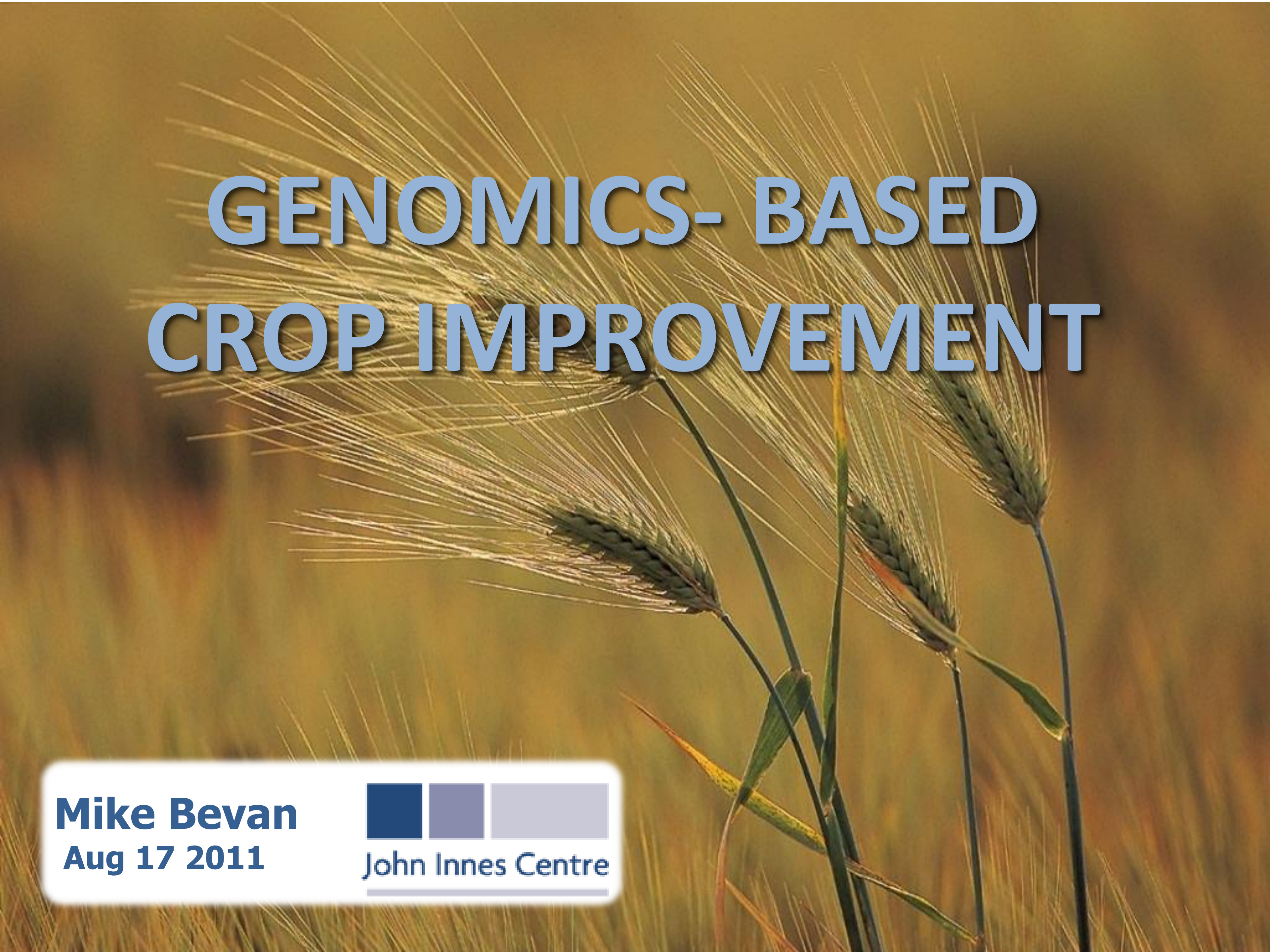
GENOMICS- BASED CROP IMPROVEMENT



Mike Bevan
Aug 17 2011

John Innes Centre

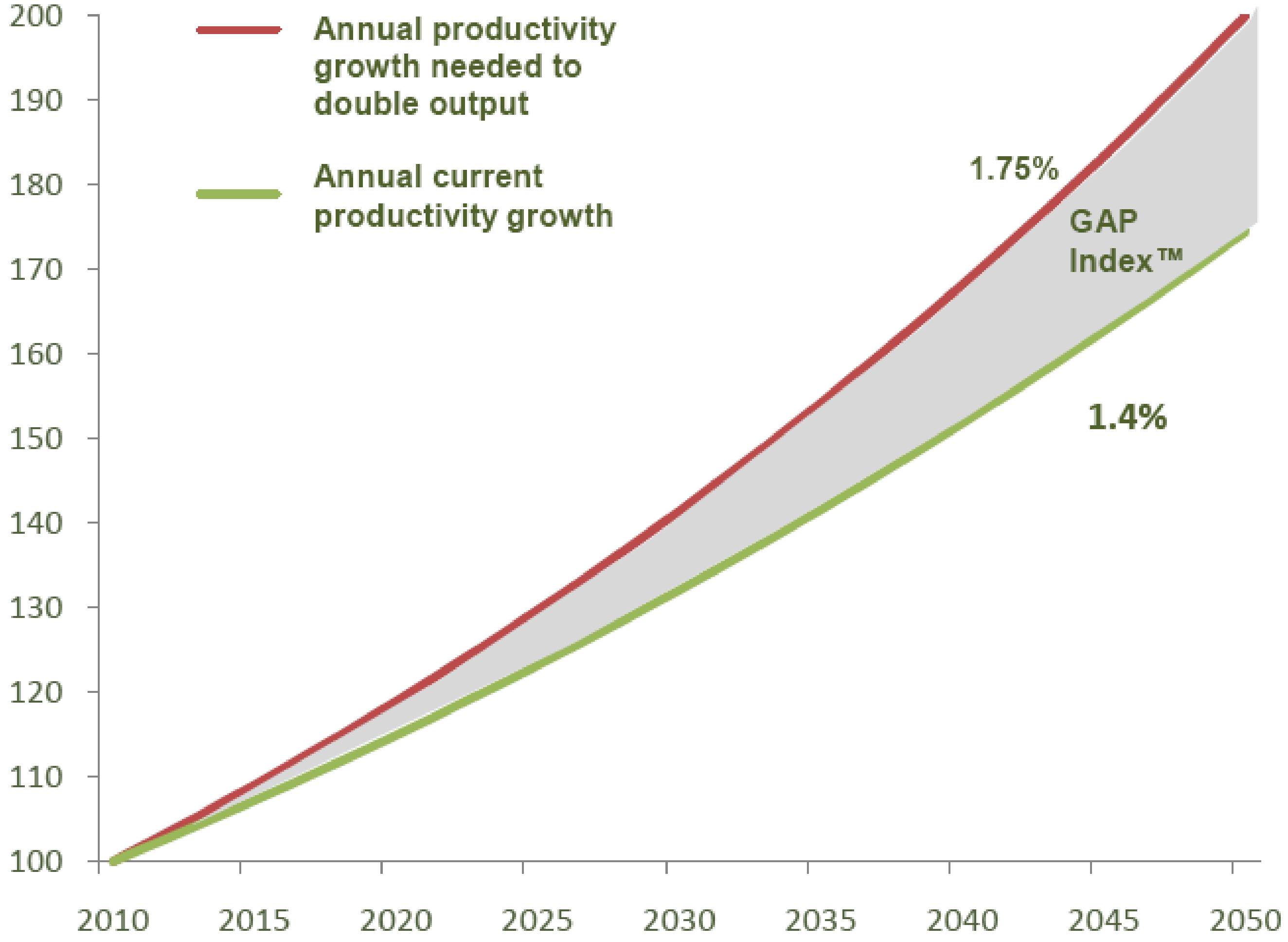
Food and Energy Security- a common Research Agenda

- The challenge of increasing crop production in a sustainable framework
- The need to domesticate new crops, eg for industrial applications
- Understanding complex traits: eg yield, yield stability and quality traits
- Working in a tight framework of time and resources: economic, population and environmental drivers.

The Global Harvest Initiative

2010 GAP Report™

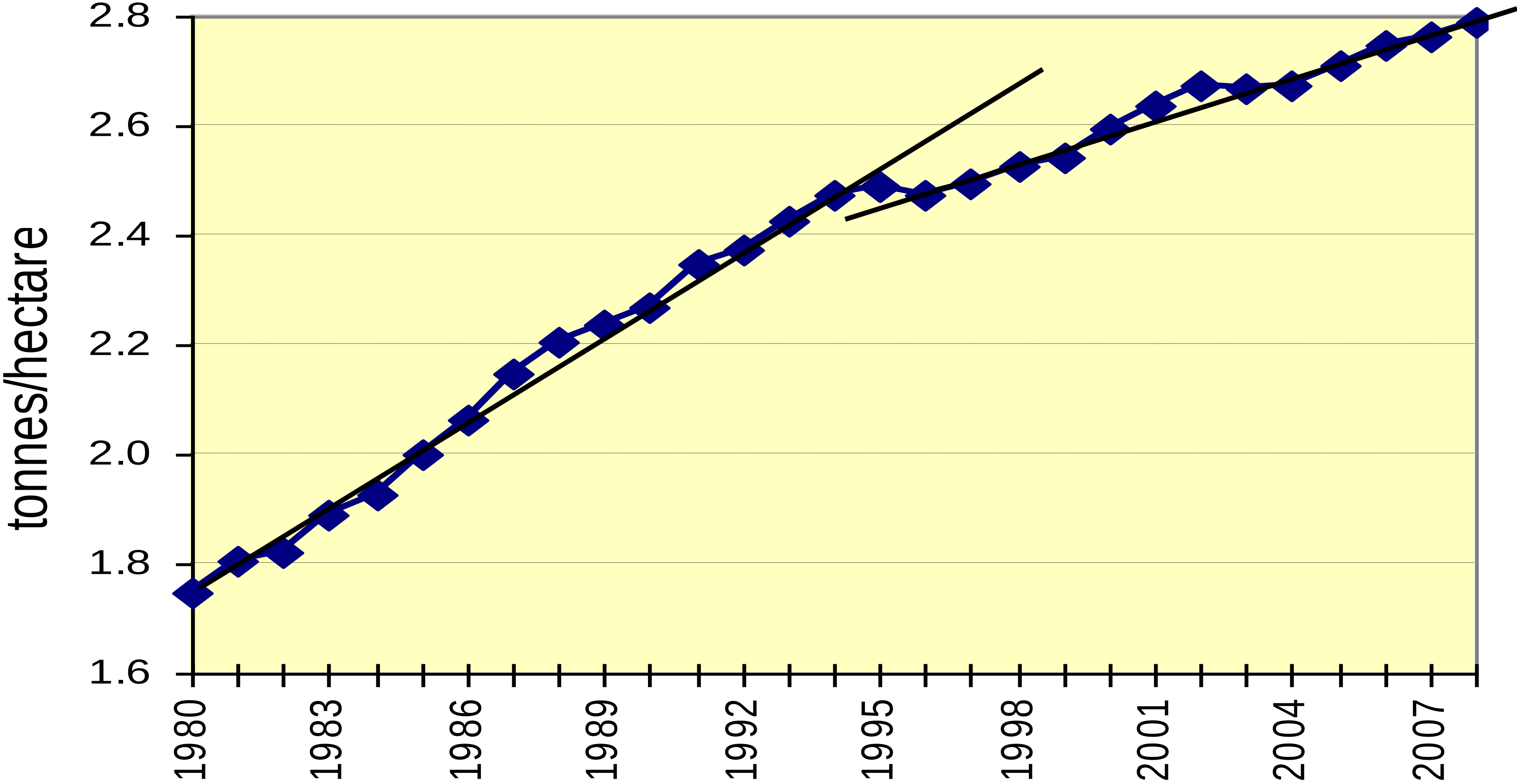
Measuring Global Agricultural Productivity



Source: Farm Foundation, NFP calculations (2010) based on USDA ERS data.

