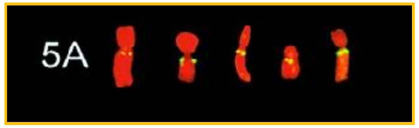
AN INTEGRATED APPROACH FOR THE PHYSICAL MAPPING OF WHEAT CHROMOSOME 5A

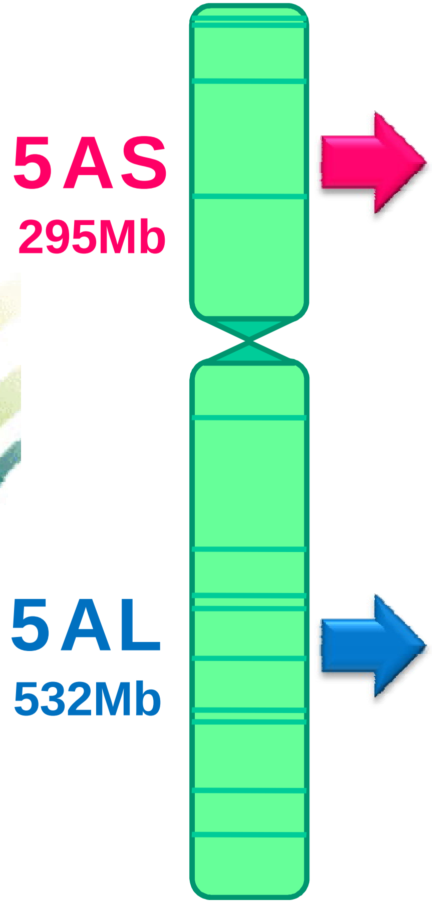
GIAMPIERO VALÈ
CRA -GENOMICS RESEARCH CENTRE
ITALY
HTTP://CENTRODIGENOMICA.ENTECRA.IT/

CRA
CONSIGLIO PER LA RICERCA
E LA SPERIMENTAZIONE
IN AGRICOLTURA
Genomics Research Centre - Fiorenzuola

Chromosome-arm specific BAC libraries

Two BAC libraries prepared from flow sorted DNA





5AS
295Mb

5AS BAC library ”TaaCsp5AShA”

46.080 clones (120 plates)

mean insert size = 120kb

5AS coverage= 16.5X

5AL
532Mb

5AL BAC library ”TaaCsp5ALhA”

90.240 clones (235 plates)

mean insert size = 123kb

5AL coverage= 18.3X

BAC Library Fingerprinting and Contig Assembly

BAC fingerprinting

✓ 44,740 clones of 5AS and 51,072 clones of 5AL were fingerprinted using SnaPshot method.
✓ 75,995 high quality fingerprints after contamination removal.

Useful fingerprints were assembled using first FPC according to IWGSC rules: a MTP (MTP-FPC) was established after automated assembly

✓ Several successive cutoffs were used from 1e-75 to 1e-25
✓ 5AS: 1,308 contigs, 4,201 clones, 342 Mb
✓ 5AL: 2,556 contigs, 6,560 clones, 601 Mb

Successively, the same useful fingerprints were also assembled using LTC according to Frenkel et al. 2010 guidelines and a second MTP (MTP-LTC) was established after automated assembly

✓ 5AS: 652 contigs, 5,412 clones, 330 Mb
✓ 5AL: 1,504 contigs, 8,709 clones, 676 Mb

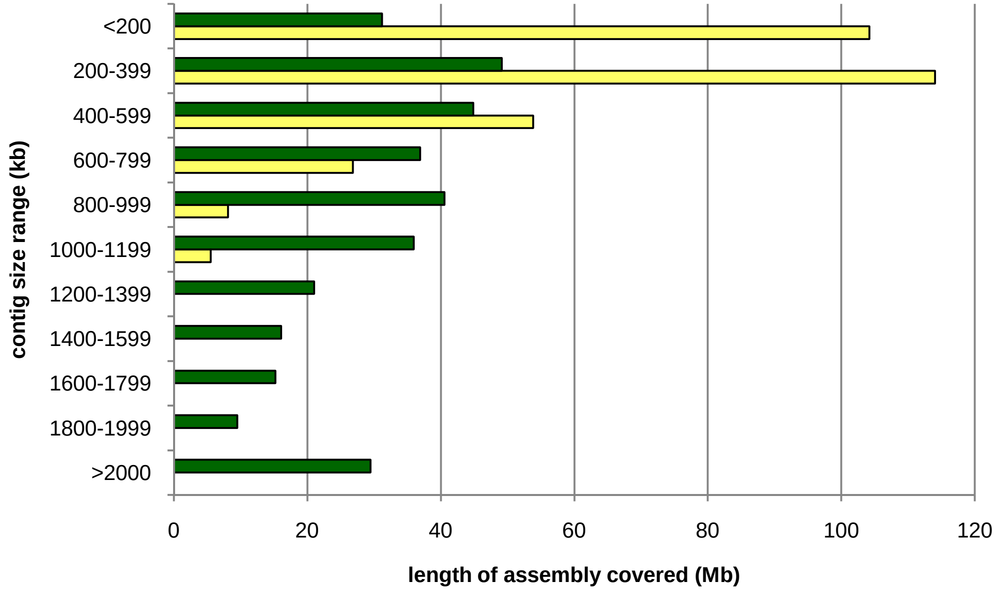
BAC Library Fingerprinting and Contig Assembly

