goal

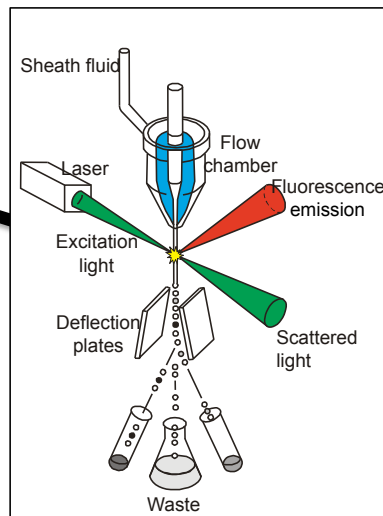
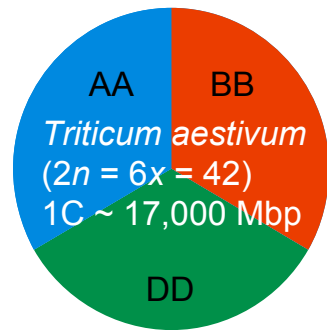
- Lay a foundation to accelerate wheat improvement
- Increase profitability throughout the industry

Vision

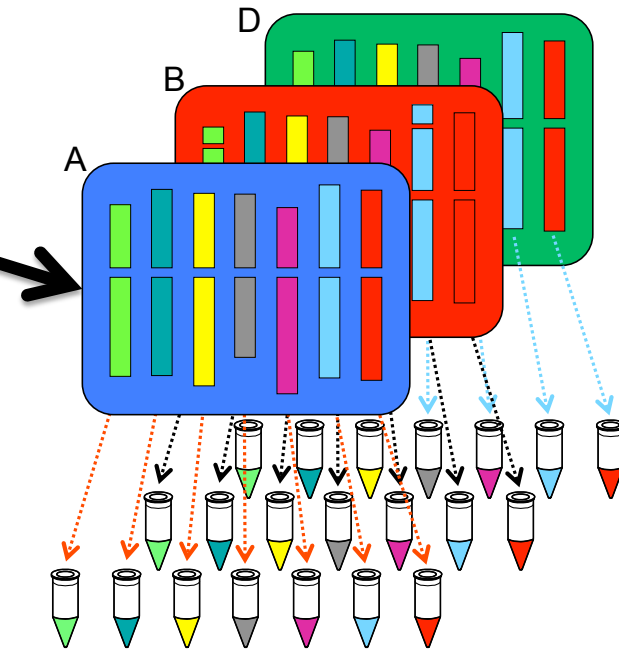
- High quality annotated genome sequence, comparable to rice
- Physical map-based, integrated and ordered sequence



A chromosome-based approach



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



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)

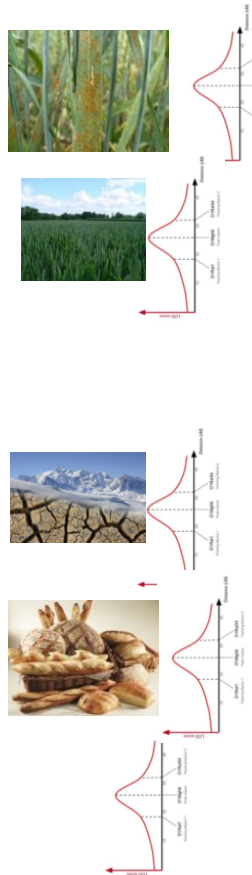
- Chromosome specific BAC libraries (2006 - 2012)
- Amplified DNA for chromosome survey (2010 - 2011)



