IWGSC-WHOLE GENOME SHOTGUN ASSEMBLY OF CHINESE SPRING


Crop Development Centre Durum Wheat Breeding and Genetics Program

UNIVERSITY OF SASKATCHEWAN
Crop Development Centre
COLLEGE OF AGRICULTURE AND BIORESOURCES
AGBIO.USASK.CA

I W G S C

Dr. Curtis Pozniak

IWGSC Workshop,
PAG Jan 2016

©Use and/or distribution of these slides is prohibited unless approval by the author is granted

Principle Investigators (*) and Co-authors

IWGSC

Kellye Eversole*
Jane Rogers *

INRA

Fred Choulet*

GIFS | GLOBAL INSTITUTE FOR FOOD SECURITY
Growing science for life
PotashCorp - a Founding Partner

Andrew Sharpe*

NRC·CNRC

David Konkin
Sateesh Kagale

NRGene

Gil Ronen*
Kobi Baruch
Omer Barad

UNIVERSITY OF SASKATCHEWAN

Curtis Pozniak*
Ron MacLachlan
Kirby Nilsen
Krysta Wiebe

IPK GATERSLEBEN

Nils Stein*
Axel Himmelbach
Ines Walde
Martin Mascher

illumina

Mike Thompson*

K·STATE
Kansas State University

Jesse Poland*

Alvaro Hernandez

ILLINOIS

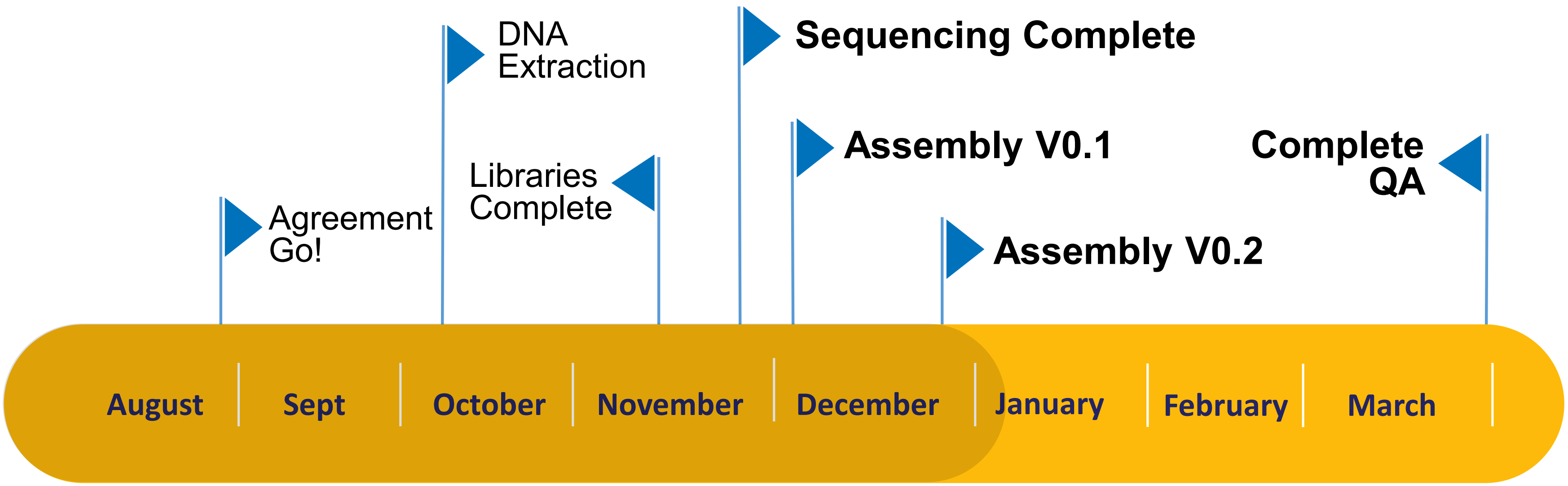
TEL AVIV UNIVERSITY

Assaf Distelfeld *

IWGSC

IWGS-Whole Genome Assembly Project

De novo assembly using **Illumnia** generated (200X) PE and MP data and **NRGene** Assembly pipline.