WORLD WHEAT YIELD TRENDS

(5 Year Moving Average - tonnes/ha)



Source: Chudleigh, 2008

Crop Genomics Research Challenges

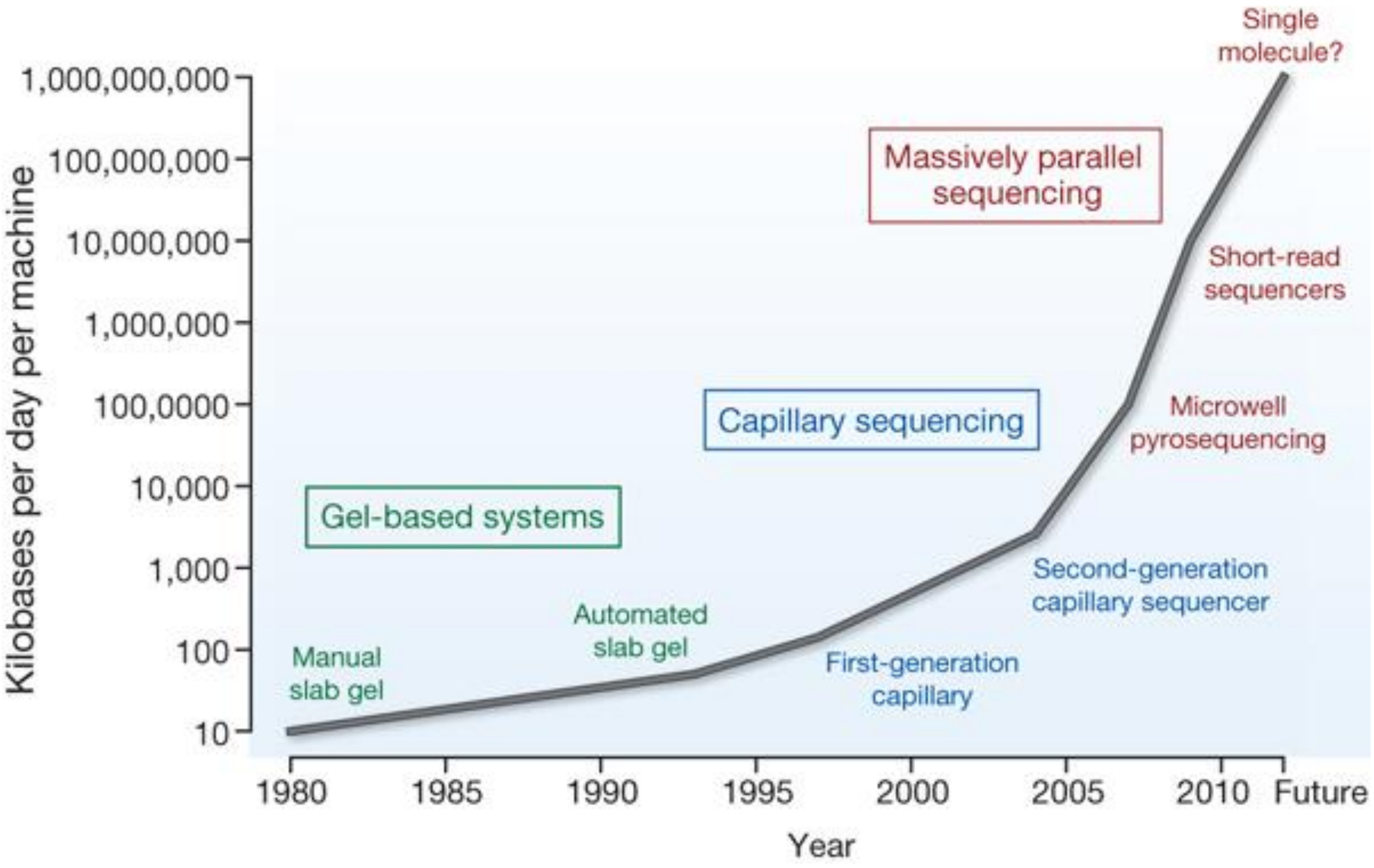
Germplasm Improvement

- Domestication of new species
- Introduction of wild germplasm into elite lines
- Creating new polyploids

Trait analysis

- Understanding and improving complex traits- such as yield
- Controlling fertility and new breeding systems- apomixis, wide crossing
- Achieving improved sustainable production- nutrient and water use.

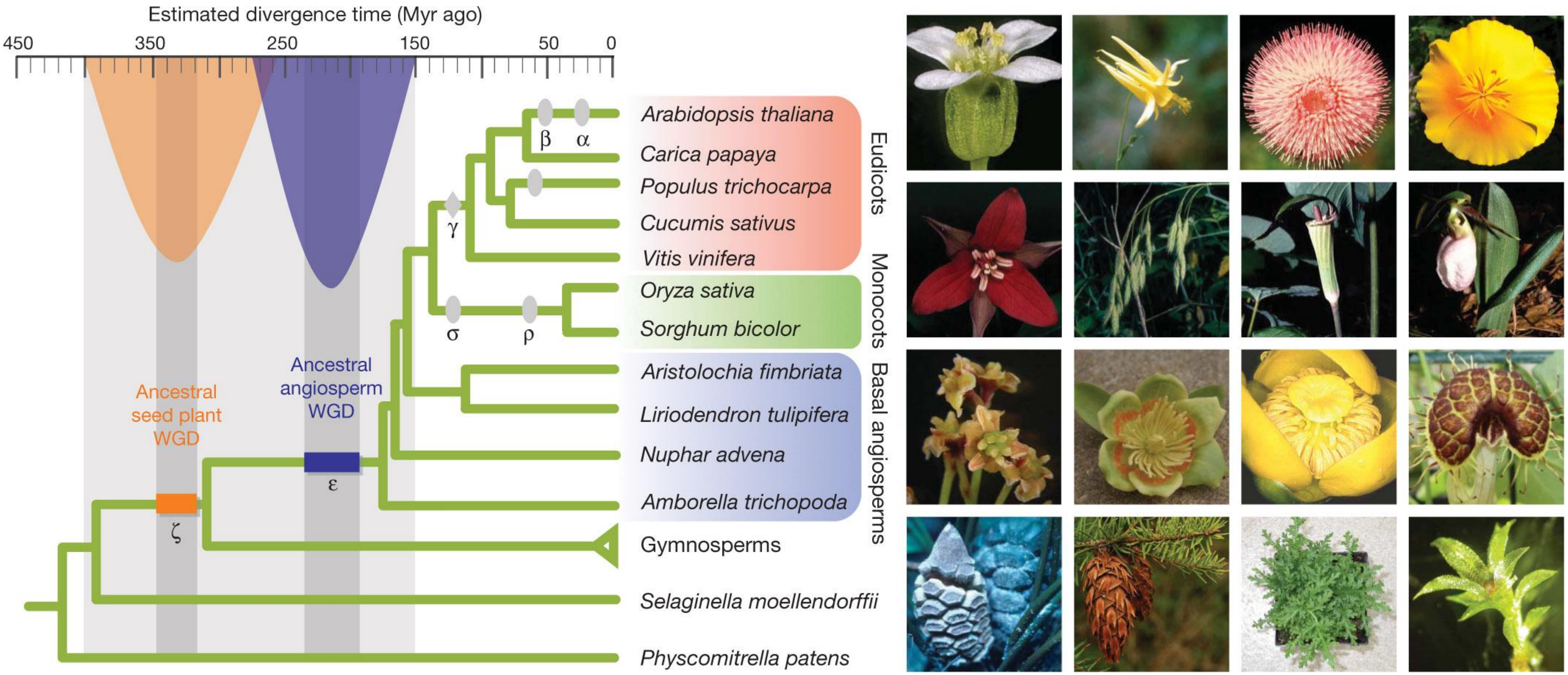
Increases in the rate of DNA sequencing underpin crop improvement



MR Stratton *et al.* *Nature* **458**, 719-724 (2009) doi:10.1038/nature07943

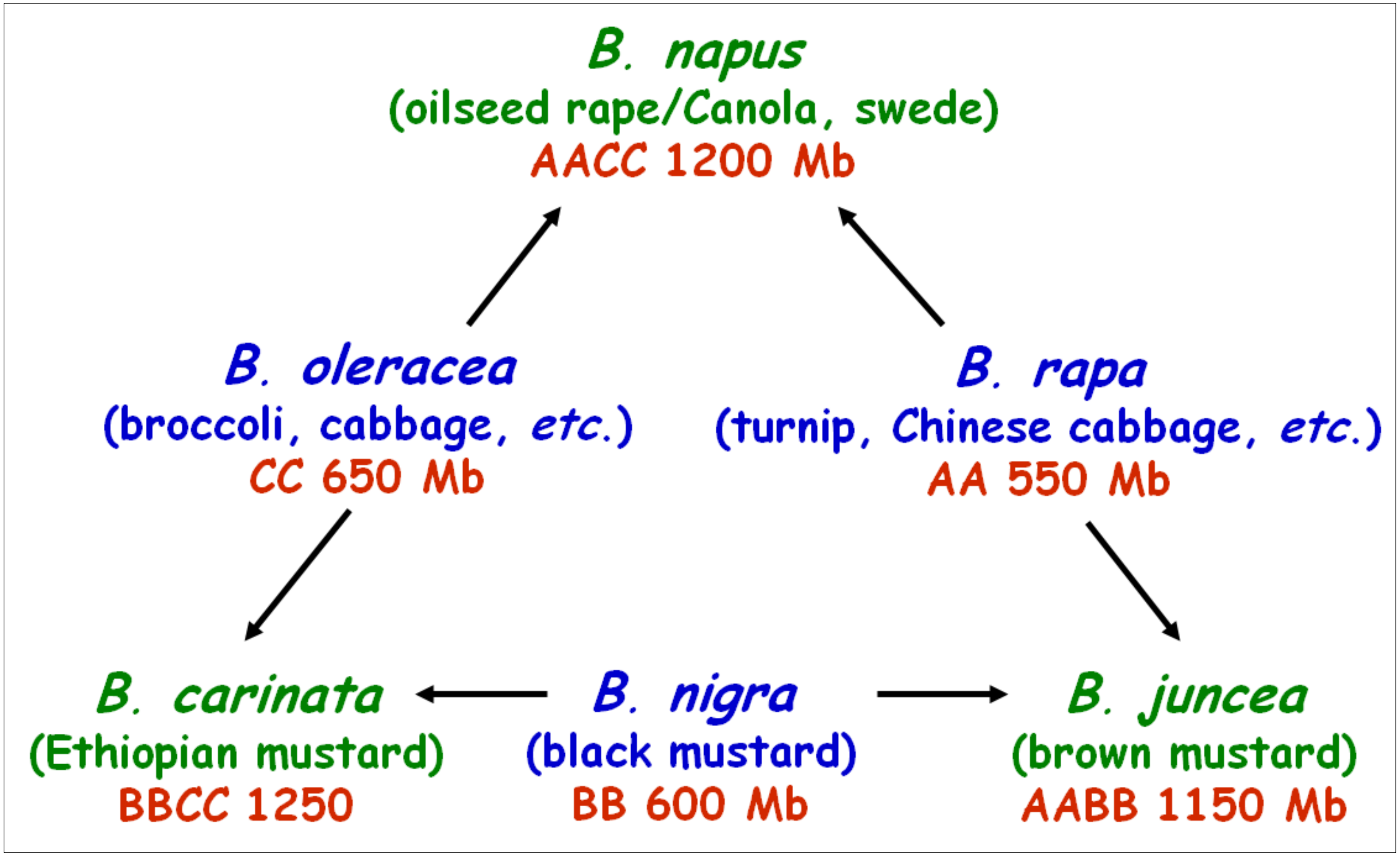
Ancestral genome duplications during land plant evolution