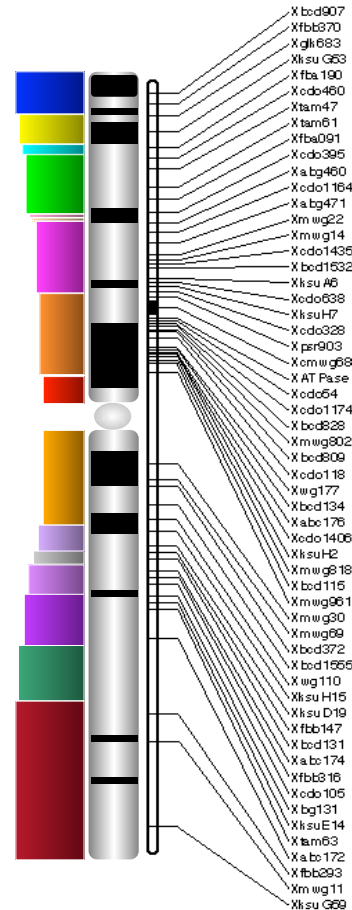


An integrated and ordered 3B reference sequence

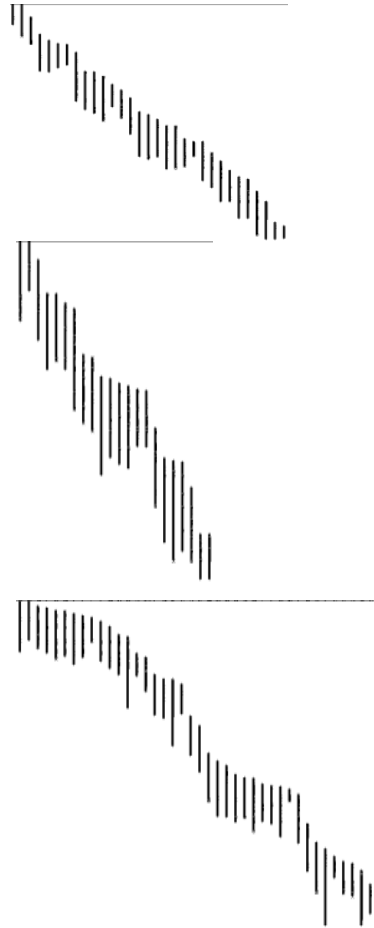
MetaQTL analysis



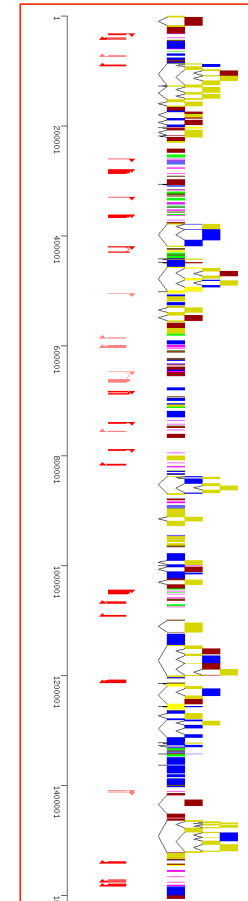
3B consensus map (5000 markers)