illumina DeNovoMAGIC™

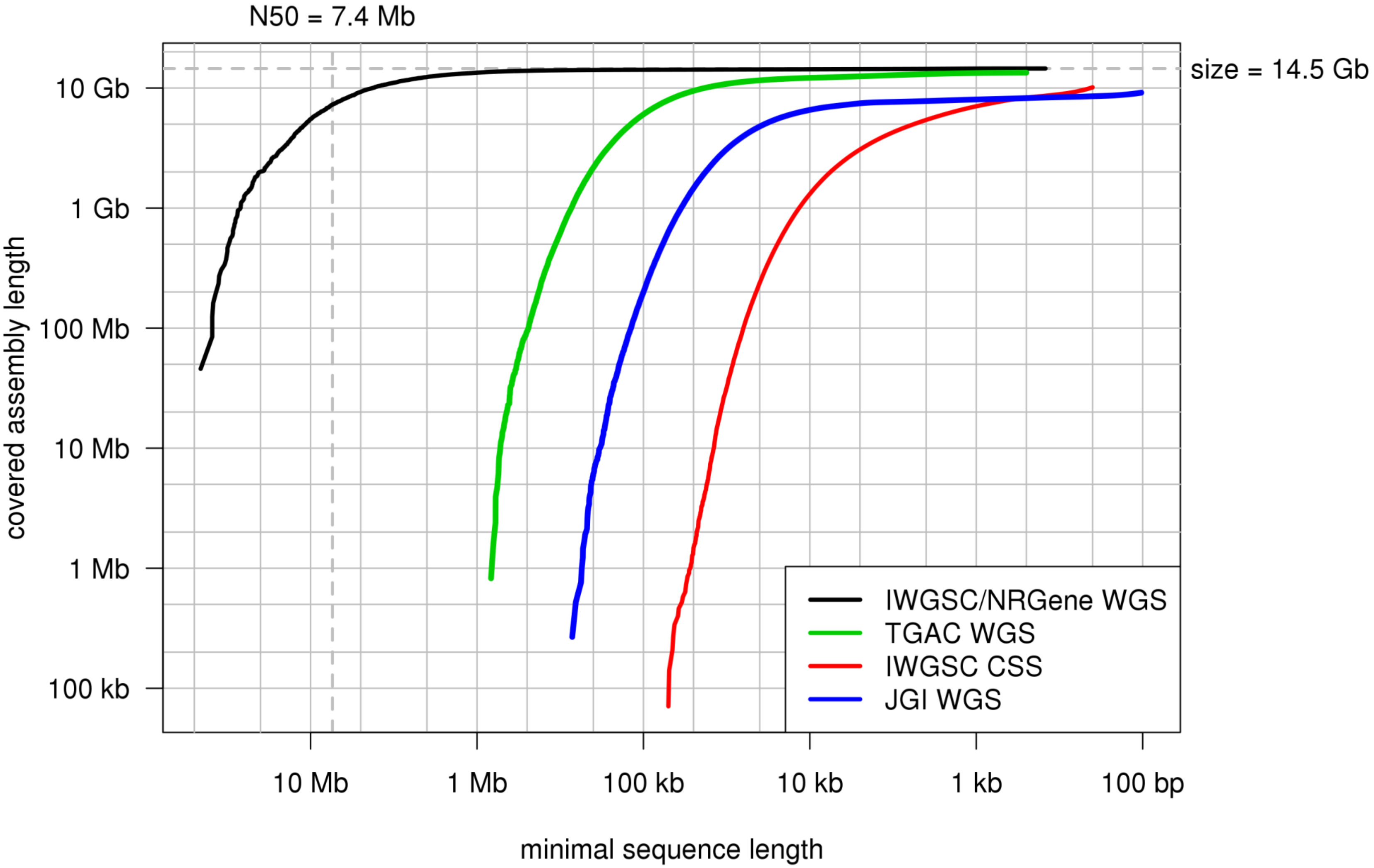
NRGene illumina IPK GATERSLEBEN UNIVERSITY OF SASKATCHEWAN K-STATE Kansas State University TEL AVIV UNIVERSITY

IWGSC-WGA v0.2 Metrics

Total No. Scaffolds	138,484
Assembly size:	14.5 Gb
Gaps size:	262 Mb
Gaps:	1.8 %
L_{50}:	**7.1 Mb**
N_{50} (#sequences):	**566**
L_{90}:	1.26 Mb
N_{90} (#sequences):	2,363
Max size	46 Mb

NRGene illumina IPK GATERSLEBEN UNIVERSITY OF SASKATCHEWAN K-STATE Kansas State University TEL AVIV UNIVERSITY