The evolution of grass genomes has involved 4 whole genome duplications



Jiao et al and dePamphilis Nature 2011

GENOME RELATIONSHIPS BETWEEN BRASSICA SPECIES
"U's TRIANGLE"



B. napus
(oilseed rape/Canola, swede)
AACC 1200 Mb

B. oleracea
(broccoli, cabbage, *etc.*)
CC 650 Mb

B. rapa
(turnip, Chinese cabbage, *etc.*)
AA 550 Mb

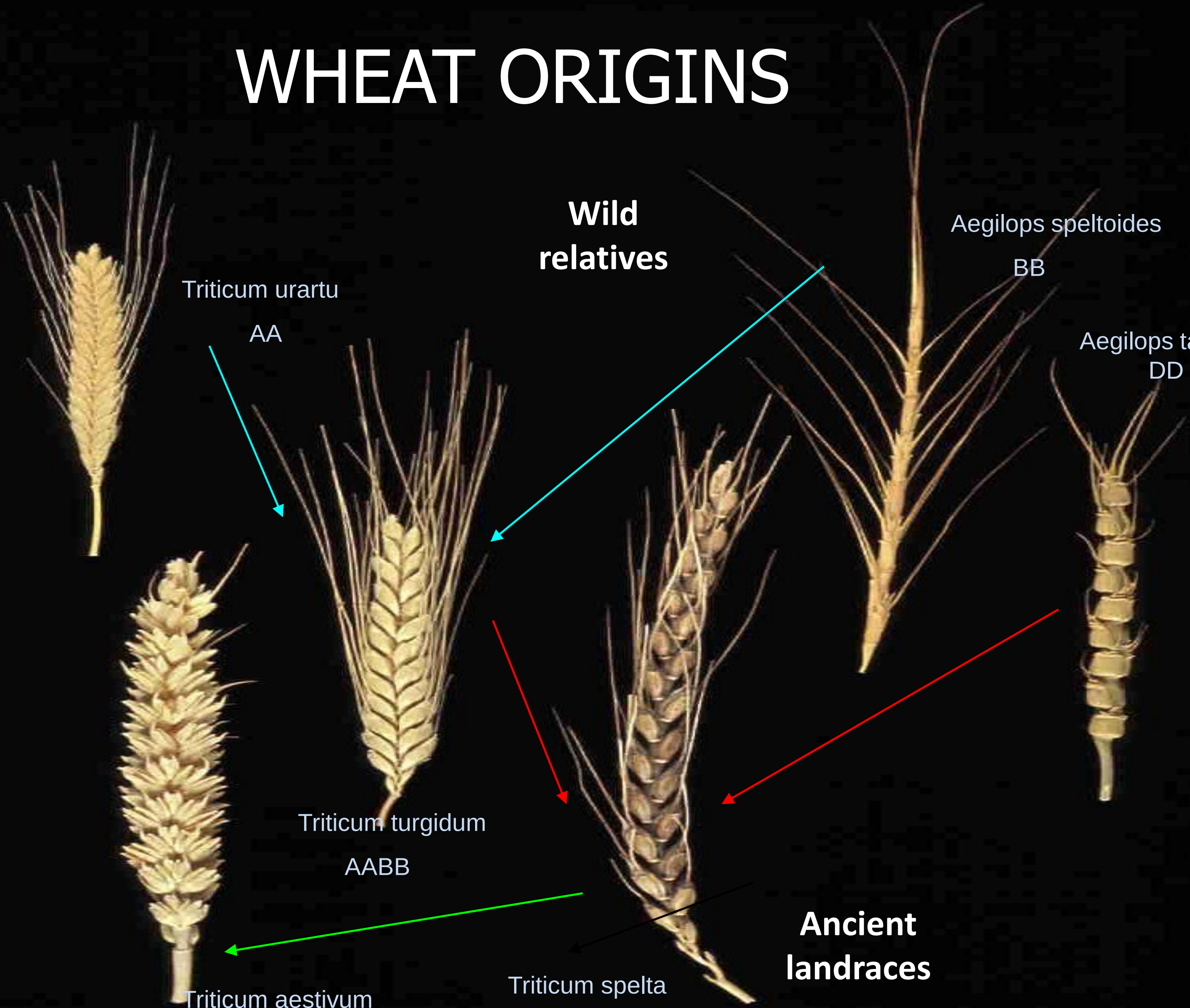
B. carinata
(Ethiopian mustard)
BBCC 1250