	5AS		**5AL**	
Chromosome arm size (Mb)	295		532	
BAC library size (clones)	46,080		90,240	
Average insert size (kb)	120		123	
Library clone depth	16.5x		10.4x	
Fingerprinting technique	SNaPShot HICF		SNaPShot HICF	
# fingerprinted clones	44,740		51,072	
Assembly method	**FPC**	**LTC**	**FPC**	**LTC**
Assembly stringency	1,00E-45	1,00E-20	1,00E-45	1,00E-20
# useful fingerprints	36,165 (80,8%_13.2 equivalents)		39,830 (76%_8.1 equivalents)	
# clones in assembly	36,165	36,165	39,830	39,830
# clones in contigs	25,084	26,659	27,764	29,610
# singletons	11,081	9,506	12,066	10,220
# MTP contigs	1,308	652	2,556	1,504
# MTP clones	4,201	5,412	6,560	8,709
Estimated chromosome arm coverage (Mb, %)	342(116%)	330 (112%)	601 (113%)	678 (127%)
Contig N50	354	128	823	407
Contig L50 (kb)	296	820	251	563
# contigs ≥ 3 clones in the assembly	1,062	625	2,166	1,453
Average contig size (kb)	265	522	228	462
Longest contig (kb)	1,297	3,391	1,027	2,303

LTC *vs* FPC assembly

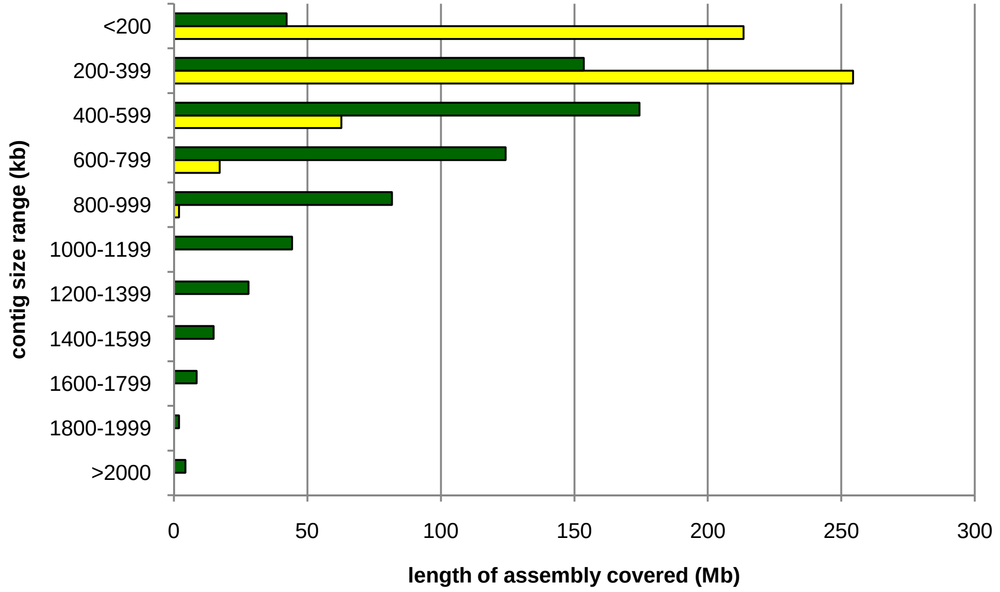
🟩 **LTC**
🟨 **FPC**

5AS
295Mb



more than 90 Mb, about a third of 5AS, were covered by contigs longer than 1.2 Mb

5AL
532Mb



more than 100 Mb, about a fifth of 5AL, were covered by contigs longer that 1 Mb

3D Pools of 5A MTP (FPC- and LTC-)