Assembly Comparisons



Preliminary QC of IWGS-WGA

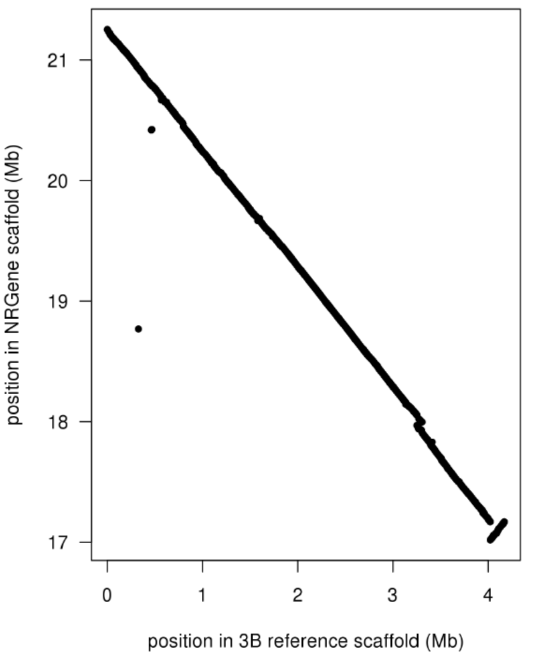
Martin Mascher,
IPK Gatersleben

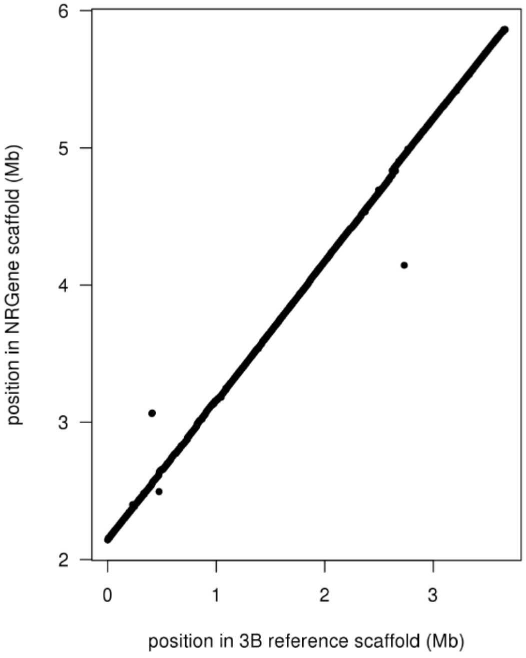
- Alignment to MTP-based assemblies
- Alignment to physical maps
- Genic/Intergenic space
- Anchoring of assembly

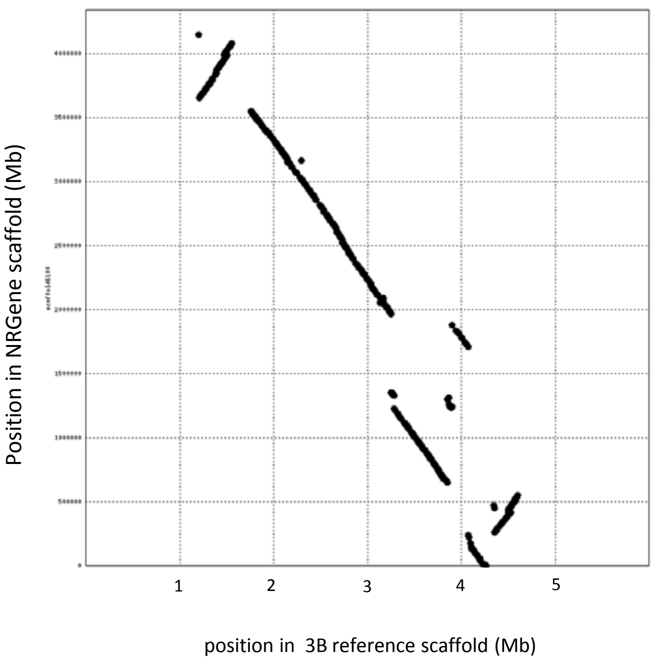
Fred Choulet
INRA Clermont

Alignment to MTP-based Assemblies

>95% of the 3B Pseudomolecule is present in the IWGSC-WGA


v443_0936 (4.2 Mb) vs. scaffold16560-2 (22.6 Mb)

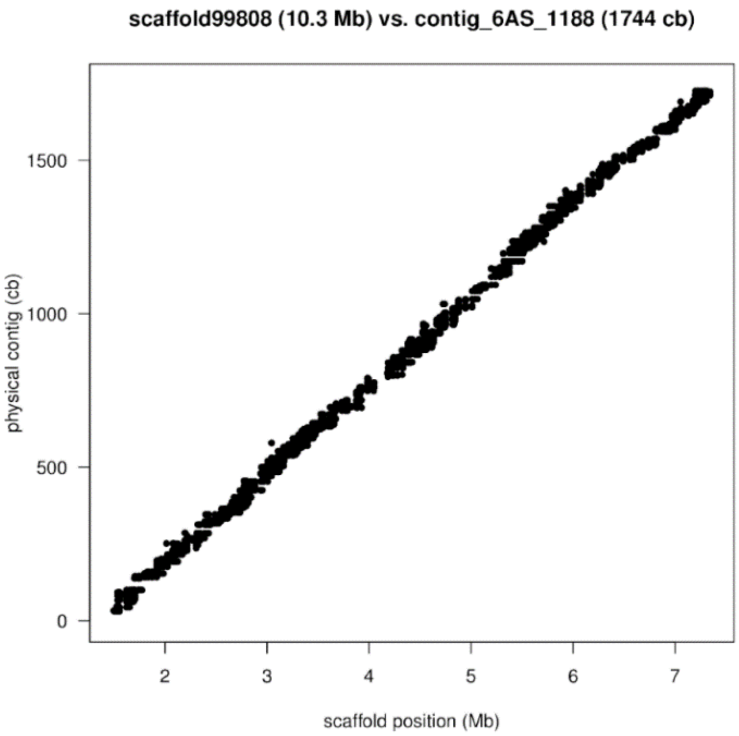

v443_0903 (3.7 Mb) vs. scaffold123840 (10.1 Mb)

