



Genome sequences for the Triticeae

Nils Stein, IPK Gatersleben



Wheat News from Hogwarts

The IWGSC whole genome assembly project

Curtis Pozniak*, Andrew Sharpe*: U of S, GIFS, CRC

Jesse Poland*: KSU, Kansas Wheat Commission

Assaf Distelfeld*: U of Tel Aviv

Gil Ronen*, Kobi Baruch, Omer Barad: NRGene

Alvaro Hernandez: U of Illinois

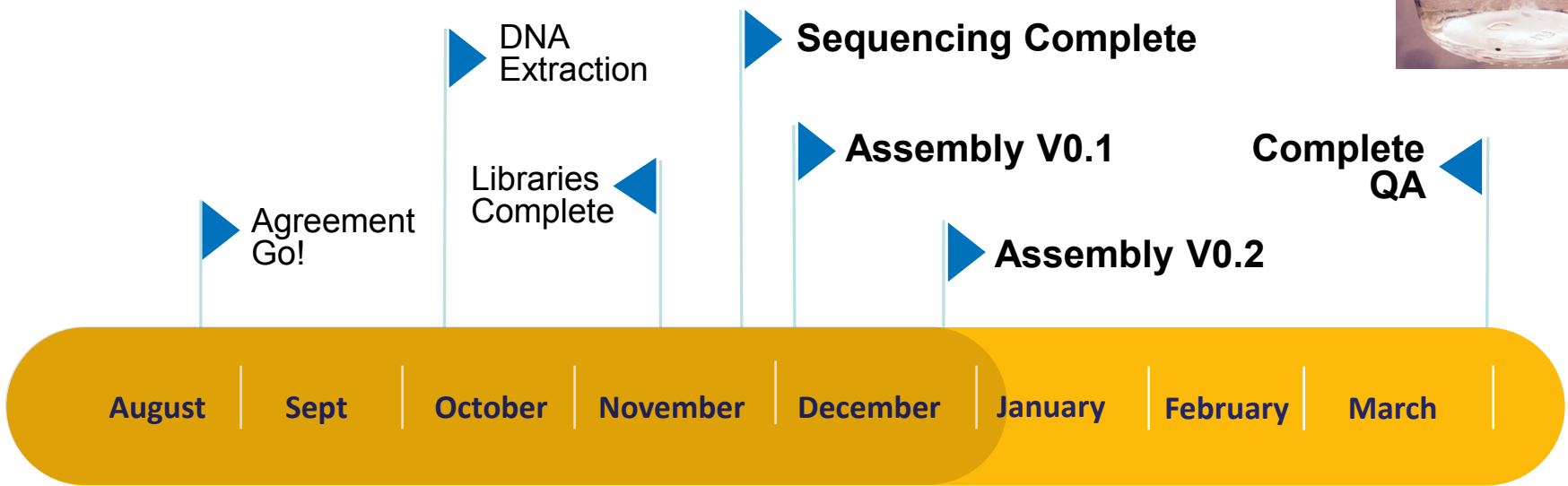
Mike Thompson*: Illumina Inc.

Kellye Eversole*, Jane Rogers*: IWGSC

Fred Choulet*: INRA GDEC, Clermont-Ferrand

Axel Himmelbach, Ines Walde, Martin Mascher, Nils Stein*: IPK Gatersleben

The IWGSC WGA project - timelines



illumina



~2 months from data accumulation to completion of first assembly



The IWGSC WGA project

De novo assembly:

- NRGene's DeNovoMagic-2 platform, total run time < 3 weeks, 1Tb RAM computer
- illumina short-reads sequencing data only (x200 coverage)

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

The IWGSC WGA – initial QA

- Comparison to CSS contigs
- Alignment to physical maps: 6A, 6D
- Alignment to BAC-based scaffolds of 3B reference sequence
- gene space, intergenic space analysis
- Comparison to bionano map of 7DS
- Hi-C map for linear ordering