MTP-FPC

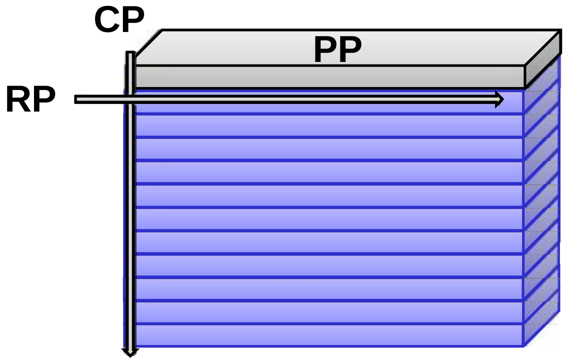
10,761 clones

5AS **5AL**

4,201 + 6,560

3D-pooling of DNAs

51 pools 58 pools



MTP-LTC

14,121 clones

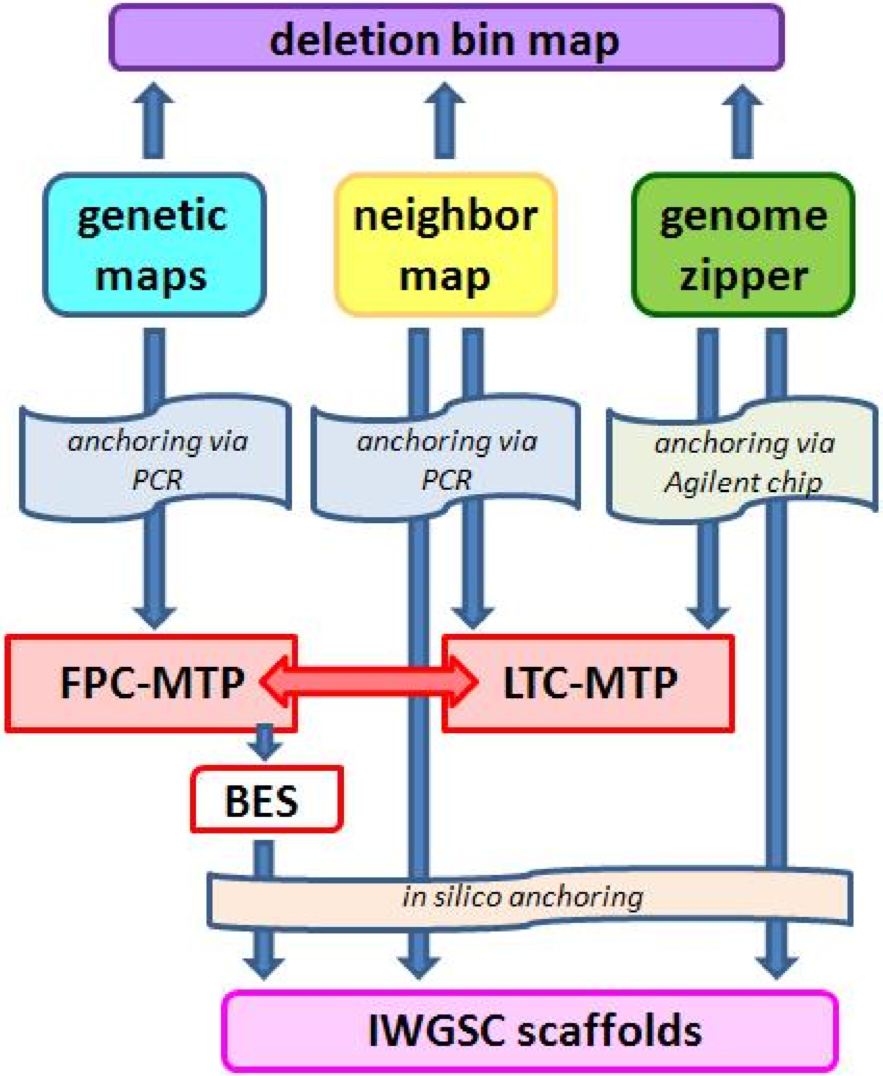
5AS **5AL**

5,412 + 8,709

3D-pooling of BACs

55 pools 63 pools

Anchoring: sources and strategies



Anchoring *via* PCR

✓ genetic maps
✓ neighbor map
✓ deletion bin map

Anchoring *via* array

✓ genome zipper

In silico anchoring

✓ BES from FPC-MTP on IWGSC Illumina scaffolds

Genetic maps

4 Segregating populations

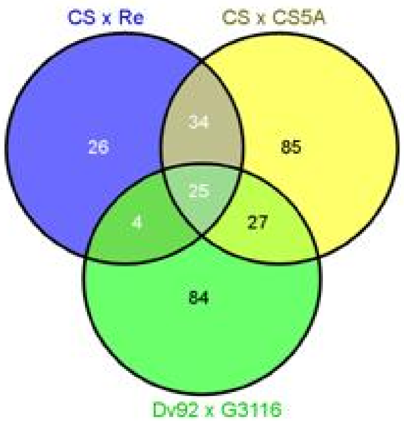
- 383 F2 Chinese Spring (CS) x Renan (*T. aestivum*)
- 188 RILs from CS x CS-*T. dicoccoides* Disomic Substitution 5A (*T. aestivum x T. turgidum dicoccoides*)
- 132 RILs from DV92 x G3116 (*T. monococcum*)
- 124 RILs from Latino x MG5323 (*T. turgidum ssp durum x T. turgidum ssp dicoccum*)

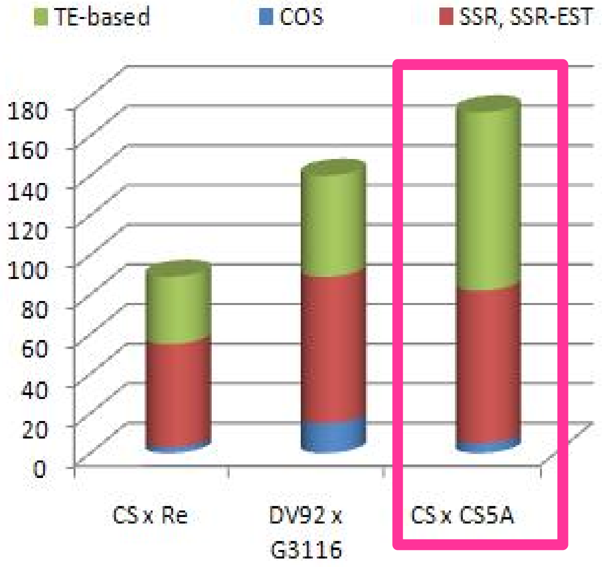
Several classes of molecular markers

- TE junction-based markers (ISBP, RJM, RJJM) from survey sequencing of 5AS
- SSRs from survey sequencing of 5AS
- SSRs and SSR-ESTs from literature
- COSs

299 polymorphic and specific markers

	LG	length	density
CS x Re	2	159.3	0.29
CS x CS5A	2	138.5	0.41
Dv x G	5	187	0.28
Lt x MG	3	116.7	0.42