B. nigra
(black mustard)
BB 600 Mb

B. juncea
(brown mustard)
AABB 1150 Mb

IAN BANCROFT JIC

WHEAT ORIGINS



Wild relatives

Triticum urartu

AA

Aegilops speltoides

BB

Aegilops tauschii
DD

Triticum turgidum

AABB

Ancient landraces

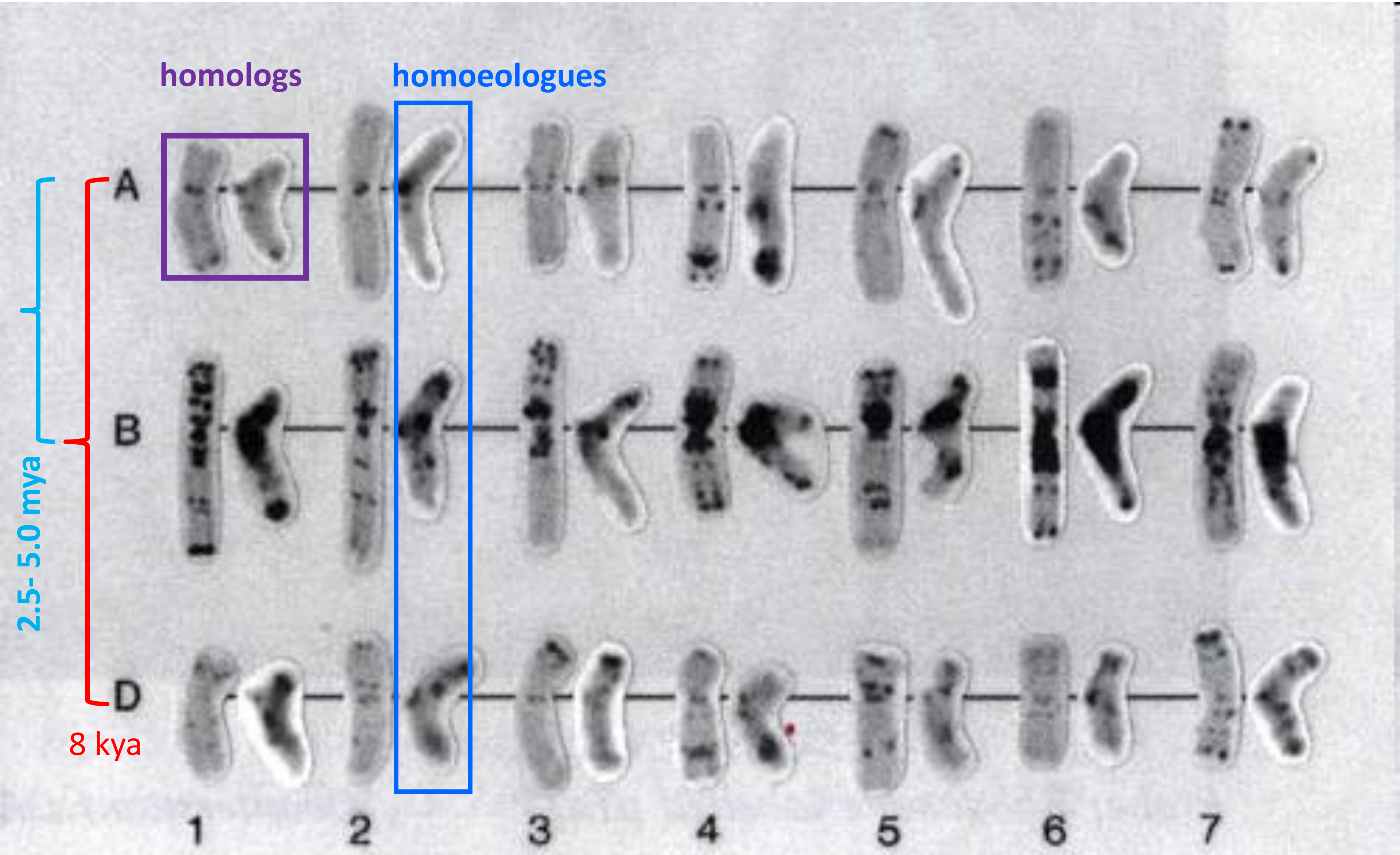
Triticum spelta

AABBDD

Triticum aestivum
AABBDD

Modern cultivars

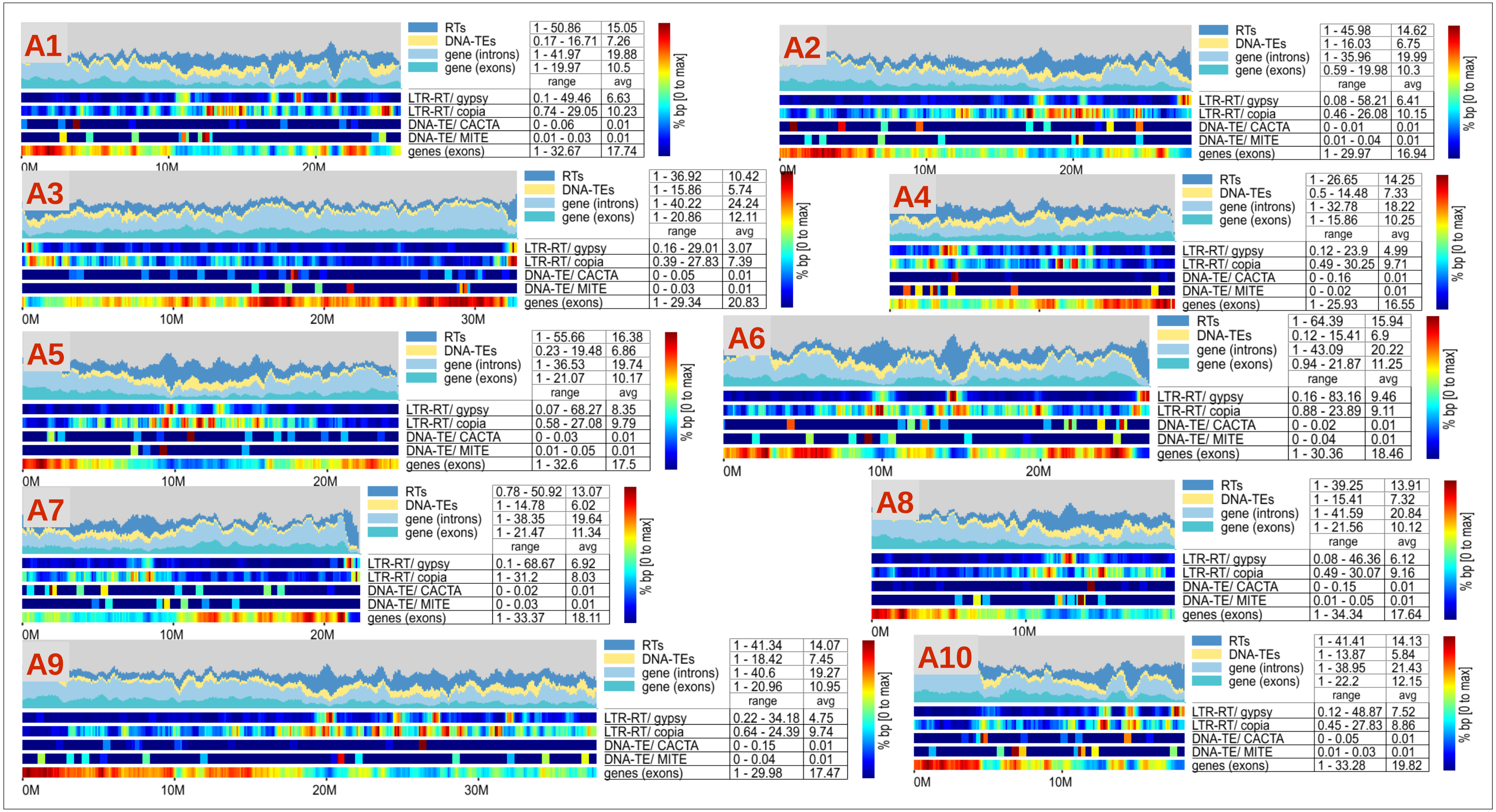
HEXAPLOID WHEAT GENOME. 2N-6X=42



THREE INDEPENDENTLY MAINTAINED GENOMES ~18 Gb.

Sequence-based approaches to genome analysis in *Brassica*

***Brassica rapa*: 284 Mb gene space sequenced 41,174 gene models**



A1

RTs	1 - 50.86	15.05
DNA-TEs	0.17 - 16.71	7.26
gene (introns)	1 - 41.97	19.88
gene (exons)	1 - 19.97	10.5
	range	avg
LTR-RT/ gypsy	0.1 - 49.46	6.63
LTR-RT/ copia	0.74 - 29.05	10.23
DNA-TE/ CACTA	0 - 0.06	0.01
DNA-TE/ MITE	0.01 - 0.03	0.01
genes (exons)	1 - 32.67	17.74

A2

RTs	1 - 45.98	14.62
DNA-TEs	1 - 16.03	6.75
gene (introns)	1 - 35.96	19.99
gene (exons)	0.59 - 19.98	10.3
	range	avg
LTR-RT/ gypsy	0.08 - 58.21	6.41
LTR-RT/ copia	0.46 - 26.08	10.15
DNA-TE/ CACTA	0 - 0.01	0.01
DNA-TE/ MITE	0.01 - 0.04	0.01
genes (exons)	1 - 29.97	16.94

A3

RTs	1 - 36.92	10.42
DNA-TEs	1 - 15.86	5.74
gene (introns)	1 - 40.22	24.24
gene (exons)	1 - 20.86	12.11
	range	avg
LTR-RT/ gypsy	0.16 - 29.01	3.07
LTR-RT/ copia	0.39 - 27.83	7.39
DNA-TE/ CACTA	0 - 0.05	0.01
DNA-TE/ MITE	0 - 0.03	0.01
genes (exons)	1 - 29.34	20.83

A4

RTs	1 - 26.65	14.25
DNA-TEs	0.5 - 14.48	7.33
gene (introns)	1 - 32.78	18.22
gene (exons)	1 - 15.86	10.25
	range	avg
LTR-RT/ gypsy	0.12 - 23.9	4.99
LTR-RT/ copia	0.49 - 30.25	9.71
DNA-TE/ CACTA	0 - 0.16	0.01
DNA-TE/ MITE	0 - 0.02	0.01
genes (exons)	1 - 25.93	16.55

A5

RTs	1 - 55.66	16.38
DNA-TEs	0.23 - 19.48	6.86
gene (introns)	1 - 36.53	19.74
gene (exons)	1 - 21.07	10.17
	range	avg
LTR-RT/ gypsy	0.07 - 68.27	8.35
LTR-RT/ copia	0.58 - 27.08	9.79
DNA-TE/ CACTA	0 - 0.03	0.01
DNA-TE/ MITE	0.01 - 0.05	0.01
genes (exons)	1 - 32.6	17.5

A6

RTs	1 - 64.39	15.94
DNA-TEs	0.12 - 15.41	6.9
gene (introns)	1 - 43.09	20.22
gene (exons)	0.94 - 21.87	11.25
	range	avg
LTR-RT/ gypsy	0.16 - 83.16	9.46
LTR-RT/ copia	0.88 - 23.89	9.11
DNA-TE/ CACTA	0 - 0.02	0.01
DNA-TE/ MITE	0 - 0.04	0.01
genes (exons)	1 - 30.36	18.46

A7

RTs	0.78 - 50.92	13.07
DNA-TEs	1 - 14.78	6.02
gene (introns)	1 - 38.35	19.64
gene (exons)	1 - 21.47	11.34
	range	avg
LTR-RT/ gypsy	0.1 - 68.67	6.92
LTR-RT/ copia	1 - 31.2	8.03
DNA-TE/ CACTA	0 - 0.02	0.01
DNA-TE/ MITE	0 - 0.03	0.01
genes (exons)	1 - 33.37	18.11

A8

RTs	1 - 39.25	13.91
DNA-TEs	1 - 15.41	7.32
gene (introns)	1 - 41.59	20.84
gene (exons)	1 - 21.56	10.12
	range	avg
LTR-RT/ gypsy	0.08 - 46.36	6.12
LTR-RT/ copia	0.49 - 30.07	9.16
DNA-TE/ CACTA	0 - 0.15	0.01
DNA-TE/ MITE	0.01 - 0.05	0.01
genes (exons)	1 - 34.34	17.64

A9

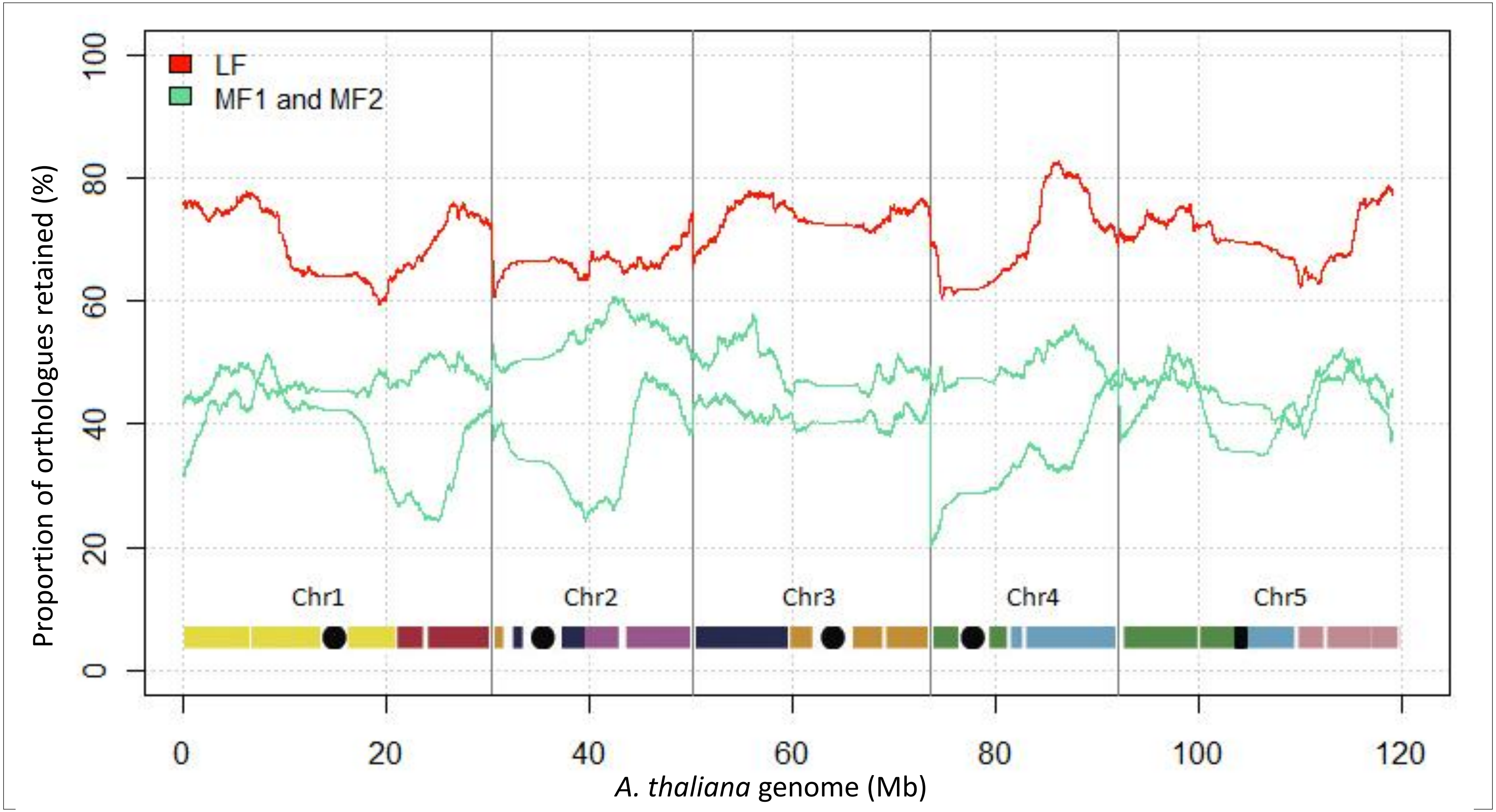
RTs	1 - 41.34	14.07
DNA-TEs	1 - 18.42	7.45
gene (introns)	1 - 40.6	19.27
gene (exons)	1 - 20.96	10.95
	range	avg
LTR-RT/ gypsy	0.22 - 34.18	4.75
LTR-RT/ copia	0.64 - 24.39	9.74
DNA-TE/ CACTA	0 - 0.15	0.01
DNA-TE/ MITE	0 - 0.04	0.01
genes (exons)	1 - 29.98	17.47