3B Physical map

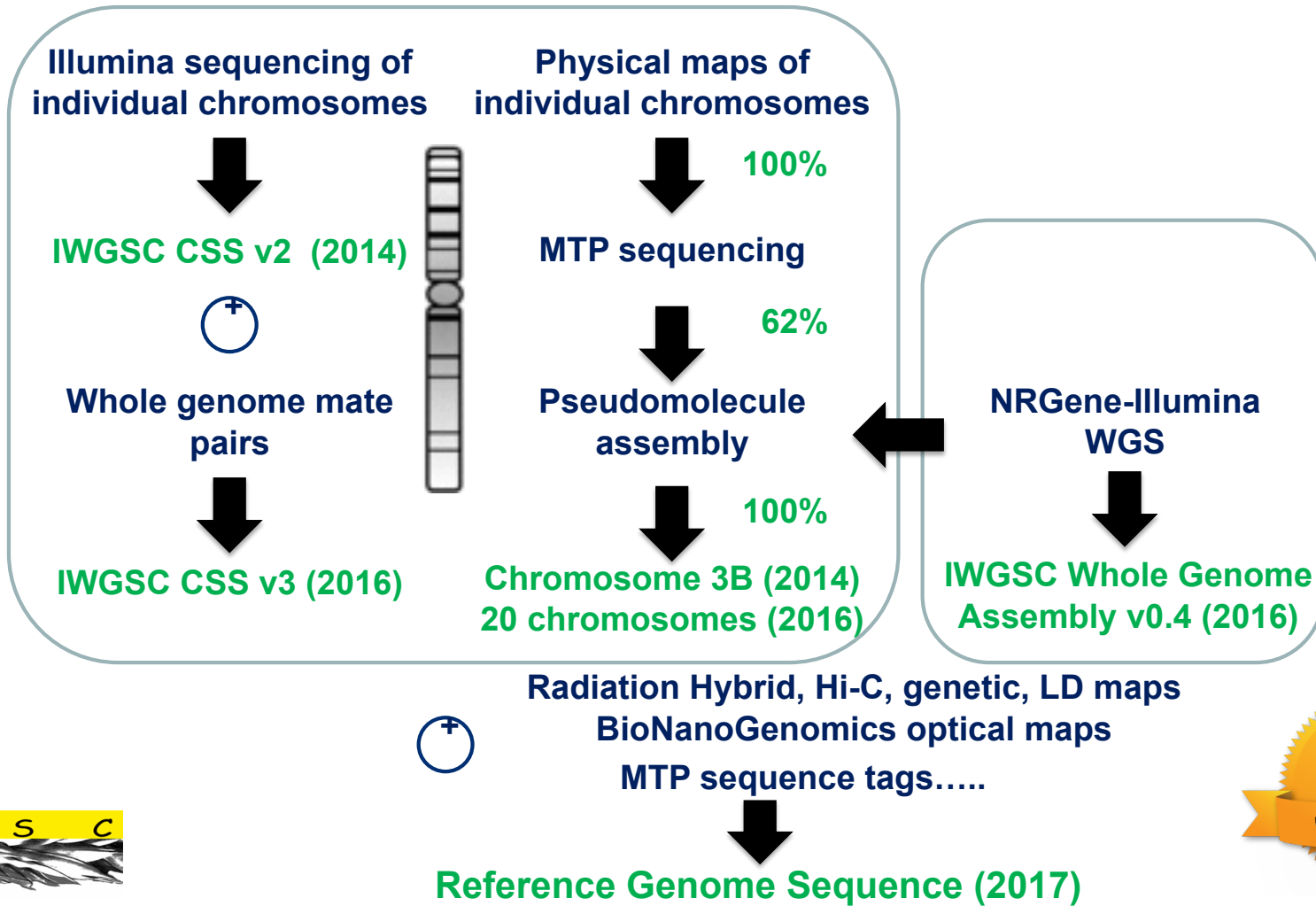


3B pseudomolecule

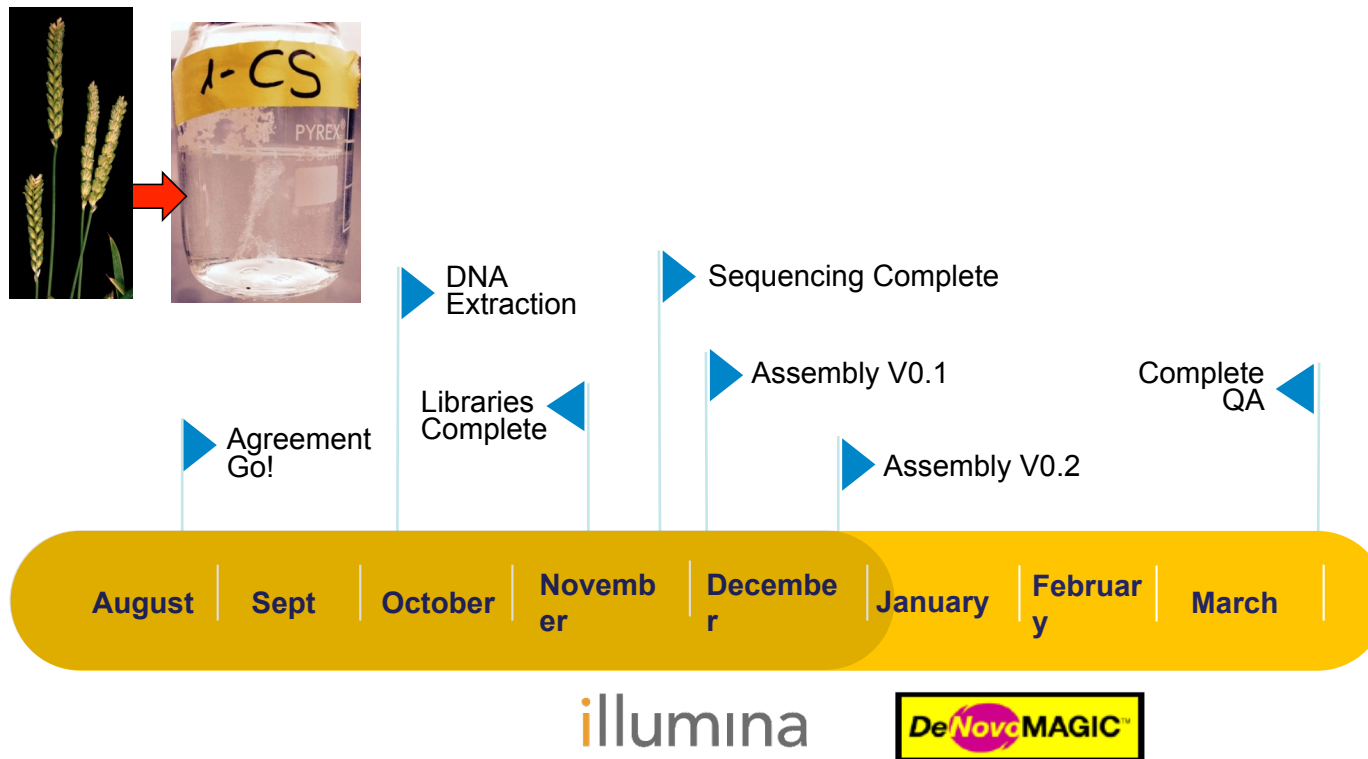


Feuillet, Paux, and Choulet, Science 2008 and Science 2014

Roadmap to the Wheat Genome Sequence



The IWGSC CS WGA Project – timeline 2015