Martin Mascher,
IPK Gatersleben

IWGSC WGA / CSS contigs and POPSeq

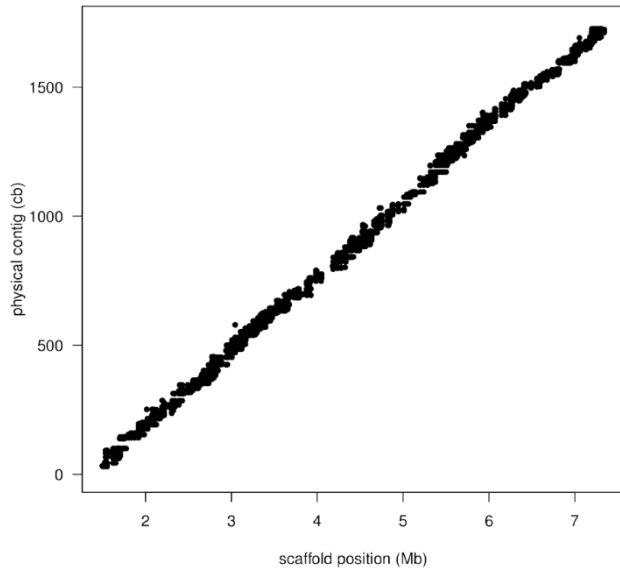
- CSS contigs mapped to IWGSC WGA
 - chromosome arm assignment
 - POPSeq positions
- 344,927 / 347,492 (99.3 %) CSS contigs mapped to the NRGene assembly
- 14.1 Gb (96.5 %) assigned to chromosome arms
- 14.0 Gb (97.2 %) anchored to the Synthetic W7984 x Opata map

IWGSC WGA / 6A and 6D WGP contigs

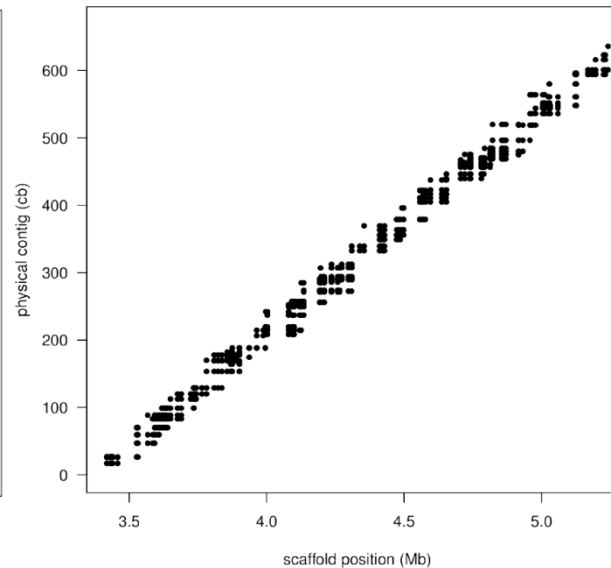


Martin Mascher,
IPK Gatersleben

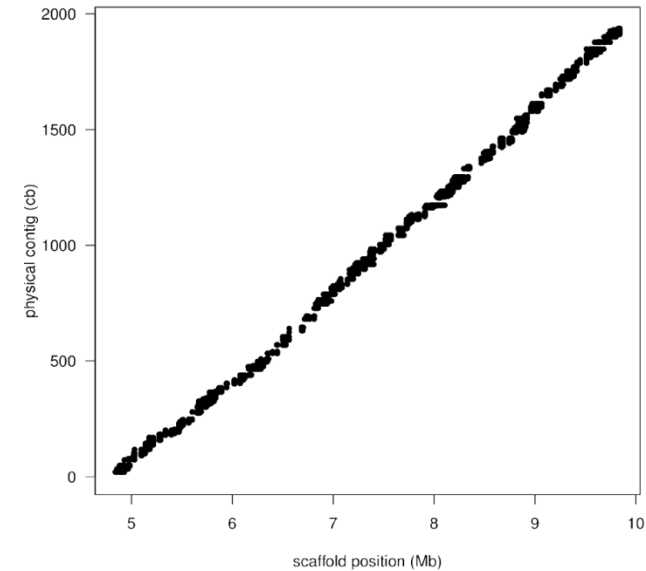
scaffold99808 (10.3 Mb) vs. contig_6AS_1188 (1744 cb)



scaffold23002 (5.3 Mb) vs. contig_6AL_627 (668 cb)