A10

RTs	1 - 41.41	14.13
DNA-TEs	1 - 13.87	5.84
gene (introns)	1 - 38.95	21.43
gene (exons)	1 - 22.2	12.15
	range	avg
LTR-RT/ gypsy	0.12 - 48.87	7.52
LTR-RT/ copia	0.45 - 27.83	8.86
DNA-TE/ CACTA	0 - 0.05	0.01
DNA-TE/ MITE	0.01 - 0.03	0.01
genes (exons)	1 - 33.28	19.82

Ian Brancroft JIC

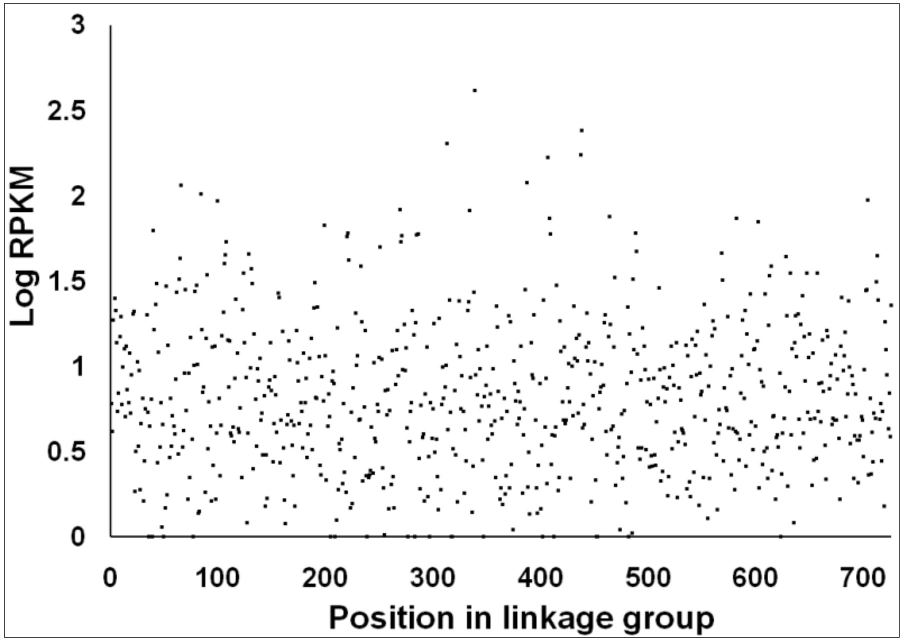
Differential gene loss from *Brassica rapa* paralogous genome segments



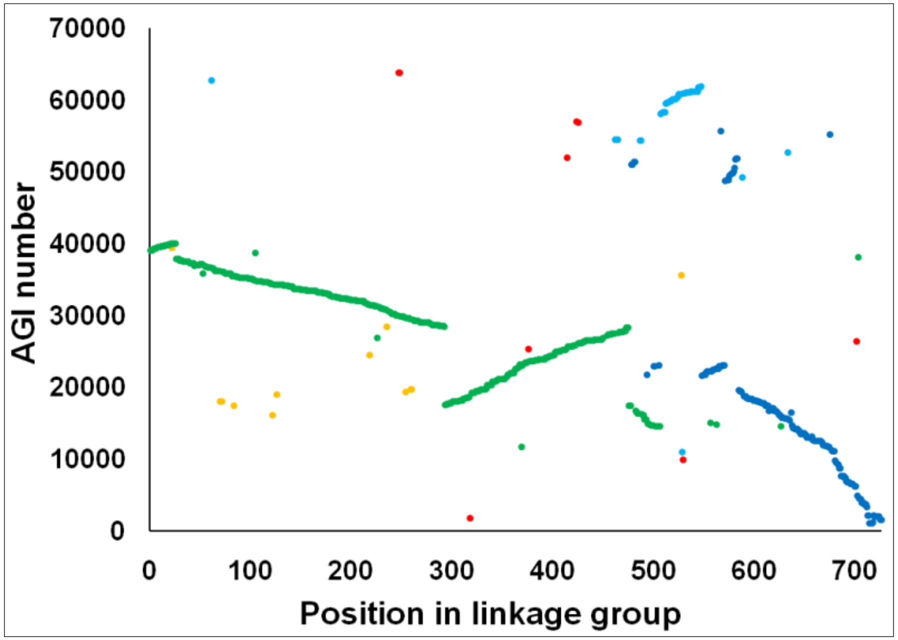
Ian Bancroft JIC

B. NAPUS GENOME ANALYSIS USING mRNA-seq

Analyse transcript abundance relative to position in genome



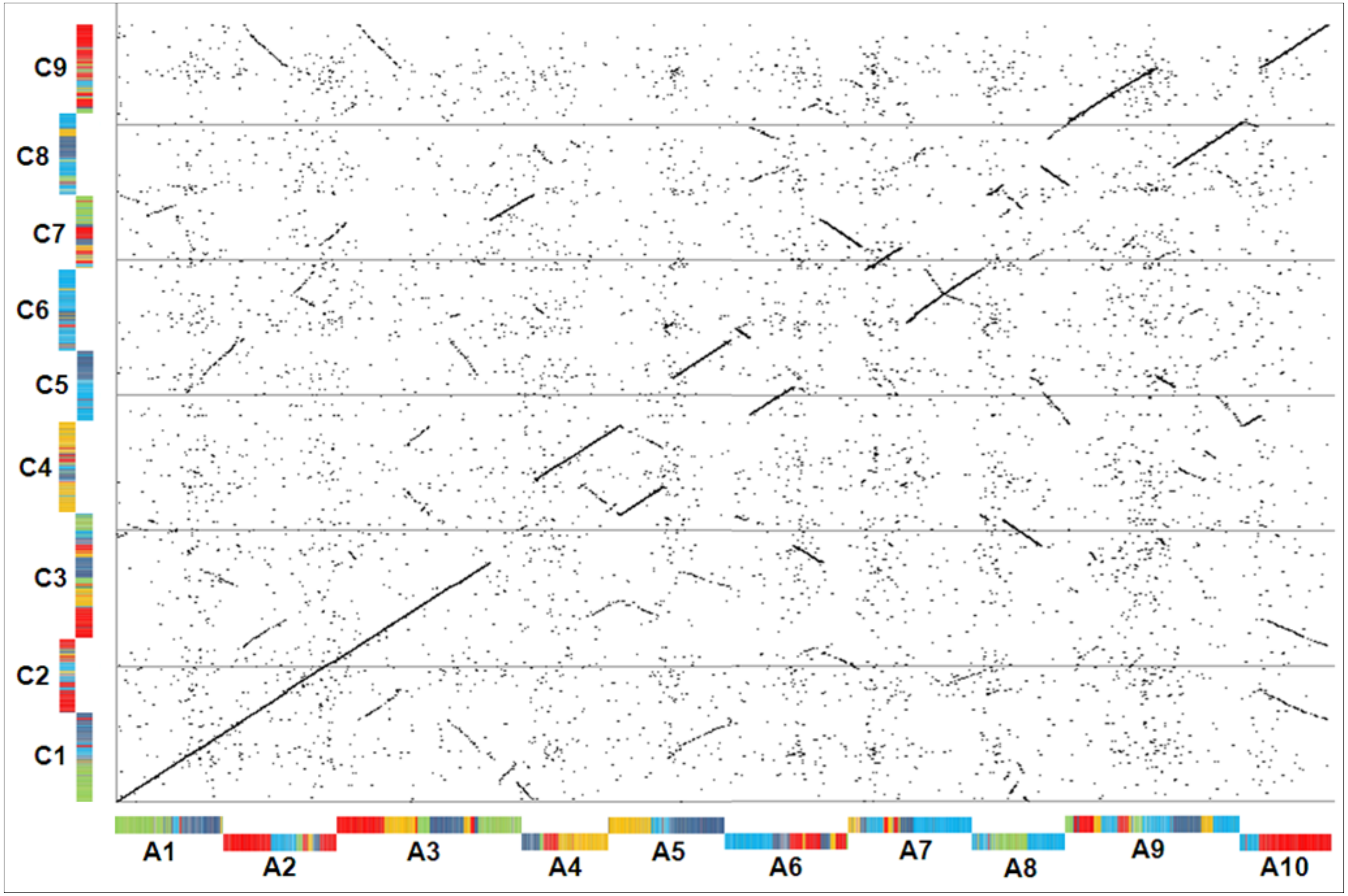
Analyse collinearity with the genome of *Arabidopsis thaliana*



Ian Bancroft JIC

BEST UNIGENE SEQUENCE MATCHES TO A AND C GENOMES

Random and paralogue matches across genomes (as represented in *B. napus*)

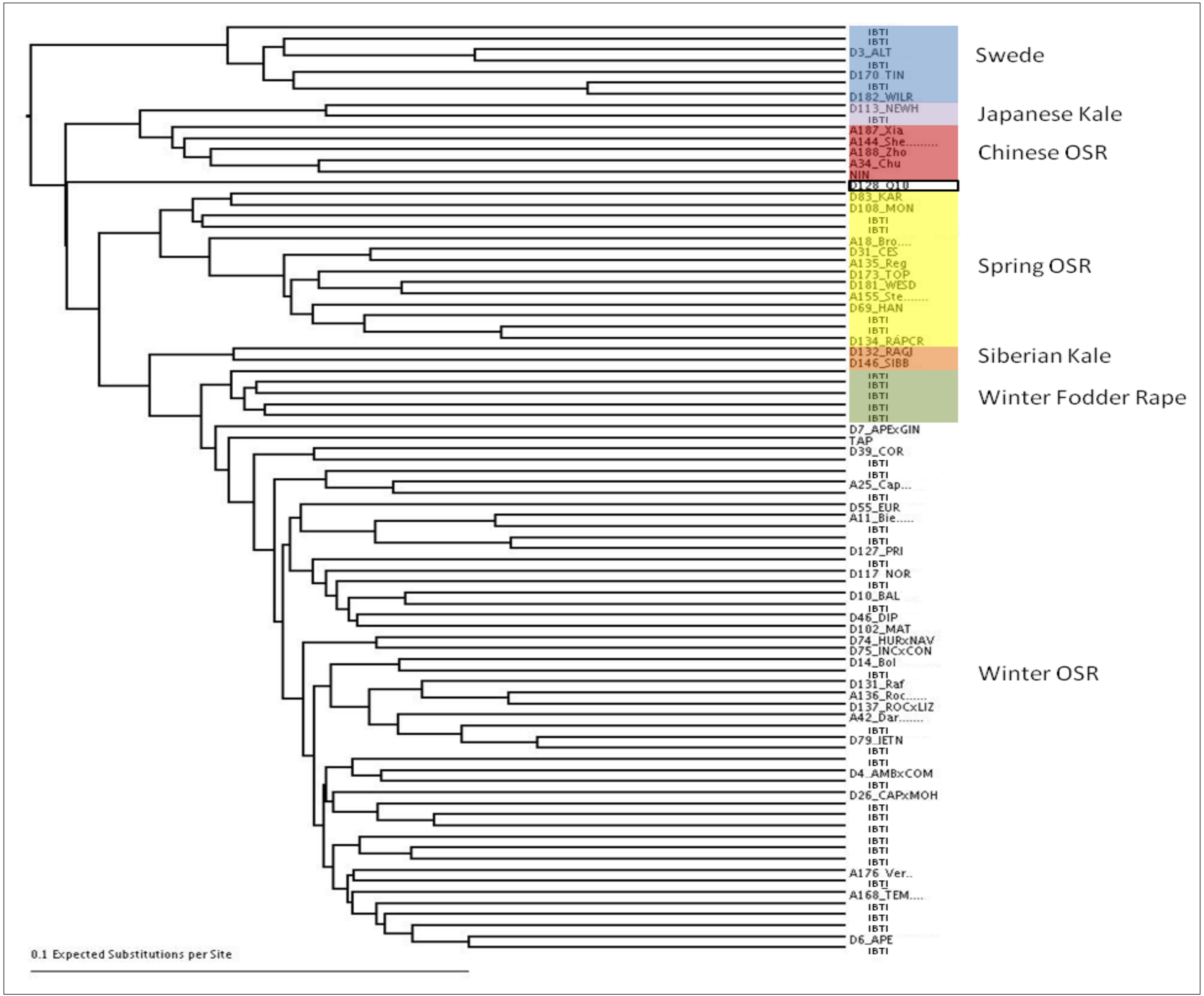


57,408 unigenes BLAST-mapped to both genomes (no filtering)