Anchoring of FPC-MTPs and integration with LTC

Screening of FPC-MTP 3D pools using PCR
with all available markers
(SSRs, COS and TE-derived)

anchoring of contigs using a dedicated scripting

178 markers anchored
(positive for at least one BAC)

168 contigs anchored by at least one marker

total length anchored:

43.2 Mb of 5AS and 12.6 Mb of 5AL

FPC ASSEMBLY	5AS	5AL
anchored markers	130	48
EST and COS	9	10
SSR and SSR-EST	33	37
STS	1	1
TE-based	87	-
FPC unique contigs	122	46
total length anchored (Mb)	43.2	12.6
total length anchored (Mb) after LTC integration	99.8	21.3
corresponding total length anchored (Mb) by LTC	106.5	27.3

Neighbor map

Map used as a scaffold

CS x CS5A 90K map (*Gadaleta et al, submitted*; 572 markers)

Map used for the integration of CSxCS5A

Latino x MG5323 90K map (*Desiderio et al, submitted*; 593 markers)
90K Illumina consensus map (*Wang et al, 2014*; 1,605 markers)
CS x Re map
SSR consensus map (*Somers et al, 2004*)
Nanda2419 x Wangshuibai (*Xue et al, 2008*)
Arina x Forno (*Paillard et al, 2003*)

neighbor map

total length: 248.7 cM

3 linkage groups

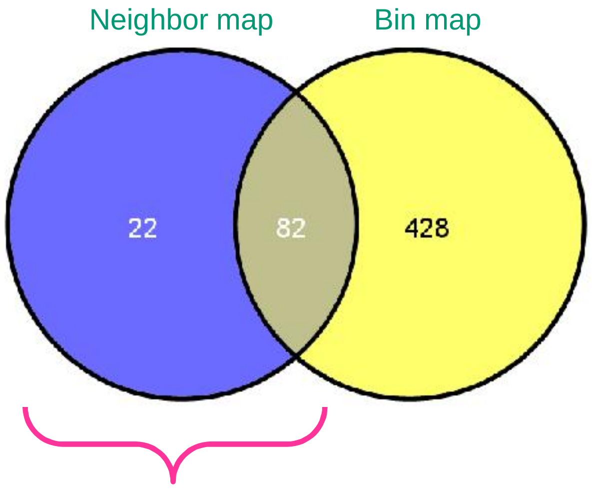
2,681 markers included

2,501	SNPs
122	SSRs
26	TE-based
18	STSs
14	ESTs

average density: 10.7 markers/cM

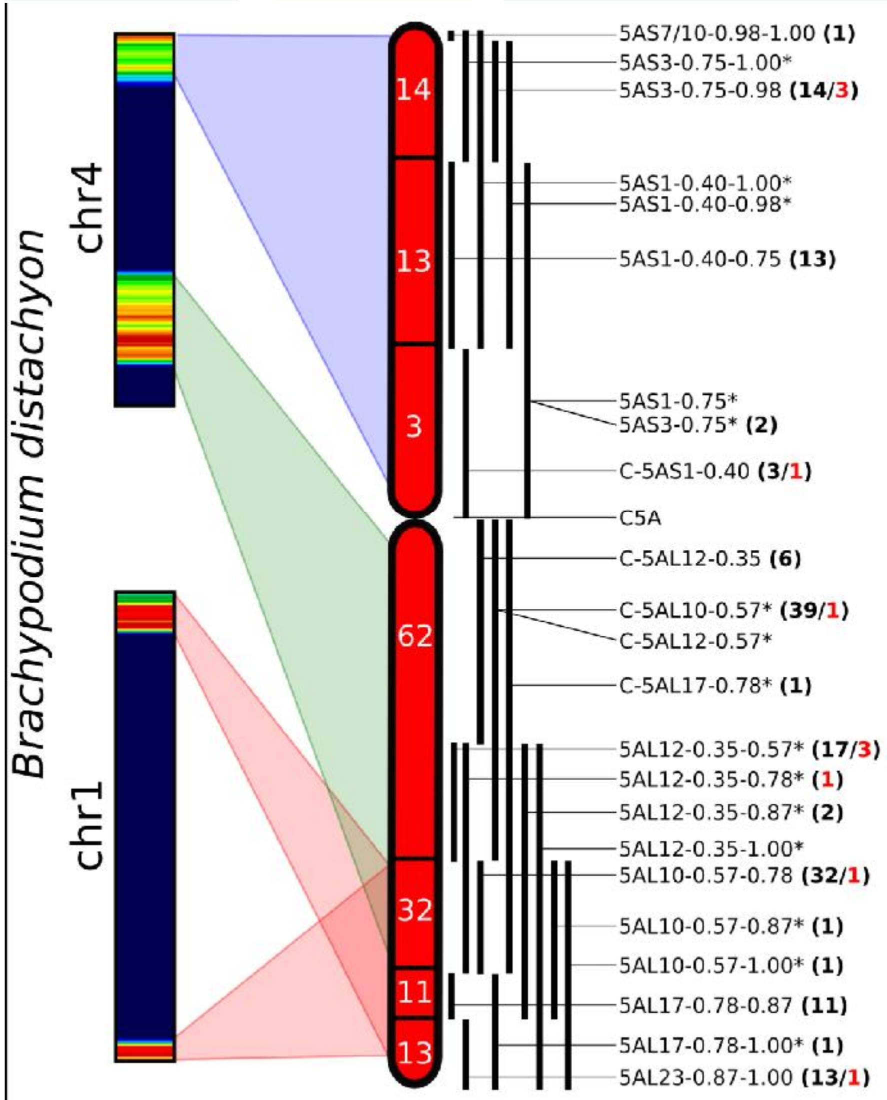
Anchoring via PCR on LTC-MTP 3D pools

deletion bin map: 511 markers



55.1 Mb 5AS
30 Mb 5AL

Anchoring of LTC-MTPs



Reference Genome Zipper for 5A

Vitulo et al, PlosOne 2011

A15K Agilent custom microarray was designed
<u>**4,722 sequences**</u>
deriving from several sources:

- Genome Zipper reads
- 5A ESTs from GrainGenes database
- RFLPs, SNPs, SSRs, TE-based junction markers

12,676 probes have been developed and 118 3D pools have been hybridized

5AS
138 different reads assigned to 89 ctgs corresponding to 58.3 Mb

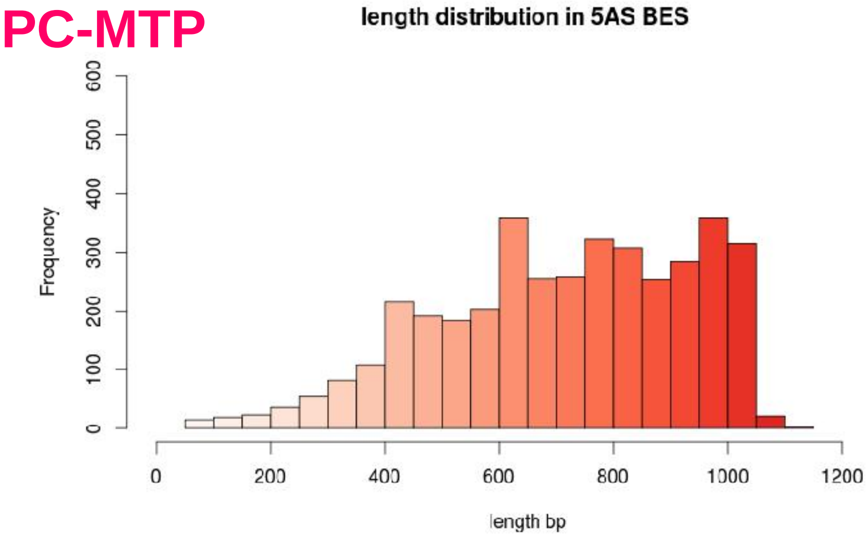
5AL
283 different reads assigned to 443 ctgs corresponding to 125.3 Mb

FPC-MTP BAC End Sequencing

5AS-FPC-MTP

2,487 clones recovered
3,861 good quality reads (78%)
2.8 Mb – GC content 44.3%
average length 715 bp

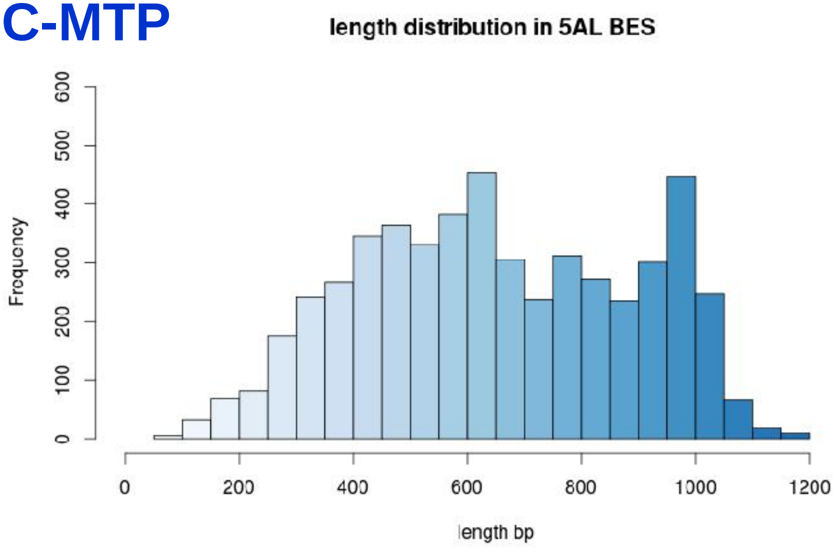
1,814 ISBP markers
72 SSRs (di- tri- tetra-nucleotide)



5AL-FPC-MTP

2,835 clones recovered
5,194 good quality reads (92%)
3.4 Mb – GC content 43.8%
average length 650 bp

2,117 ISBP markers
52 SSRs (di- tri- tetra-nucleotide)



1,964 ISBPs (considering only one marker/sequence + 124 SSRs → Ready for RH panel

A radiation hybrid panel for 5A

Chinese Spring
pollen

X

nulli5Atetra5**B**
nulli5Atetra5**D**

Approx. 390 crosses performed for each nullisomic line

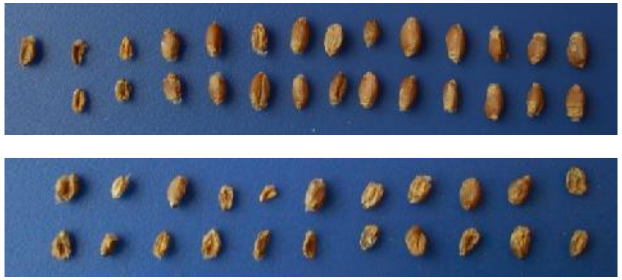
Nulli5Atetra5B → 3500 seeds
Nulli5Atetra5D → 5600 seeds