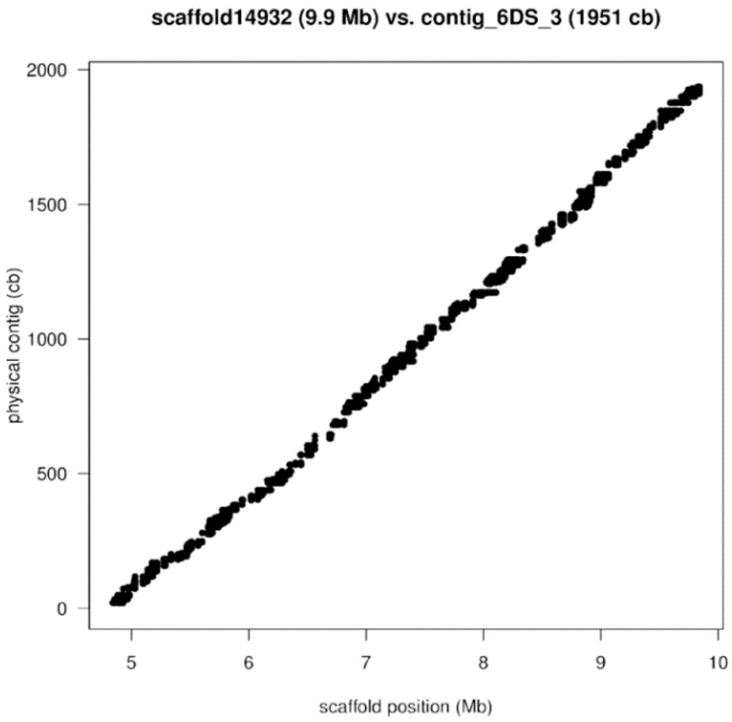

SSt1 ta3B Interval (4 Mb)

IWGSC

Comparison of Physical Maps to CS-WGS Assembly

Mapping of whole genome profiling tags of LTC-derived physical maps to IWGSC-WGA reveals near perfect colinearity


scaffold99808 (10.3 Mb) vs. contig_6AS_1188 (1744 cb)


scaffold14932 (9.9 Mb) vs. contig_6DS_3 (1951 cb)

NRGene illumina IPK GATERSLEBEN UNIVERSITY OF SASKATCHEWAN K-STATE Kansas State University TEL AVIV UNIVERSITY

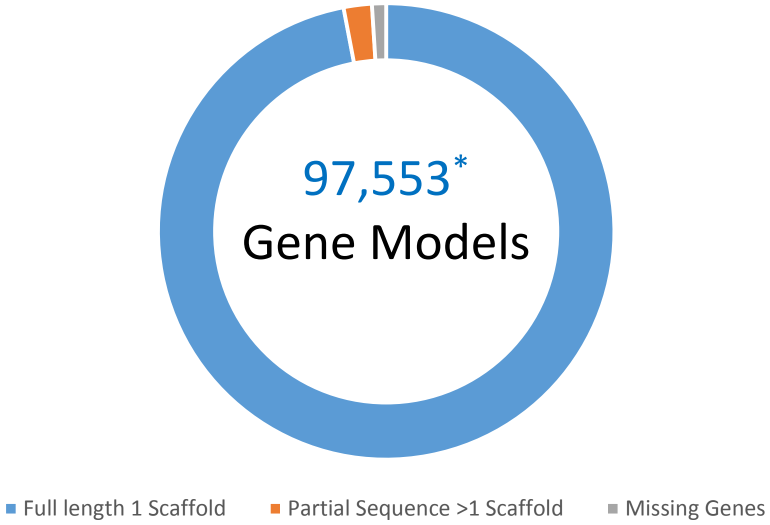
IWGSC

Analysis of Genic / Intergenic Space

>97% of the CSS coding exons are present on a single scaffold

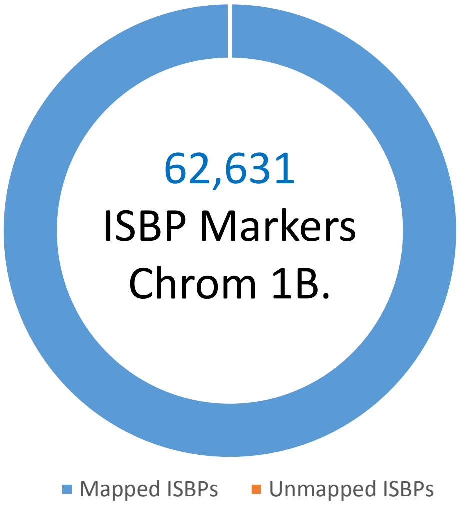
>98% of ISBP markers map to the IWGSC-WGA