Ian Bancroft JIC

PHYLOGENETIC ANALYSIS

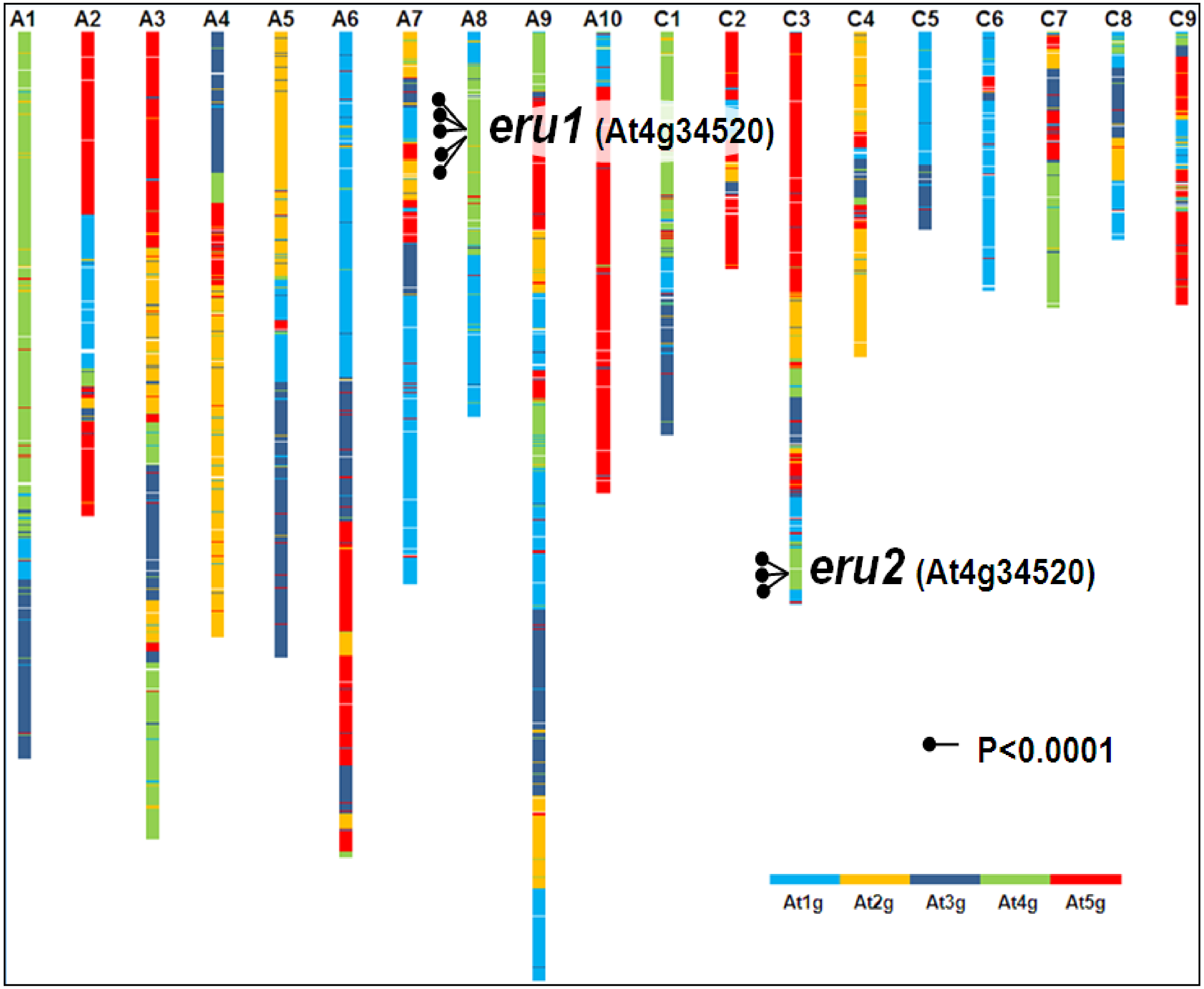
Phylogenetic analysis across 83 *B. napus* accessions with 101,644 transcriptome SNPs.



Ian Bancroft JIC

ASSOCIATION GENETICs

Linkage disequilibrium analysis for erucic acid content of seed oil across 53 *B. napus* accessions, using mixed linear models with 43,744 transcriptome SNPs.



Ian Bancroft JIC

GRASS CHROMOSOME CHANGES AID GENOMIC ANALYSIS

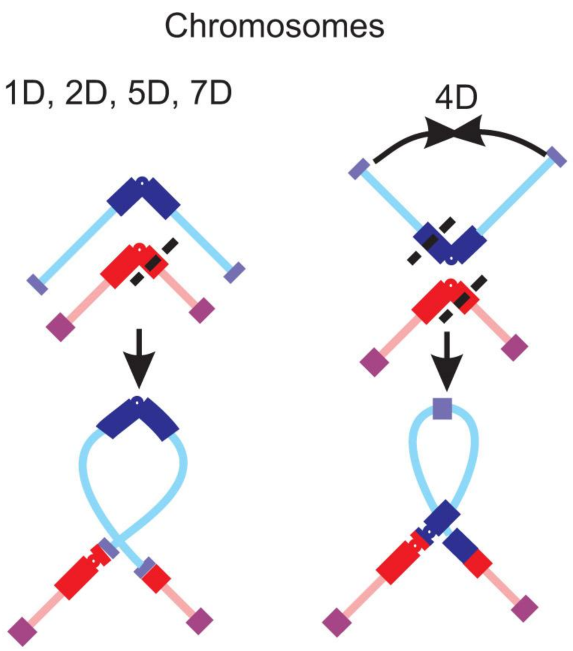
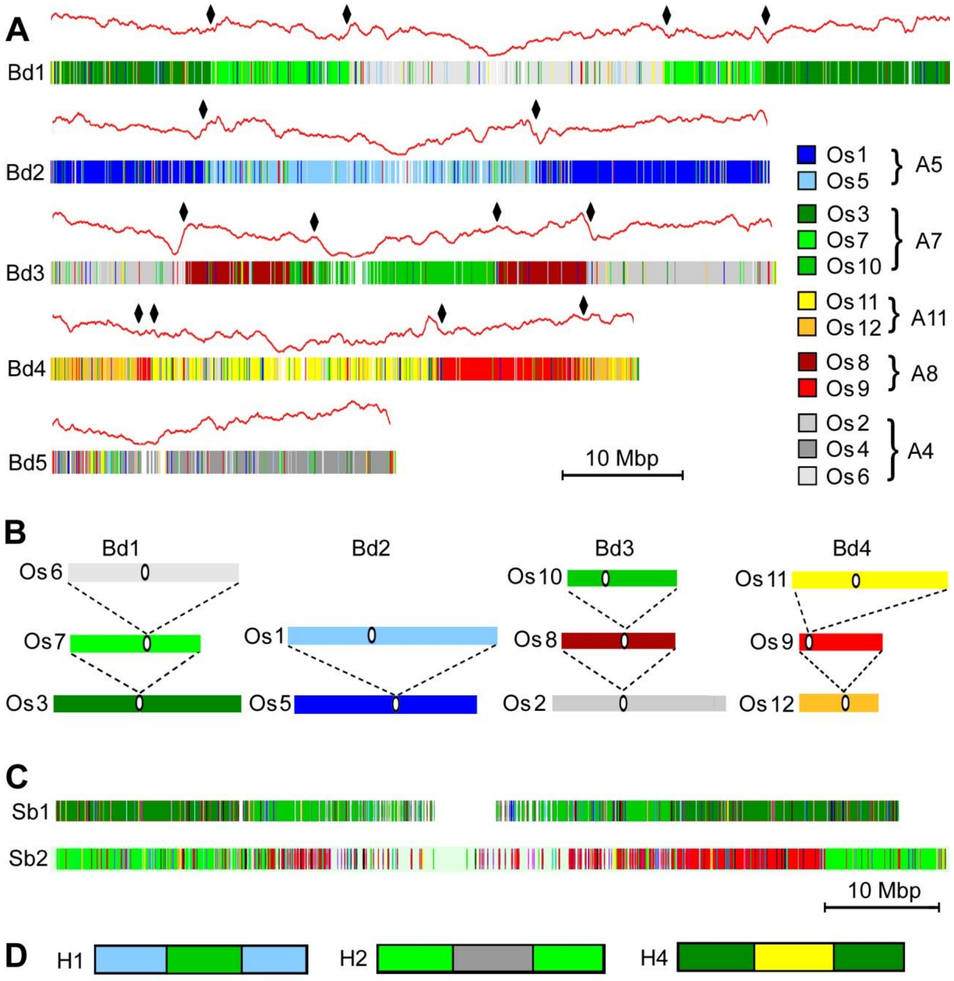


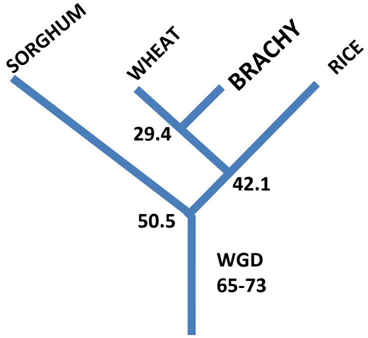
Fig. 2. Insertional model of dysploid reduction of basic chromosome number. Centromeres are indicated by opened circles, and heterochromatin is indicated by boxes. A break is indicated by a dashed line.