~2 months from data accumulation to completion of first assembly



IWGSC Whole Genome Assembly Project

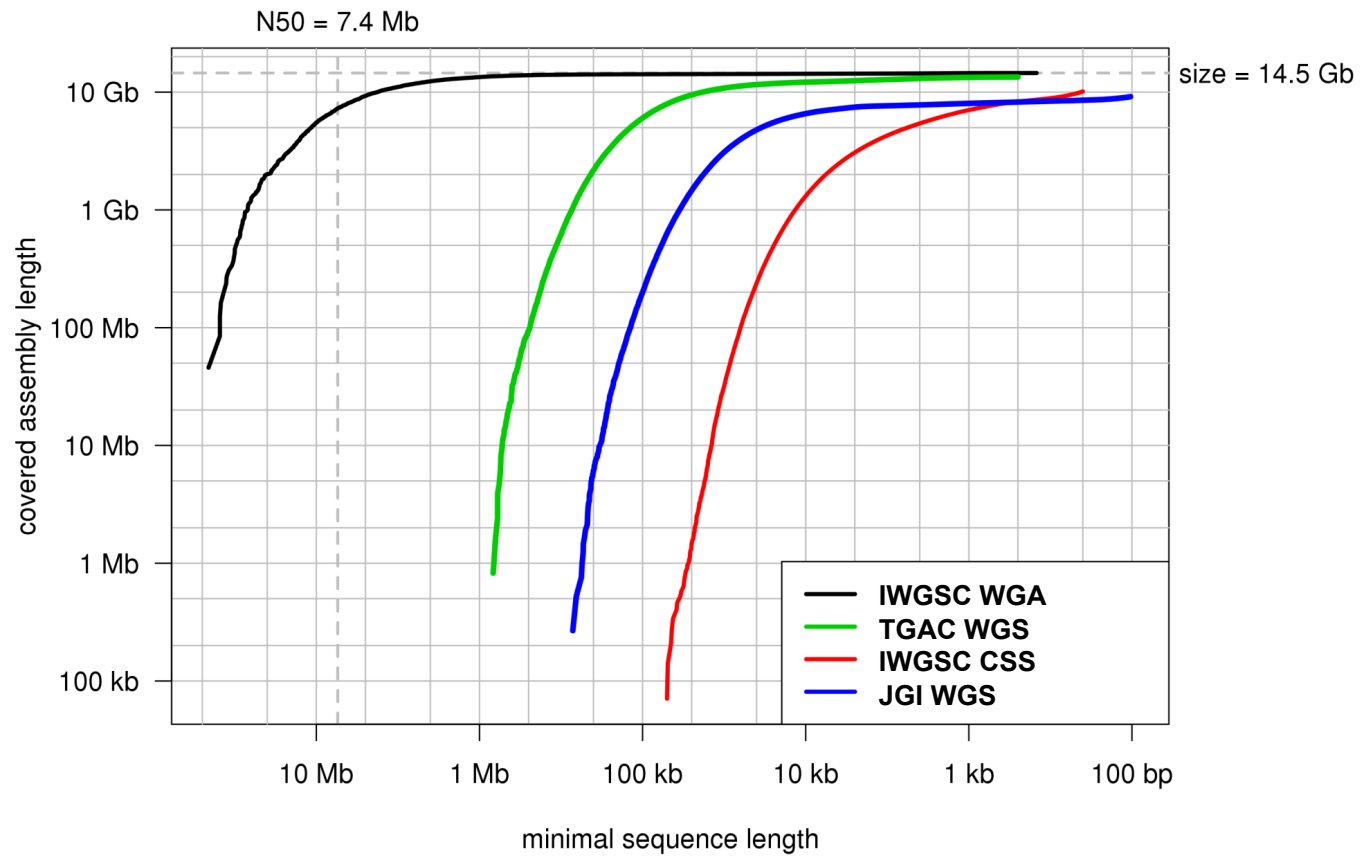
De novo assembly:

- NRGene's DeNovoMagic-2 platform, total run time < 3 weeks, 1Tb RAM computer