IWGSC-CSS Gene Models



**99% identity, 90% coverage*

TE junction-based markers



4315 scaffolds (14.0 Gb) with 1 + CSS genes

3197 scaffolds (13.6 Gb) with 3 + CSS genes

NRGene illumina IPK GATERSLEBEN UNIVERSITY OF SASKATCHEWAN K-STATE Kansas State University TEL AVIV UNIVERSITY

I W G S C

Anchoring of IWGSC-WGA

High density marker data allows anchoring of 14.0 Gb to genetic maps

Source	No. markers	% Mapped
CSS Survey Sequence	(337,000 contigs)	99.3%
Cs x Re	272,218	92.1%
Syn x Opata	(437,973 contigs*)	98.3%
90K iSelect	81,987	97.6%
U of Bristol Axiom**	819,572	98.1%

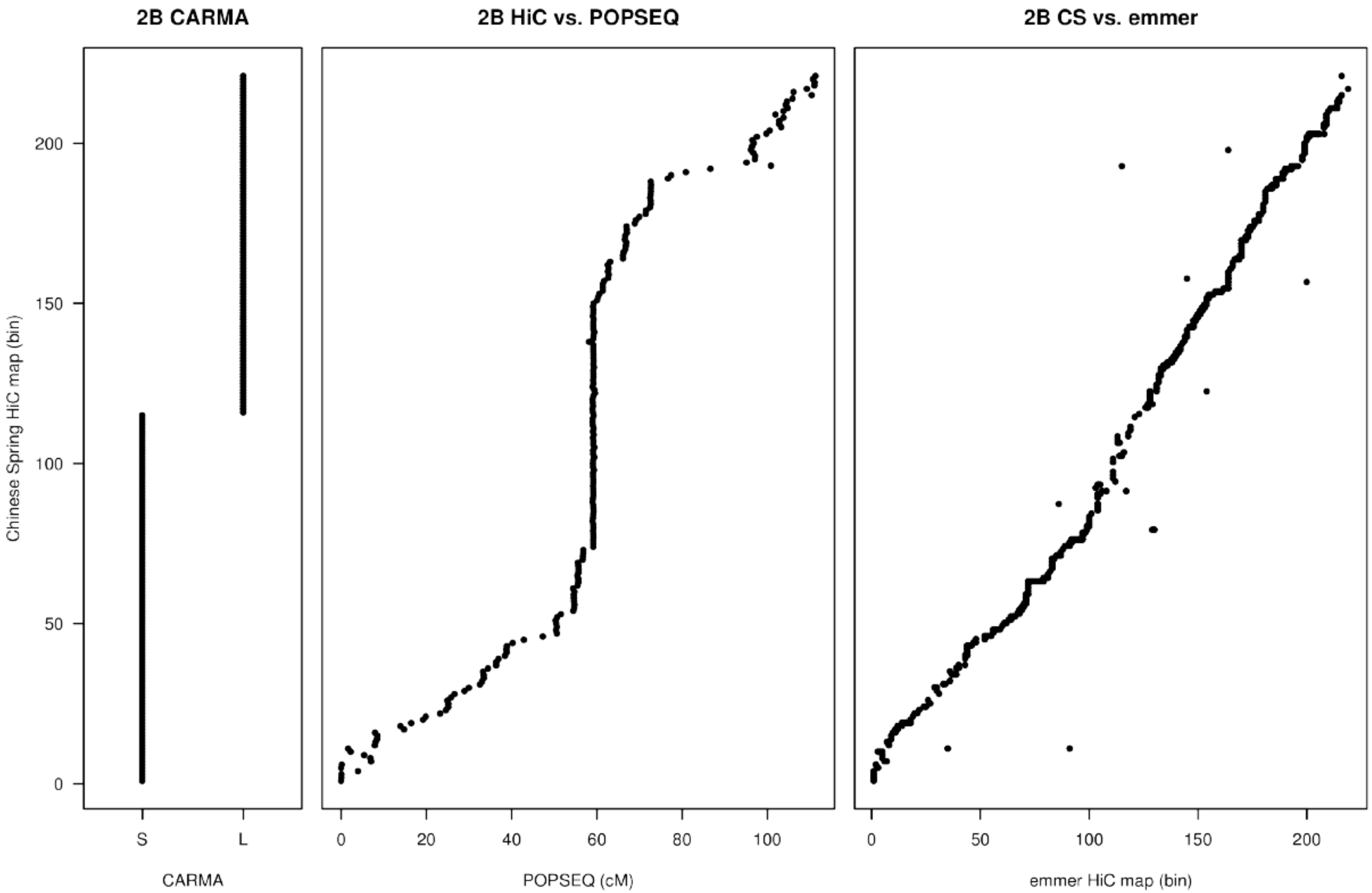
>14.1 Gb (**96.5 %**) assigned to chromosome arms