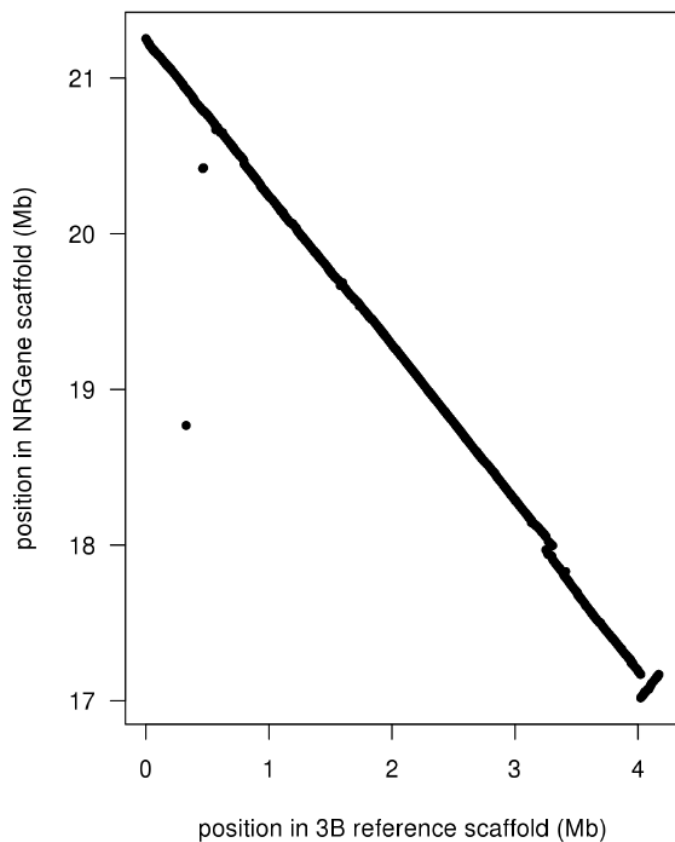
scaffold14932 (9.9 Mb) vs. contig_6DS_3 (1951 cb)



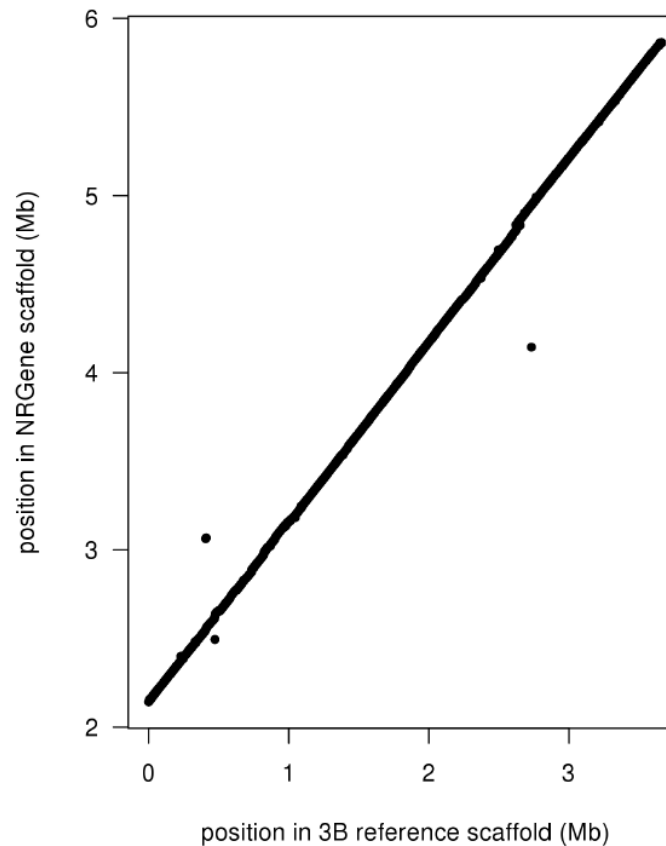
Whole genome profiling tags of entire LTC / FPC contigs mapped against IWGSC WGA scaffolds