- Illumina short-read sequencing data only (200 x coverage)

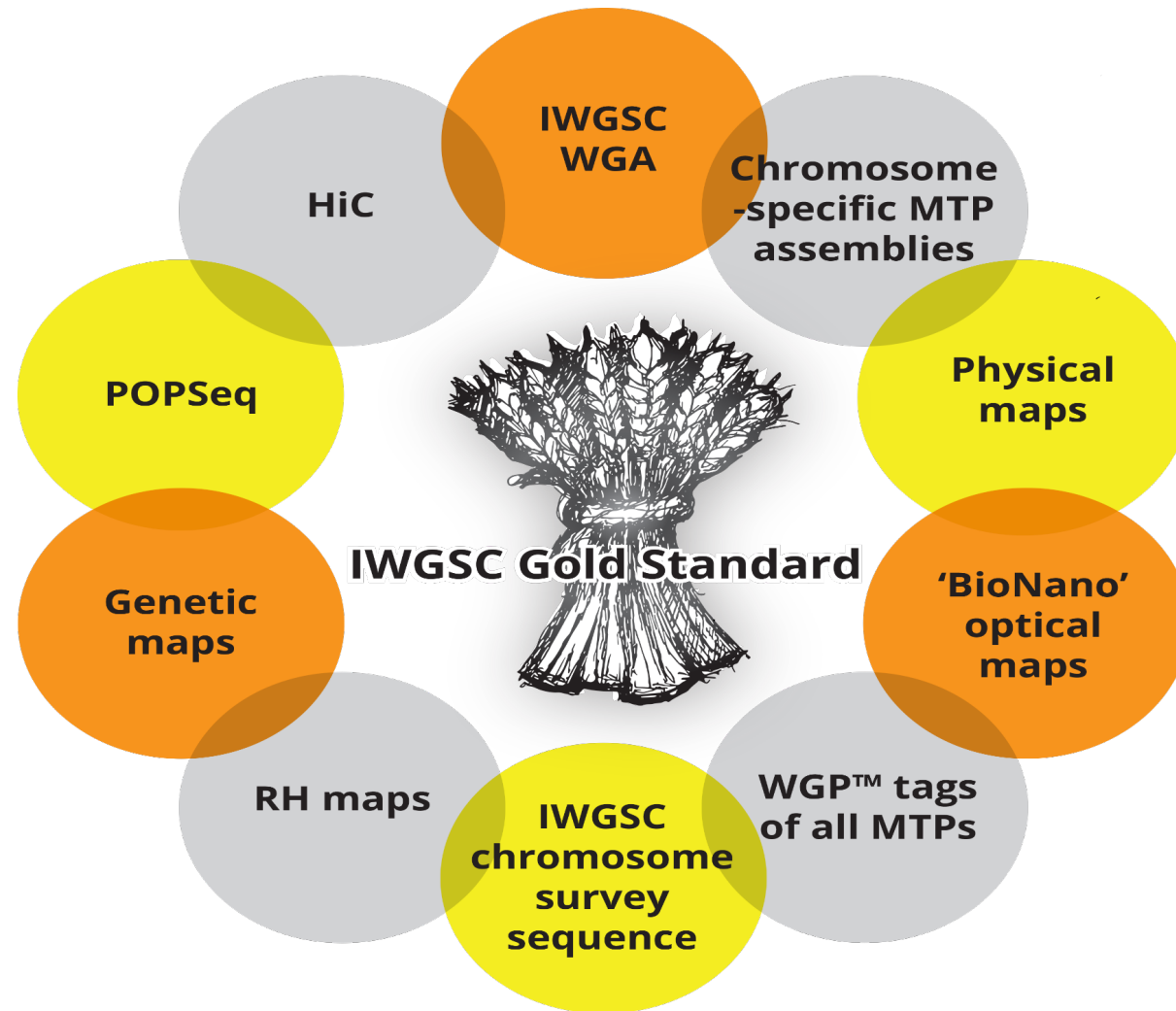
Assembly size:	14.5 Gbp
Est. gaps size:	262 Mbp
Gaps %:	1.80
Total # scaffolds:	138,484
N50:	7.1Mbp
L50 (#sequences):	566
N90:	1.3 Mbp
L90 (#sequences):	2,363
MAX Scaffold:	45.8 Mbp