>431,617 anchored scaffolds map to Syn x Opata

*Number of JGI Contigs anchored to Syn x Opata genetic map
** U of Bristol WISP Project (www.cerealsdb.uk.net)

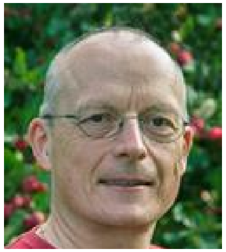
NRGene illumina IPK GATERSLEBEN UNIVERSITY OF SASKATCHEWAN K-STATE Kansas State University TEL AVIV UNIVERSITY

Chromosome arm assignment and Hi-C



IPK GATERSLEBEN


Nils Stein


Axl Himmelbach


Ines Walde


Martin Mascher

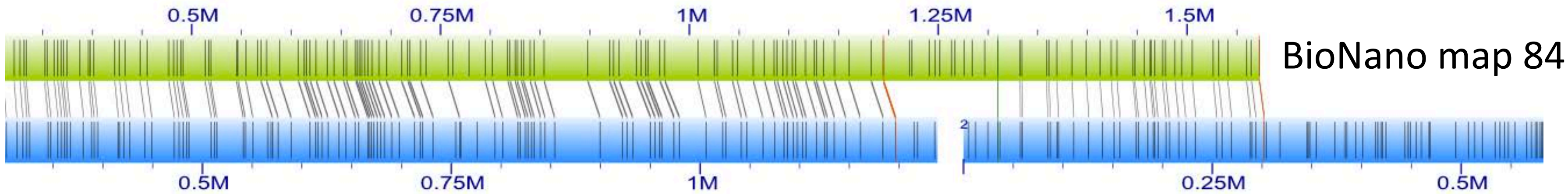
With support from

Federal Ministry
of Food
and Agriculture

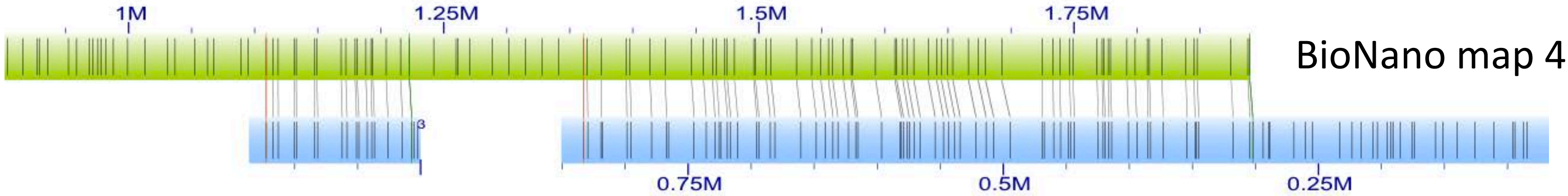
by decision of the
German Bundestag

Optical Mapping

Bionano single molecule maps of chrom. 7A confirm assembly order and allow super-scaffolding of IWGSC-WGA