Nulli5Atetra5 ⊗

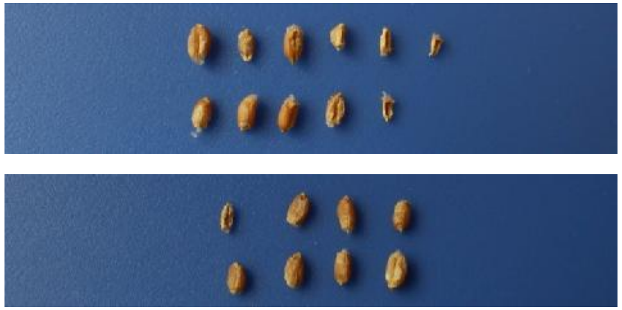
Nulli5Atetra5 ⊗

Nulli5Atetra5B x CS
Nulli5Atetra5D x CS

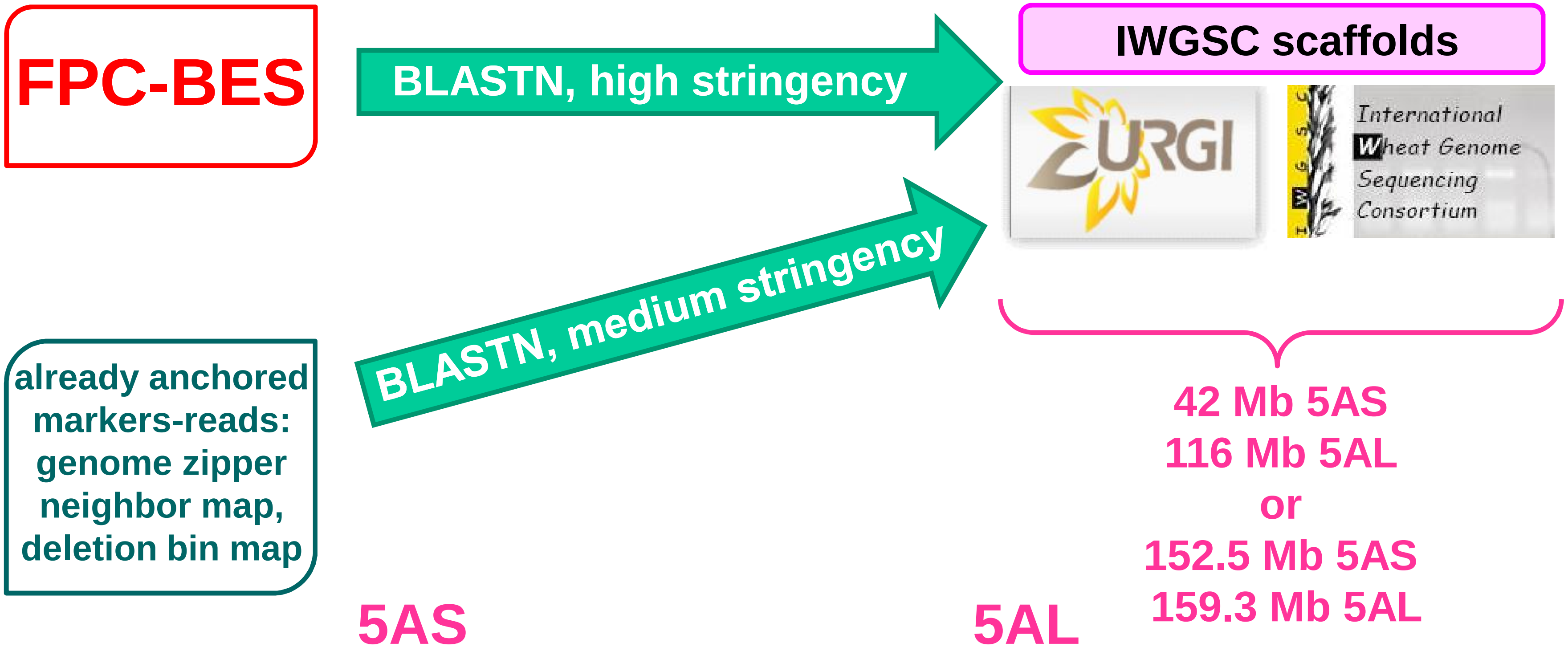
Nulli5Atetra5B
x
CS 1 Kr



Nulli5Atetra5D
x
CS 1 Kr



In silico anchoring



FPC-BES

BLASTN, high stringency

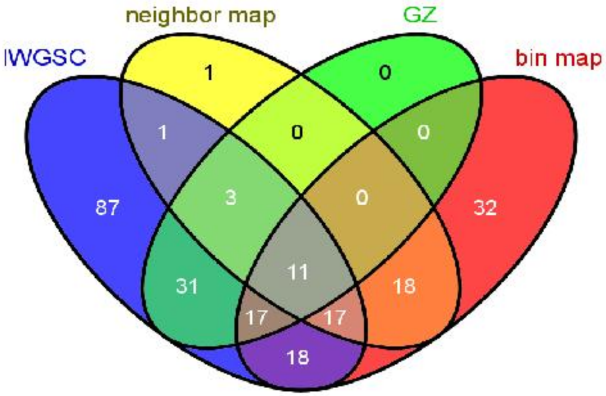
IWGSC scaffolds

already anchored markers-reads:
genome zipper
neighbor map,
deletion bin map

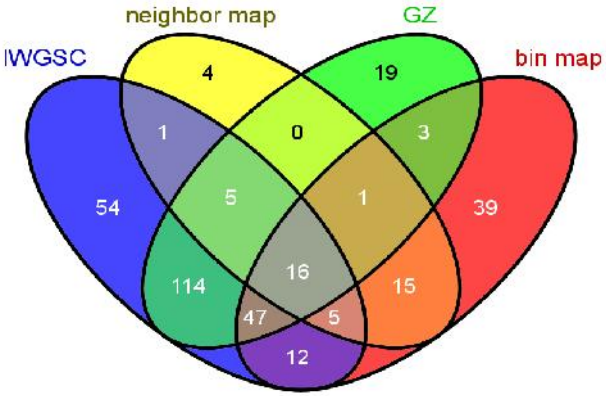
BLASTN, medium stringency

42 Mb 5AS
116 Mb 5AL
or
152.5 Mb 5AS
159.3 Mb 5AL

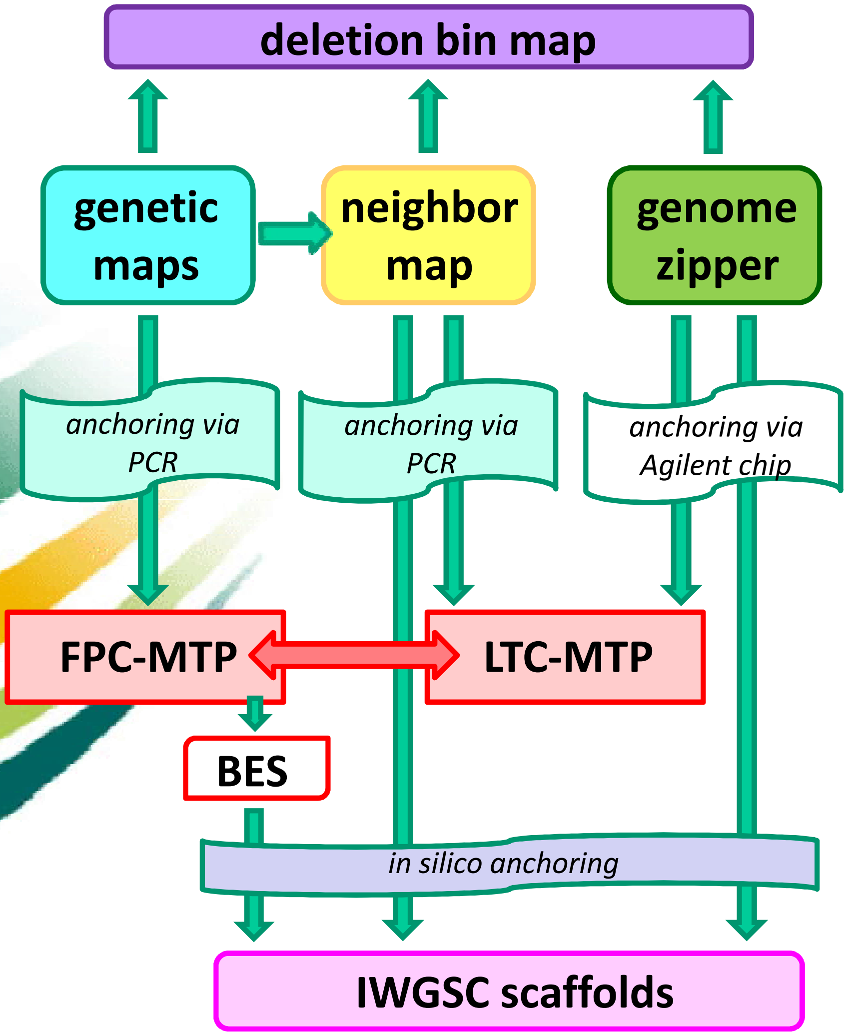
5AS

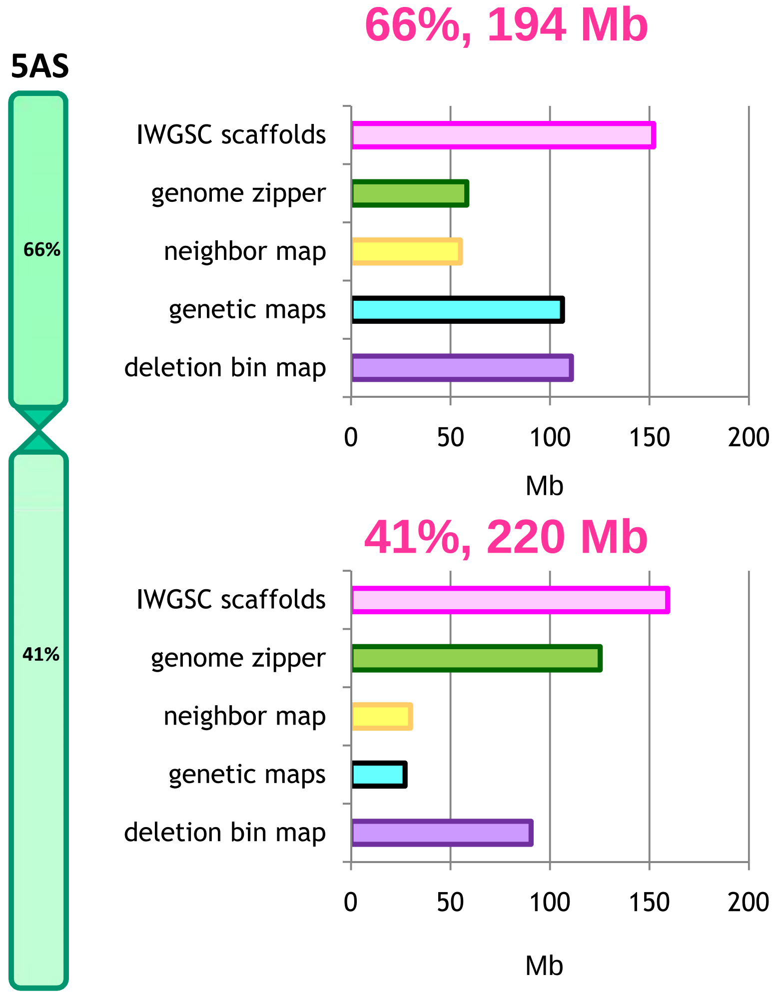


5AL



Summarizing..





Further improvement from Physical Scaffolding

⇩

manual contig elongation and end to end merging using LTC assembly

Collaborations..

***Yellow rust* (5AL)**

with Jorge Dubcovsky
University of California,