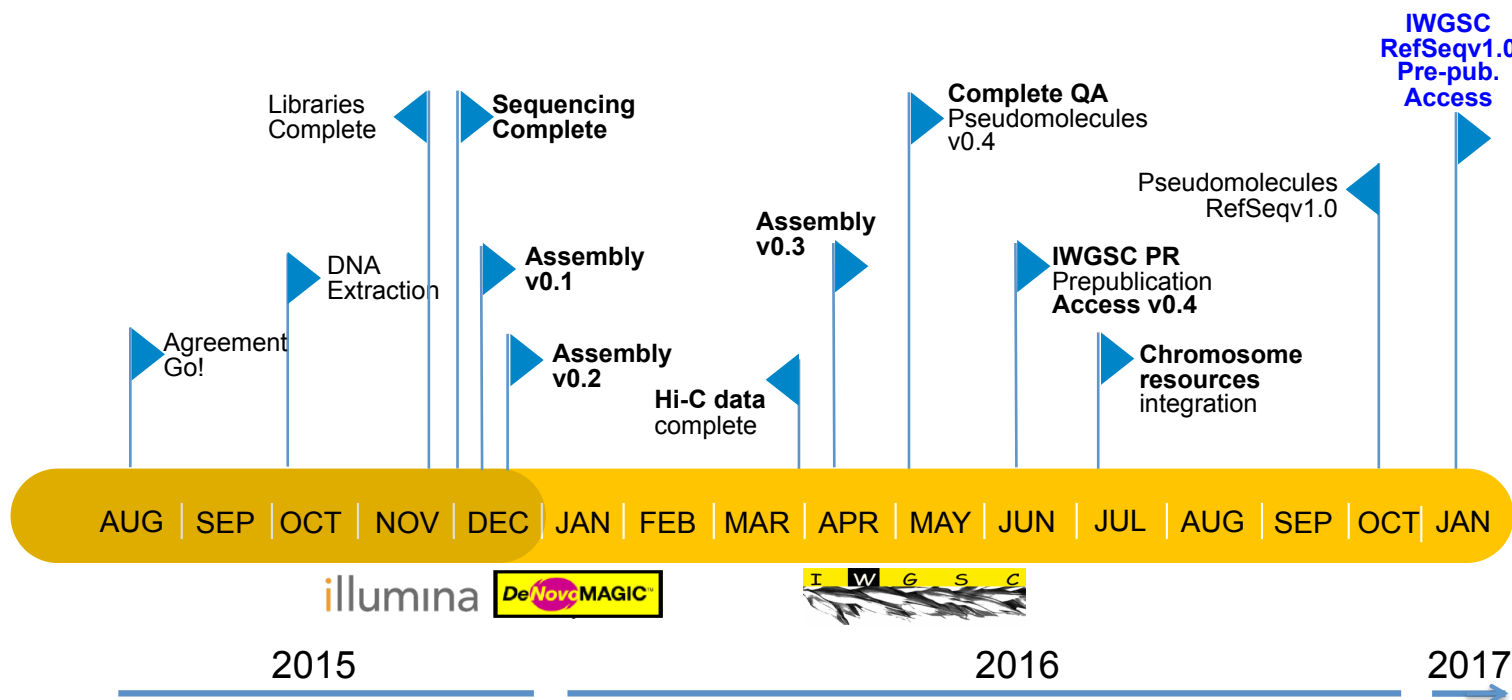
WGA Assembly Statistics



Concerted integration of resources: RefSeq v1.0



IWGSC RefSeq v1.0 Project Timeline



IWGSC RefSeq Project

- Physical maps for all chromosomes
 - ▶ 1,839,128 BACs, 47,810 contigs, 380,675 singletons
- WGP tags (mostly from MTP BACs) for all chromosomes except 3B
 - ▶ 4,305,249 unique tags, 693,697 BACs
- BAC sequence assemblies for 8 chromosomes (1A, 1B, 3B, 3D, 6B, 7A, 7B, 7D) and partial MTP data for two arms (4AL, 5BS)
 - ▶ 52,890 BACs (9.7 Gb), N50 - 68 kb
- Optical maps for 7A, 7B and 7DS
 - ▶ 1,335 BioNanoGenomics contigs aligned to the WGA assembly
- GBS map of the SynOp RIL population
 - ▶ 179 RILs, 4074 markers