BioNano map 84



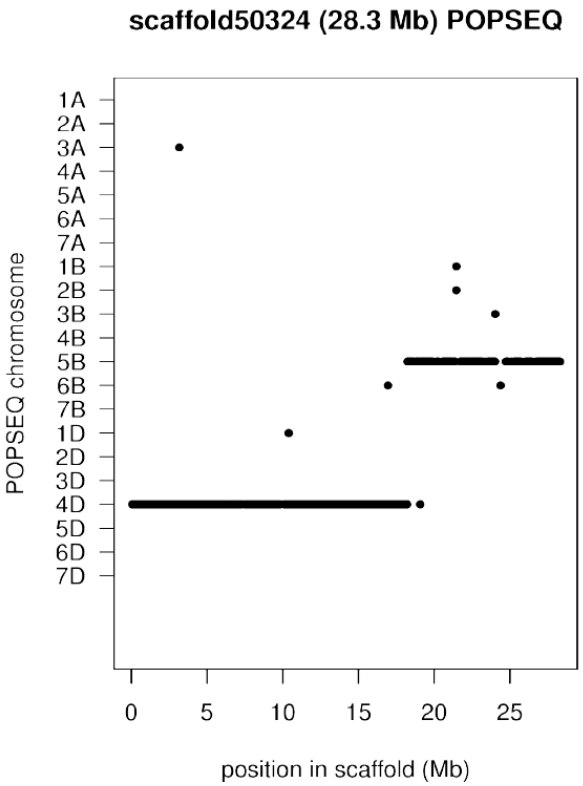
BioNano map 49

Analysis courtesy of Rudi Appels and Gabriel Keeble

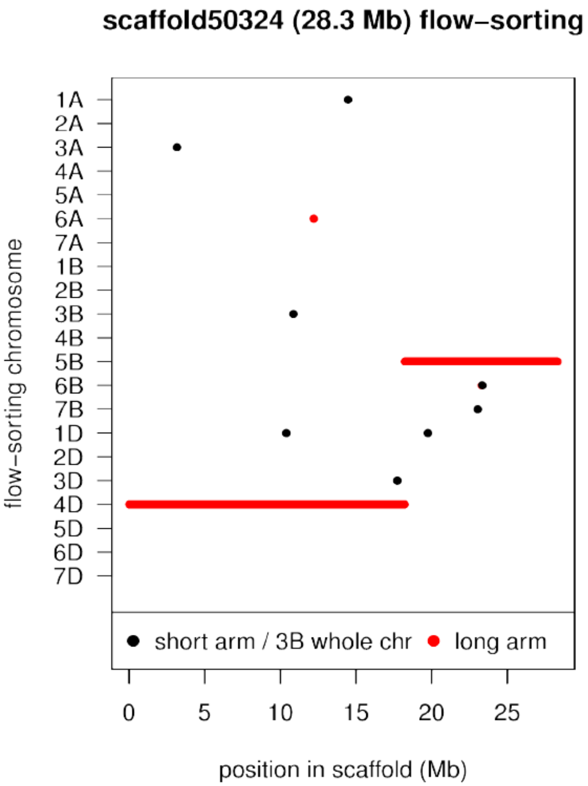
NRGene illumina IPK GATERSLEBEN UNIVERSITY OF SASKATCHEWAN K-STATE Kansas State University TEL AVIV UNIVERSITY

Evidence for Chimeric Assemblies

Approx. 5 % scaffolds >1 Mb are putative chimeras



scaffold50324 (28.3 Mb) POPSEQ

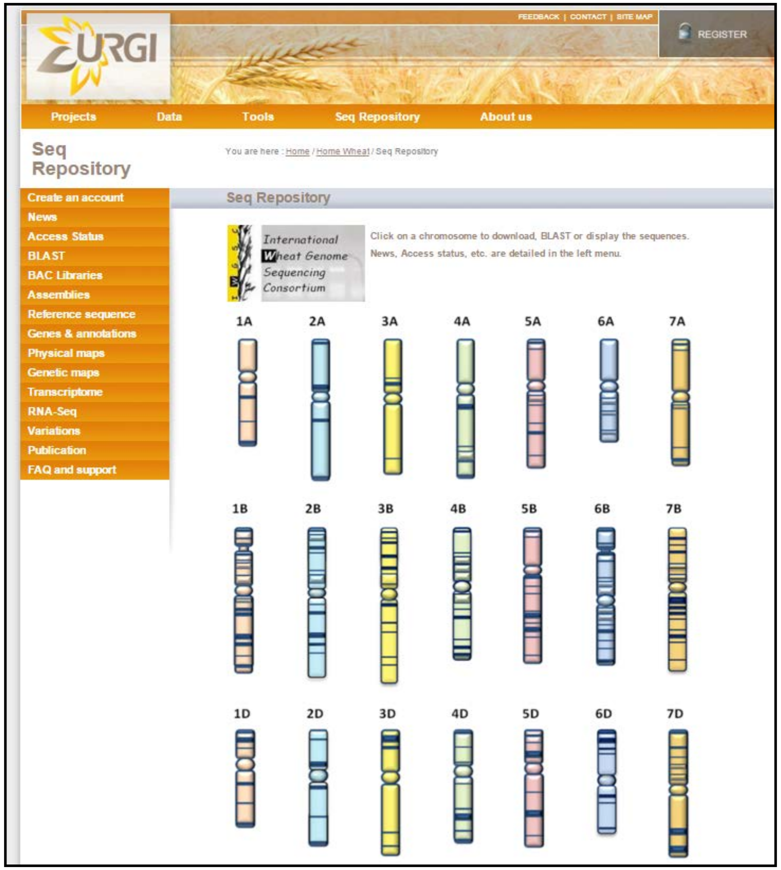


scaffold50324 (28.3 Mb) flow–sorting

IWGSC Whole Genome Assembly Data Release

Initial QA suggests that the IWGSC-WGA is impressively complete

IWGSC Data Repository hosted by INRA-URGI
http://wheat-urgi.versailles.inra.fr/Seq-Repository



Seq Repository

You are here : Home / Home Wheat / Seq Repository

Seq Repository

Click on a chromosome to download, BLAST or display the sequences.
News, Access status, etc. are detailed in the left menu.