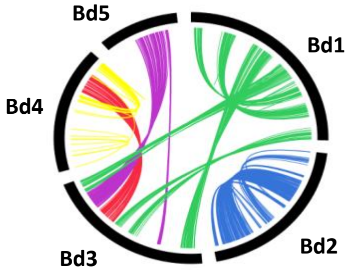
Luo et al PNAS 2009

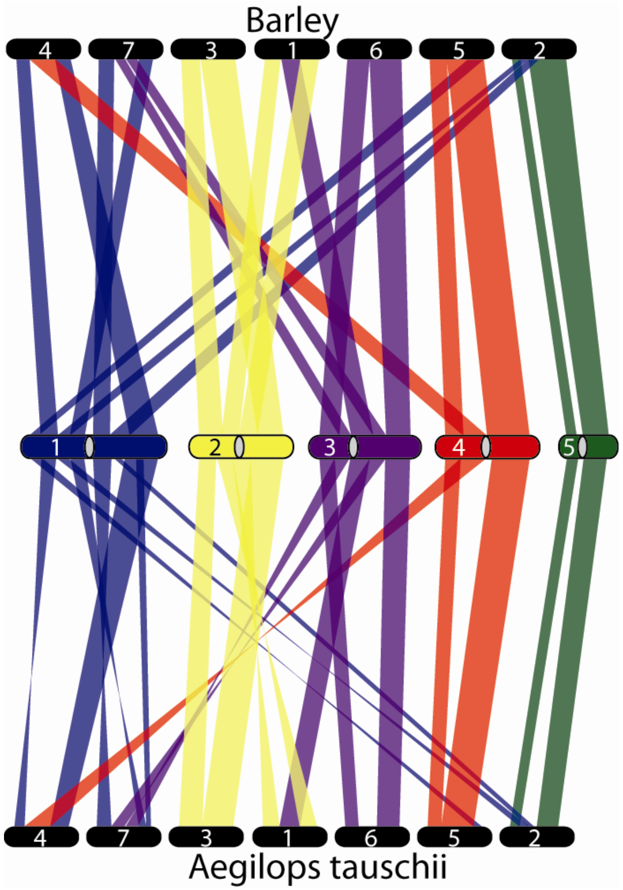


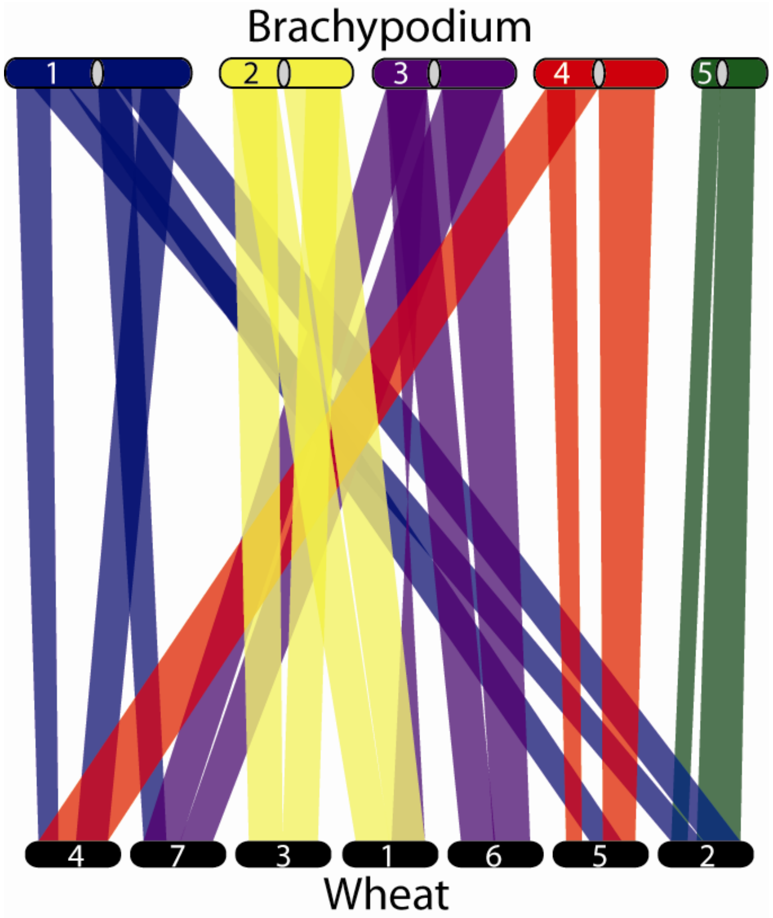
Thomas Wicker

Whole genome synteny analysis

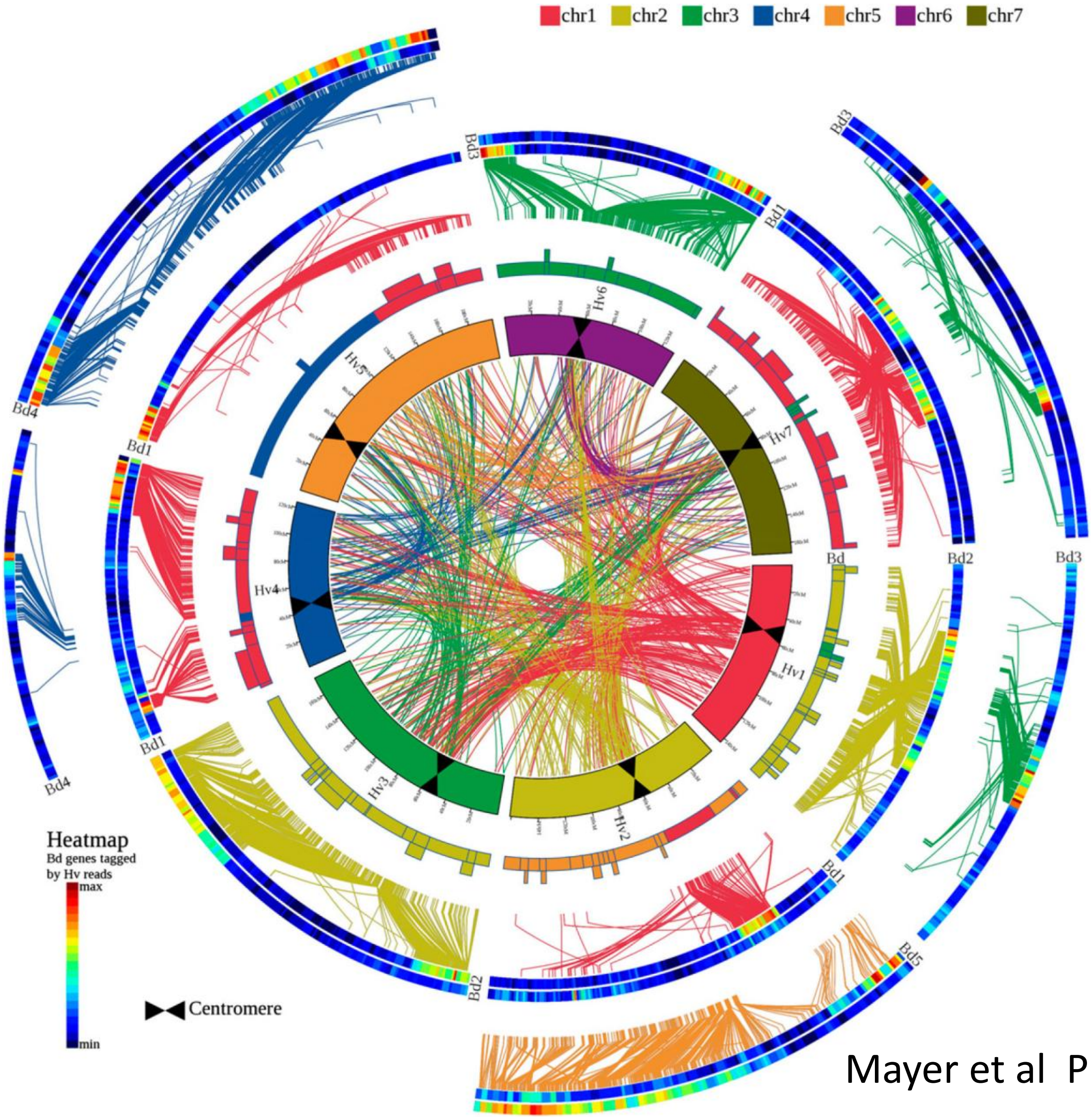








The genome zipper concept exploits conserved gene order to build An initial gene order – for example in barley.