Alignment to 3B reference contigs

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



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





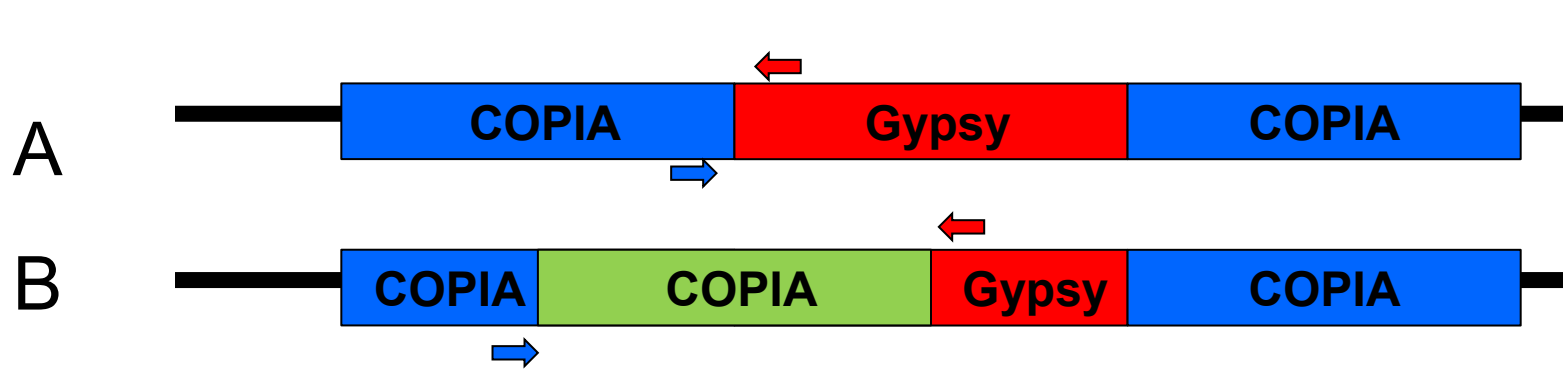
Martin Mascher,
IPK Gatersleben

Completeness of gene space

Analysis against wheat full length cDNAs:

- 5,992 out of 6,137 (97.6%) fl-cDNAs align to IWGSC WGA (99% identity, 80% sequence coverage)
- 305 out of 309 (98.7%) cDNAs aligned to the 3B scaffolds are present in IWGSC WGA (same criteria as above)
- 31 cDNAs aligned to both IWGSC WGA scaffolds and CSS contigs and anchored to 3B are missing or are incomplete in the 3B scaffolds

The intergenic space – ISBP markers



Fred Choulet
INRA Clermont

- 62,631 1B-ISBPs (=TE junctions)
 - ▶ 11 (0.02%) ISBP unmapped on IWGSC WGA scaffolds
- 4,247,440 CSS-ISBPs
 - map 151 bp-ISBPs with BWA
 - filter fully mapped ISBPs with 0-1 mismatch
 - ▶ 97.7% (4,150,743) mapped