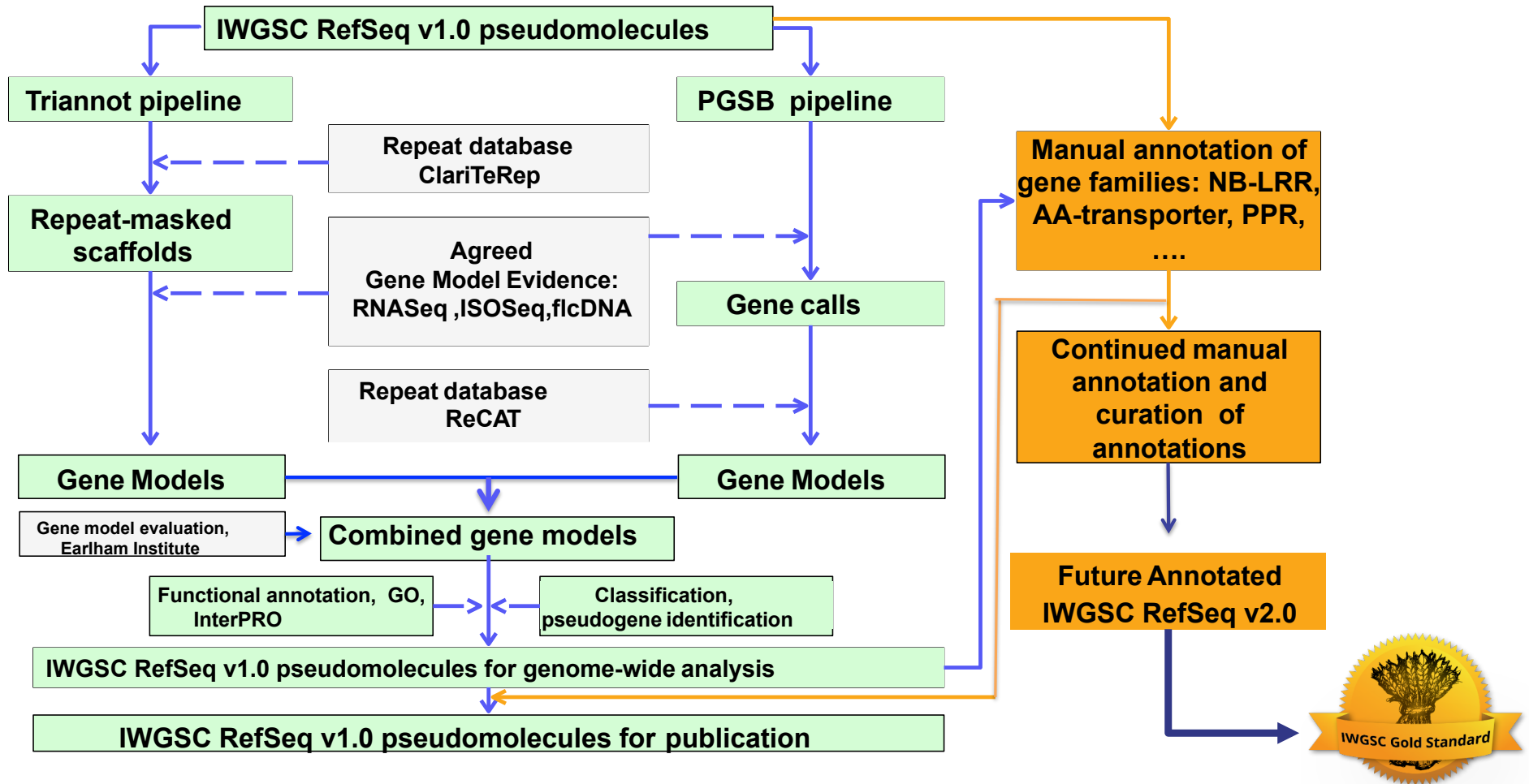
Comparison of IWGSC Assembly Releases

	IWGSCv0.4	RefSeqv1.0
Number / coverage of scaffolds/contigs	138,607 /14.5 Gb	138,665 /14.5 Gb
Number / coverage of scaffolds/contigs >=100kb	4,442 / 14.2 Gb	4,443 / 14.2Gb
N50 scaffolds / superscaffolds	7.0 Mb	22.8 Mb
L50 (no. sequences →N50)	566	166
N90 scaffolds / superscaffolds	1.3 Mb	4.1 Mb
L90 (no. sequences →N50)	2363	718
Gaps filled with BAC sequences		183 (1.7 Mb)
Average size of inserted BAC sequences		9.5 kb
Sequence assigned to chromosomes	14.1 Gb (96.8%)	14.1 Gb (96.8%)
Sequence assigned to chromosomes (>=100kb)	14.1 Gb (99.1%)	14.1 Gb (99.1%)
No. scaffolds / superscaffolds on chromosomes	3,975	1,601
No. oriented scaffolds / superscaffolds	2,464	1,243
Oriented sequence	13.1 Gb (90.2%)	13.8 Gb (95%)
Oriented sequence >=100kb	13.1 Gb (92.4%)	13.8 Gb (97.3%)

**RefSeq
v1.0
contains
~ 75
scaffolds
per
chrom.**



IWGSC RefSeq v1.0 Annotation



IWGSC RefSeq Data Access & Availability



<https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies>

Pre-publication data access:

IWGSC WGA v0.4: June 13, 2016

IWGSC RefSeq v1.0: January 14, 2017

Gene models completed: March 2017

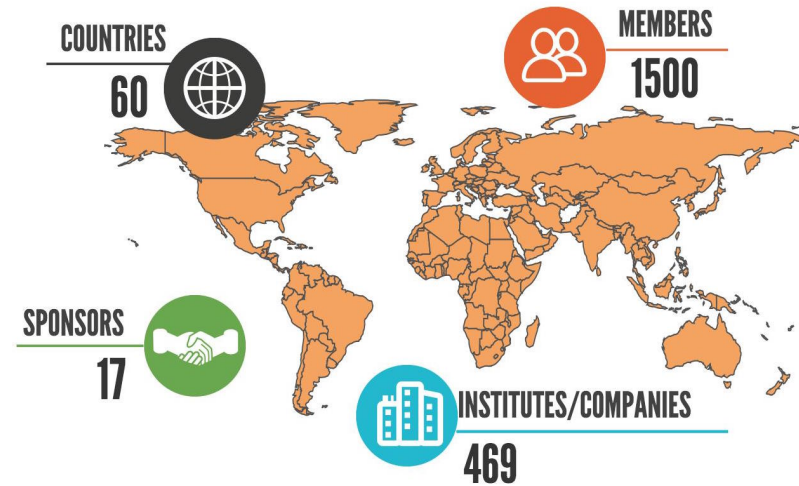
Final analyses completed: April/May 2017

Manuscript submission: Summer 2017



The IWGSC Today

2017



www.wheatgenome.org



TAKING
ACTION
FOR YOU

HAPPENING NOW

ARRIVAL IN PORT OF HISTORIC IWGSC CRUISE
1ST REFERENCE SEQUENCE OF BREAD WHEAT ACHIEVED

abc ACTION NEWS
9:02 75°



IWGSC 2.0

- Manual and functional annotation of the sequence to empower gene discovery and gene cloning to understand the molecular basis of traits
- Coordination of re-sequencing activities for diversity panels that represent the breadth of geographic distribution of germplasm for breeders
- Support the development of user-friendly, integrated databases



Lessons learned

- At least one high quality, manually annotated reference sequence
- Physical map-based for adaptability to any technology
- Maintain flexibility for new technologies without losing sight of quality
- Stay on the course towards your vision



Acknowledgments

IWGSC Leadership: Rudi Appels, Kellye Eversole, Catherine Feuillet, Beat Keller, Jane Rogers

IWGSC Chromosome Leaders:



Etienne Paux, Frédéric Choulet



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

Jaroslav Dolezel, Hana Simkova, Miroslav Valarik, Jan Bartos



Hirokazu Handa



Luigi Cattivelli

Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria



Institute of Cytology and Genetics
OF SIBERIAN BRANCH OF THE RUSSIAN ACADEMY OF SCIENCES