Davis, CA, USA

***Fusarium* *headblight, FHB* QTL (5AS)**

with Hermann Buerstmayr
Institute for Biotechnology in Plant Production,
Tulln, AUSTRIA

***Yellow mosaic virus (WYMV)* (5AL)**

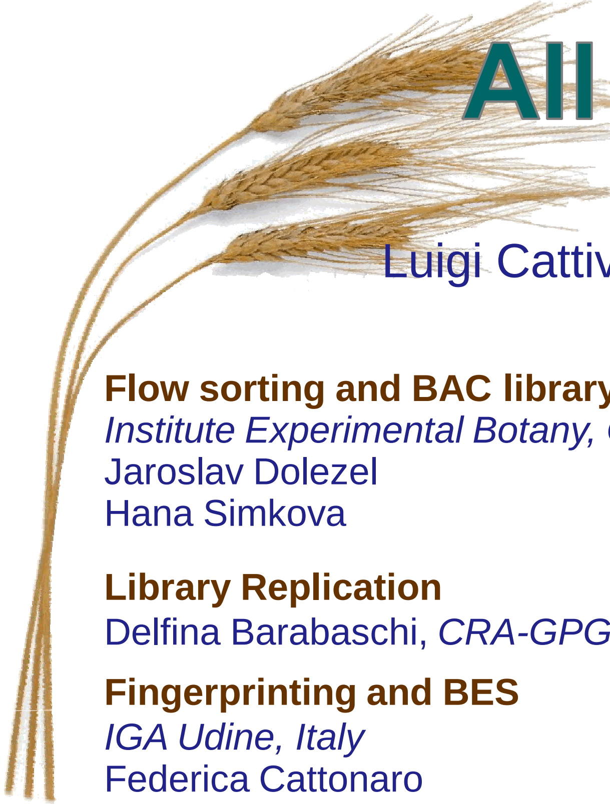
with Xiue Wang
Nanjing Agricultural University,
Nanjing, CHINA

5AL-4AL breakpoints

with Chunji Liu
CSIRO Plant Industry, Queensland
Bioscience Precinct, Australia

Oxalate oxidase (5AS)

with Renato D’Ovidio
Universita’ degli Studi della Tuscia
Viterbo, ITALY



All people involved

Coordinators (*CRA-GPG, Italy*)

Luigi Cattivelli, Antonio Michele Stanca, Giampiero Valè

Flow sorting and BAC library preparation

Institute Experimental Botany, Olomouc, Czech Republic
Jaroslav Dolezel
Hana Simkova

Library Replication

Delfina Barabaschi, *CRA-GPG*

Fingerprinting and BES

IGA Udine, Italy