Linking to 7DS bionano map

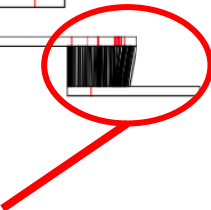
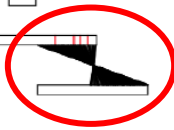
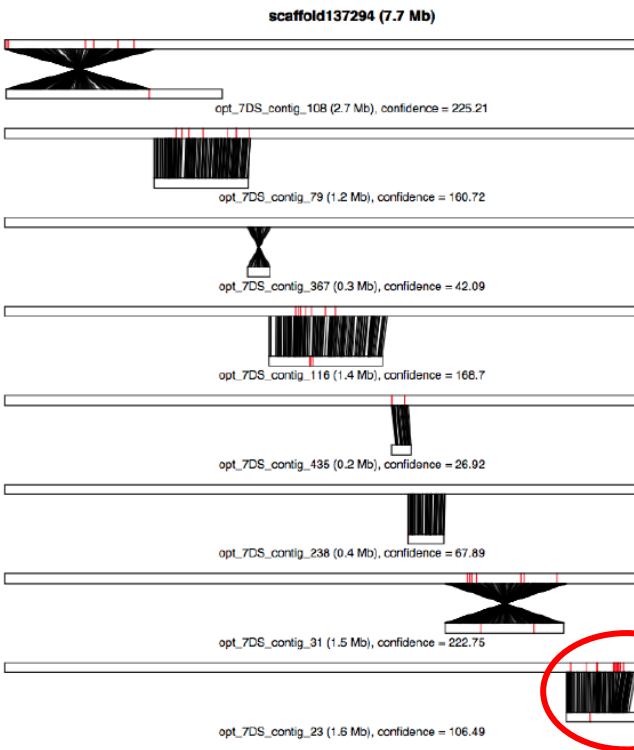
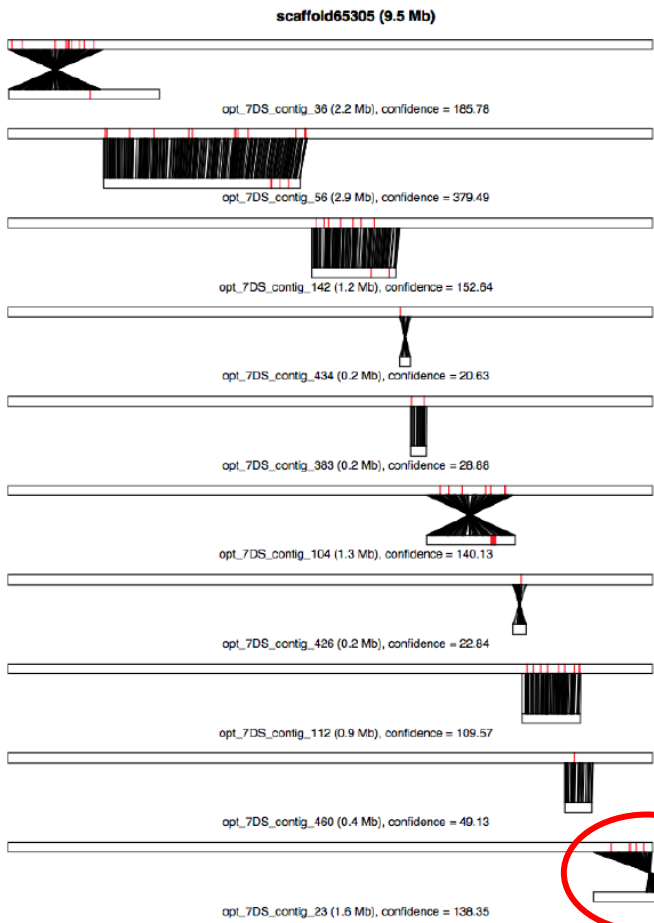
Stankova et al., Plant Biotechnology Journal (2015), doi: 10.1111/pbi.12513