Elena Salina



Nikolai Ravin



Bikram Gill



Bayer CropScience

**Catherine Feuillet
John Jacobs**



**Universität
Zürich^{UZH}**

Beat Keller



University of Haifa

Abraham Korol



Odd-Arne Olsen



**Murdoch
UNIVERSITY**

Rudi Appels



Hikmet Budak



**Nils Stein
Thorsten Schnurbusch**



**UNIVERSITY OF
SASKATCHEWAN**

**Curtis Pozniak
Andrew Sharpe**



Kuldeep Singh



NORTHWEST A&F UNIVERSITY

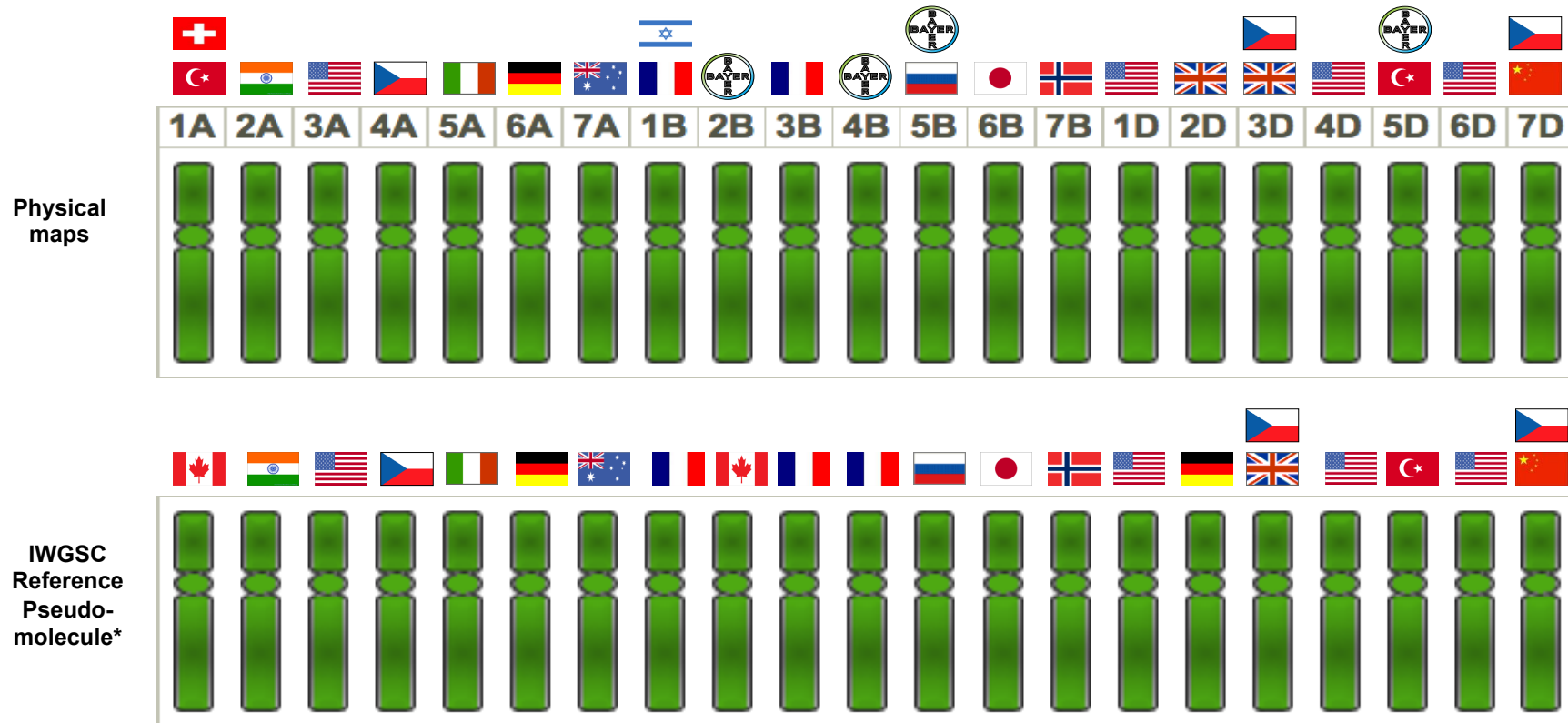
Song Weining



**Earlham
Institute**

Matt Clark

Acknowledgments



All physical maps and pseudo-molecule sequences available at IWGSC repository: <https://wheat-urgi.versailles.inra.fr>

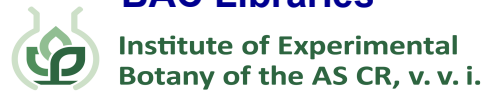


IWGSC RefSeq v1.0 Team Leaders

IWGSC Sequence Repository

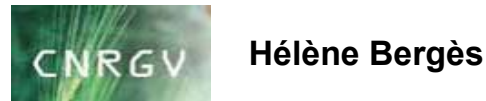


BAC Libraries

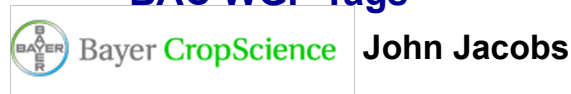


Jaroslav Dolezel, Hana Simkova

BAC Library Pools



BAC WGP Tags



Genetic Maps



RH Mapping



WGA PIs



Curtis Pozniak
Andrew Sharpe



NRGene Gil Ronen



Kellye Eversole
Jane Rogers

Pseudomolecule Team



Gabriel Keeble-Gagnere



Annotation Team



HelmholtzZentrum münchen
Deutsches Forschungszentrum für Gesundheit und Umwelt

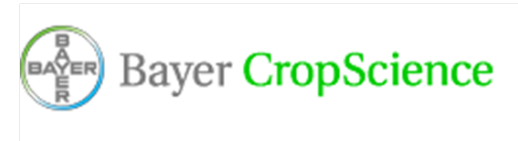
Manuel Spannagl, Klaus Mayer



RNASeq



IWGSC Sponsors



Thank you for your attention!