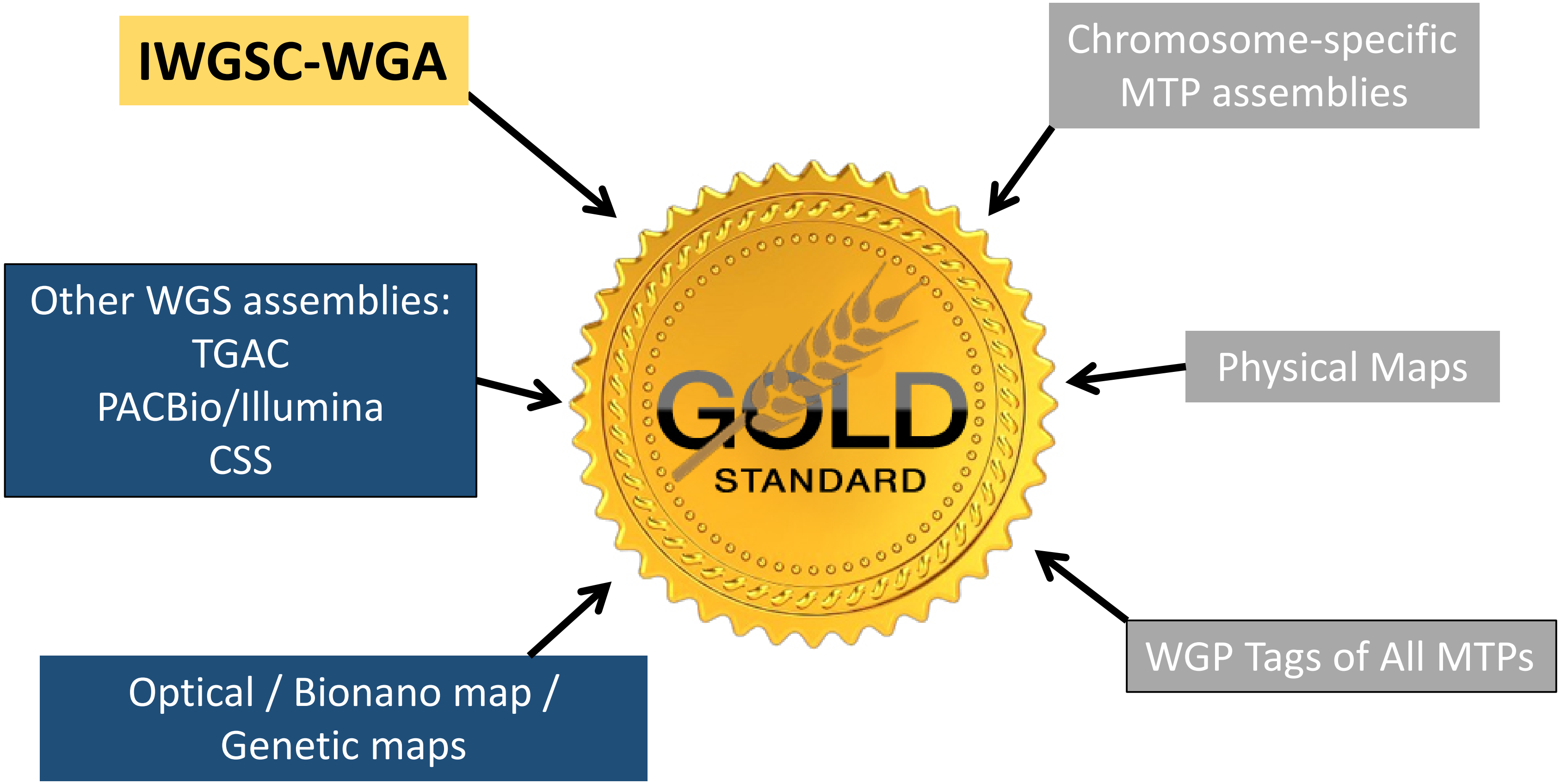
After QC, IWGSC-WGA will be released under **Toronto agreement** through the IWGSC sequence repository

April 2016

NRGene illumina IPK GATERSLEBEN UNIVERSITY OF SASKATCHEWAN K·STATE Kansas State University TEL AVIV UNIVERSITY GIFS | GLOBAL INSTITUTE FOR FOOD SECURITY Growing science for life PotashCorp - a Founding Partner

A Path Forward: Integration of all Resources

A Complete Wheat Genome Sequence by 2017



IWGSC-WGA

Chromosome-specific
MTP assemblies

Other WGS assemblies:
TGAC
PACBio/Illumina
CSS

Physical Maps

WGP Tags of All MTPs

Optical / Bionano map /
Genetic maps

Discussion initiated and ongoing!

Thanks to Our Funders!

CTAG²
Canadian Triticum Applied Genomics

NSF

USDA

GenomeCanada

GenomePrairie

Federal Office for
Agriculture and Food

SaskWheat
DEVELOPMENT COMMISSION

Alberta Wheat
COMMISSION
GROWTH.
INNOVATION.
RESULTS.

Government
of
Saskatchewan

Western
Grains
Research Foundation

illumina

NRGene

illumina

IPK
GATERSLEBEN

UNIVERSITY OF
SASKATCHEWAN

K-STATE
Kansas State University

TEL AVIV UNIVERSITY