Hi-C Position: rank 69, 81.8 cM

rank 70, 82.0 cM



Martin Mascher, IPK Gatersleben



Gap size: 8.8 kb with no labels between adjacent alignments

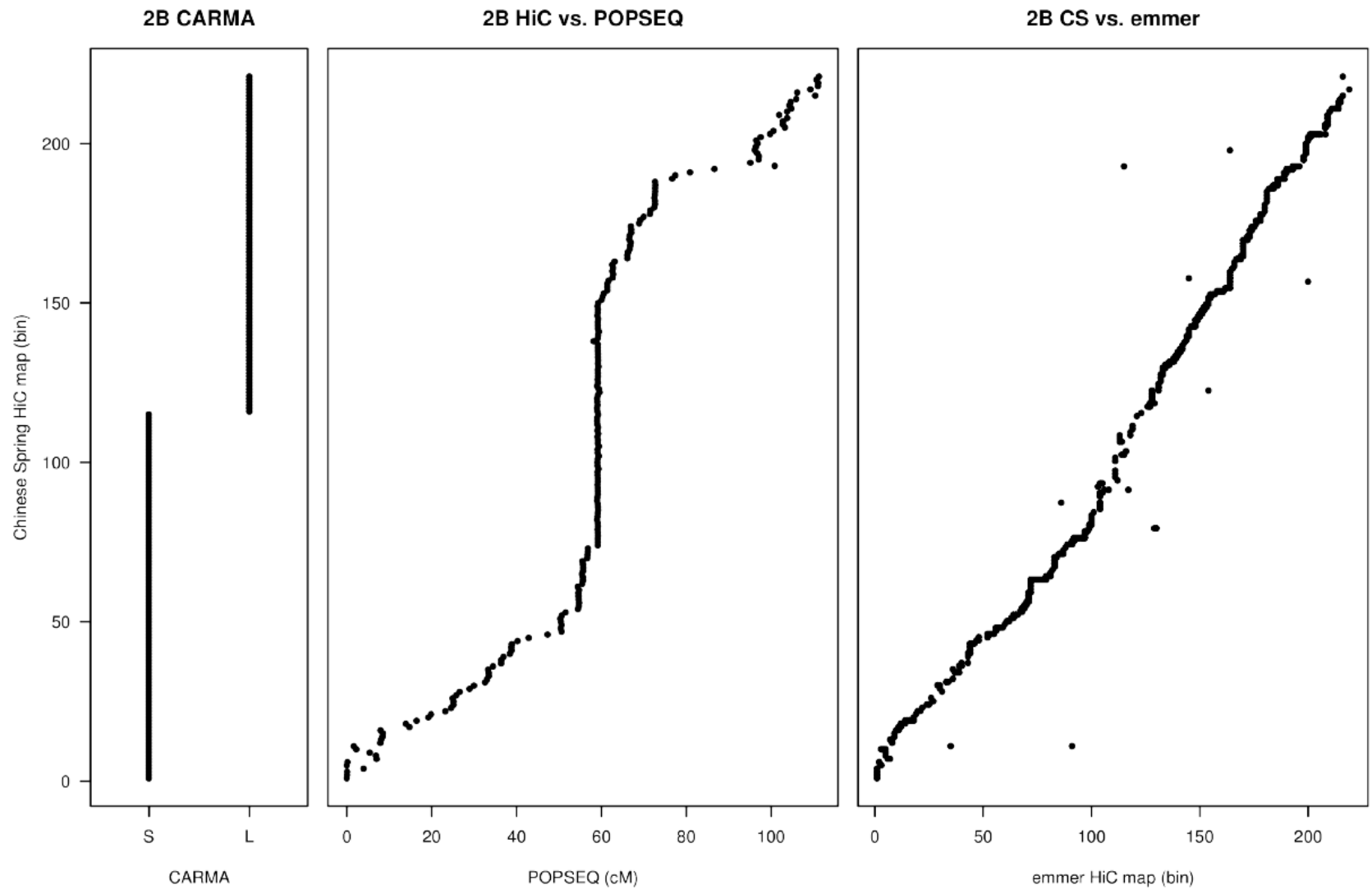


Martin Mascher,
IPK Gatersleben

Linking to 7DS bionano map

- 7DS bionano map: 339/346 high quality maps aligned to scaffolds with >95% of their length (cumulative length of 339 Mbp)
- Un-aligned maps: short or low bionano coverage
- 460 7DS NRGene scaffolds (345.5 Mbp): 335.4 Mbp aligned to bionano maps

Chromosome arm assignment and Hi-C



Summary of initial QA of IWGSC WGA

The WGA looks impressingly complete:

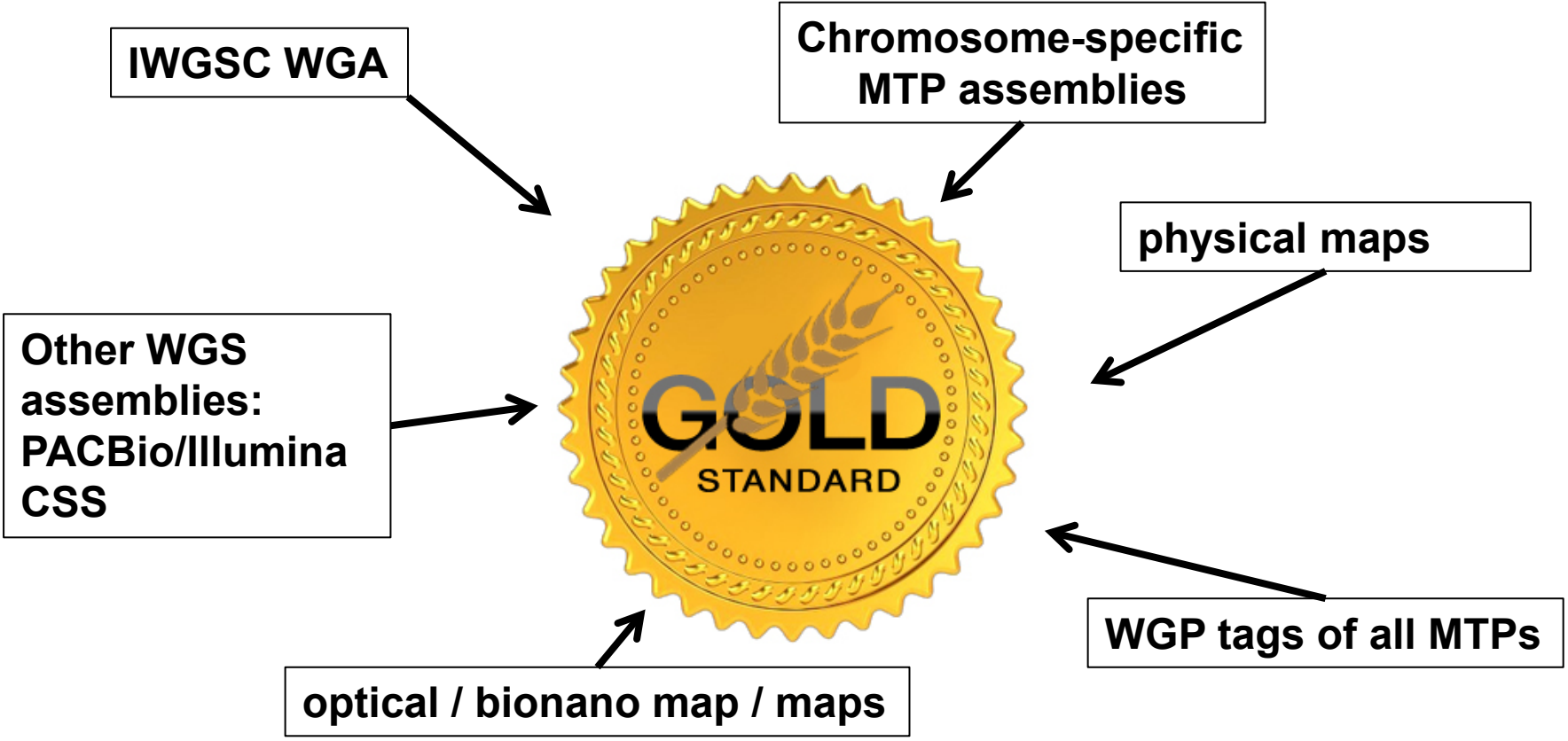
- > 95 % representation of map-based reference sequence of 3B and the physical map of 6A.
- 98 % of fl-cDNA present
- 98 % of CSS coding exons found
- >99% of 1B ISBPs present

But:

- A number of chimeric scaffolds
- Only 14.5 Gbp ▶ Is the CS genome actually smaller than expected?

Some further curation and integration with other resources needed!

How to move forward? Concerted integration of resources!



A reference sequence anchored to the genetic and phenotypic maps in 2016

Thanks to Funders of the WGA project!



With support from



by decision of the German Bundestag