Federica Cattonaro
Federica Magni
Simone Scalabrin
Michele Morgante

Genetic Mapping

Delfina Barabaschi, *CRA-GPG*
Andrea Volante, *CRA-GPG*
Katia Lacrima, *CRA-GPG*
Vania Michelotti, *CRA-GPG*
Luigi Orrù, *CRA-GPG*
Francesca Desiderio, *CRA-GPG*
Enrico Francia, *UniMore, Reggio Emilia, Italy*
Agostino Fricano, *PTP, Lodi Italy*
Annamaria Mastrangelo, *CRA-CER, Foggia, Italy*

Radiation Hybrid panel

Andrea Volante, *CRA-GPG*
Ajay Kumar, *NDSU, USA*

Citogentic Mapping

UniBari, Italy
Agata Gadaleta
Antonio Blanco

Anchoring Genetic to Physical

Delfina Barabaschi, *CRA-GPG*
Andrea Volante, *CRA-GPG*
Lucia Prazzoli, *CRA-GPG*
Paola Tononi, UniVerona, Italy
Massimo Delledonne, UniVerona, Italy

Survey Sequencing and GZ development

CRIBI Padova
Giorgio Valle
Nicola Vitulo

Bioinformatic Support

CRA-GPG
Paolo Bagnaresi