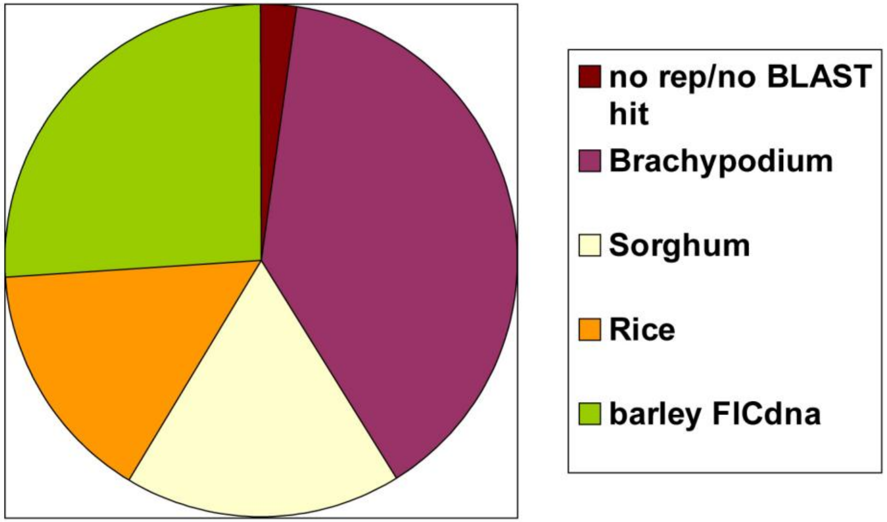


Mayer et al Plant Cell 2011

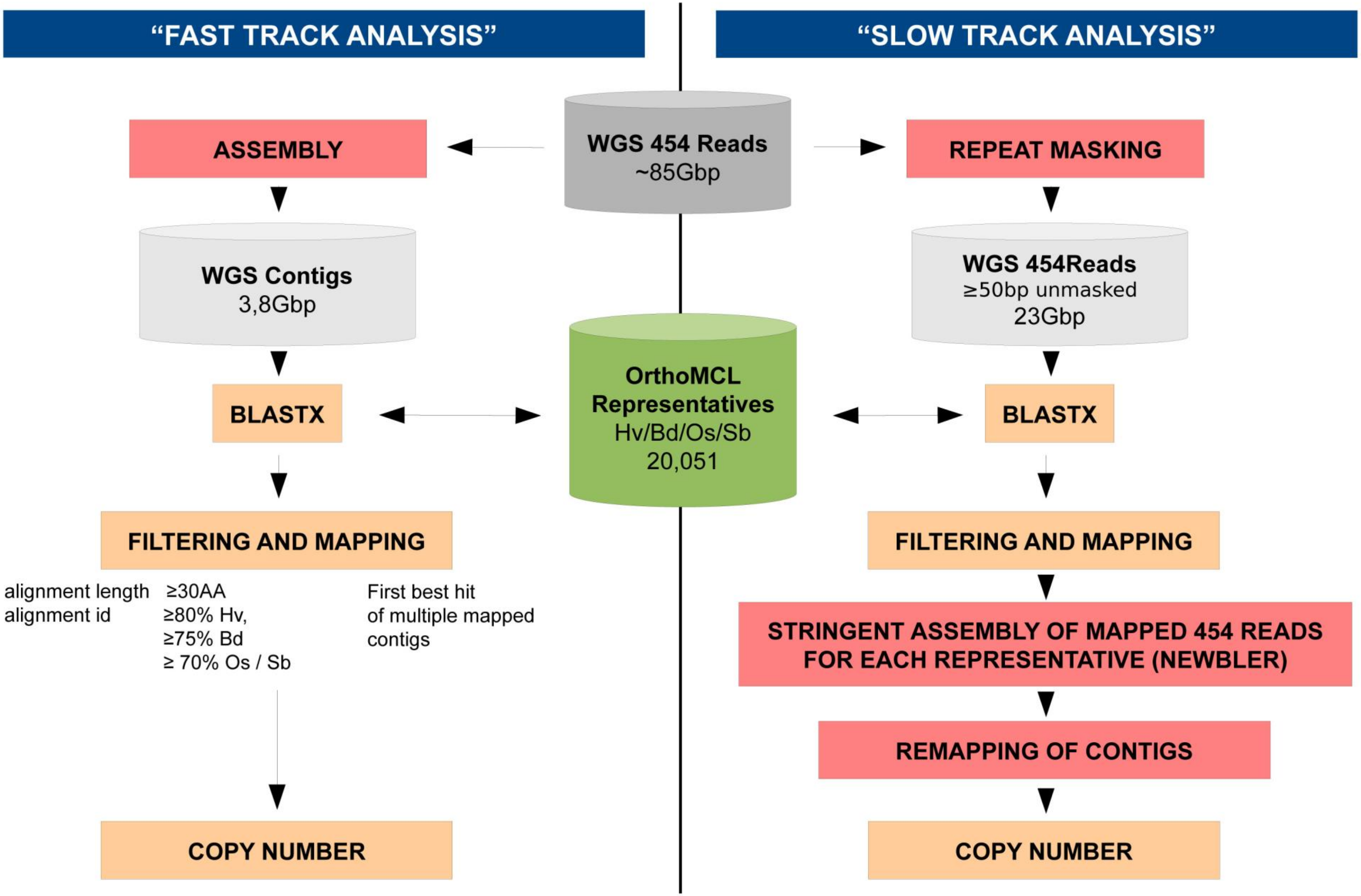
DEFINING A SET OF CONSERVED OTHOLOGOUS GENES FOR THE TRITICEAE

- 7996 Brachypodium reps
- 3582 Sorghum reps
- 3136 Rice reps
- 5337 barley flcDNA rep
- 445 with no rep due to NO BLAST hits from wheat

- Total 20,496 clusters



Workflow for determining the copy numbers



"FAST TRACK ANALYSIS"

"SLOW TRACK ANALYSIS"

WGS 454 Reads
~85Gbp

OrthoMCL Representatives
Hv/Bd/Os/Sb
20,051

ASSEMBLY

WGS Contigs
3,8Gbp

BLASTX

FILTERING AND MAPPING

alignment length	≥30AA	First best hit of multiple mapped contigs
alignment id	≥80% Hv, ≥75% Bd ≥ 70% Os / Sb	

COPY NUMBER

REPEAT MASKING

WGS 454Reads
≥50bp unmasked
23Gbp

BLASTX

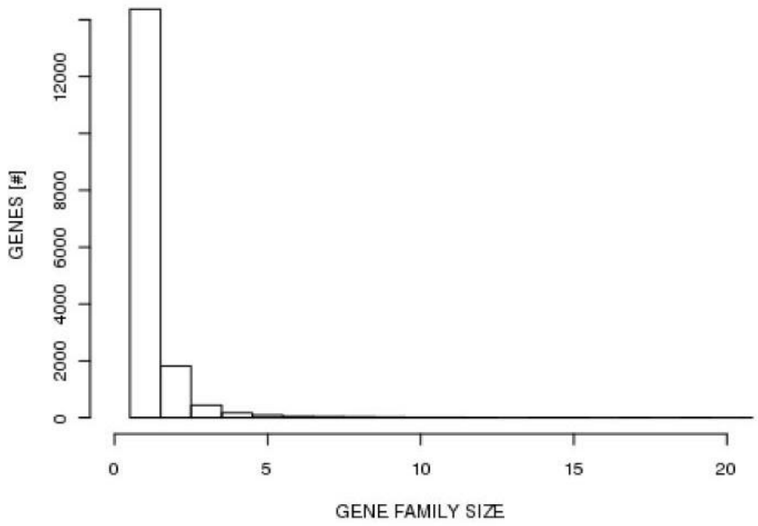
FILTERING AND MAPPING

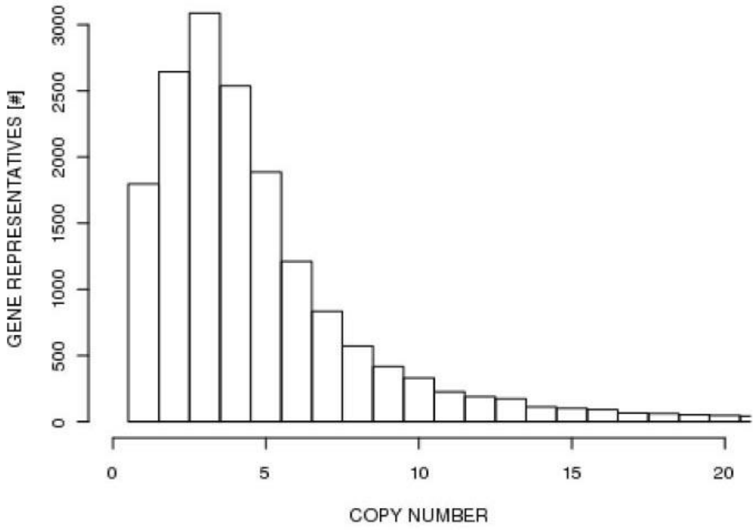
STRINGENT ASSEMBLY OF MAPPED 454 READS FOR EACH REPRESENTATIVE (NEWBLER)

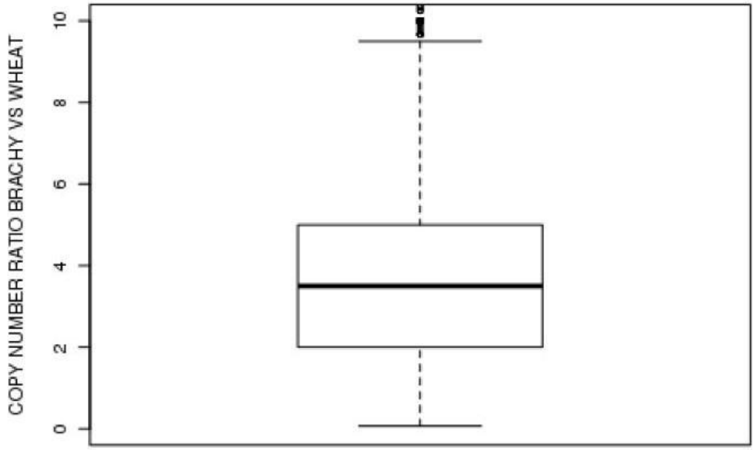
REMAPPING OF CONTIGS

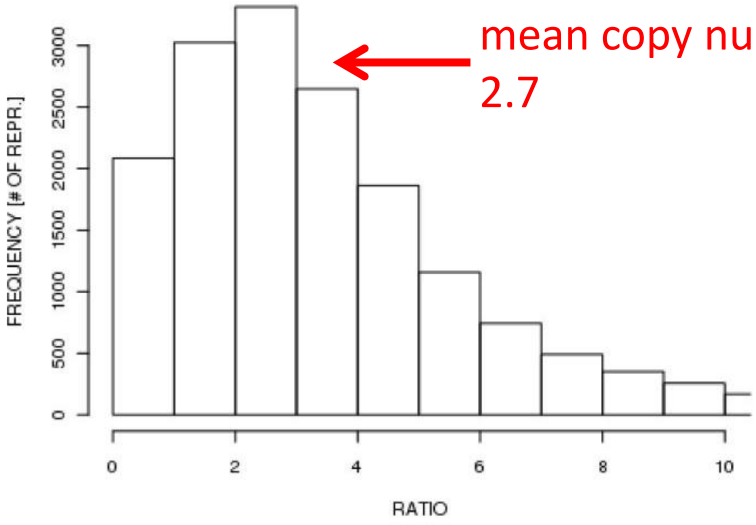
COPY NUMBER

COPY NUMBER OF WHEAT ASSEMBLIES


GENE FAMILY SIZE DISTRIBUTION IN BRACHY

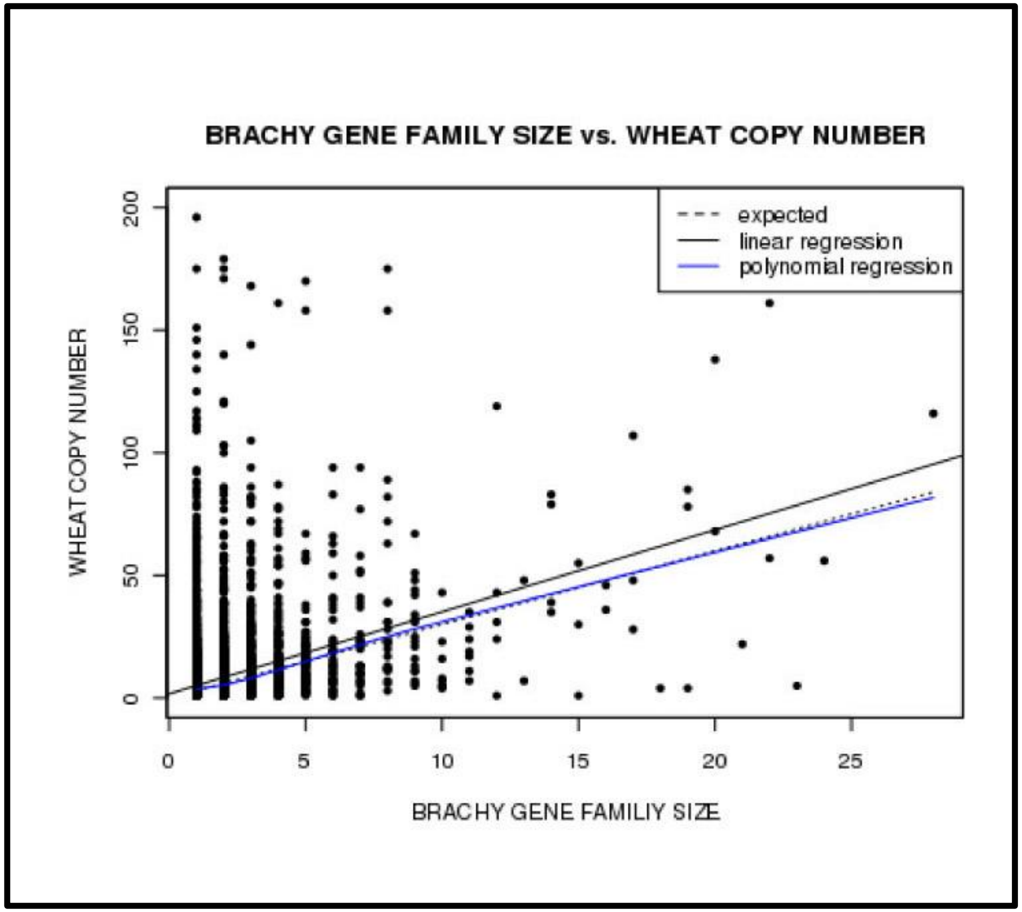

COPY NUMBER DISTRIBUTION WHEAT


DISTRIBUTION OF COPY NUMBER RATIO BRACHY VS WHEAT

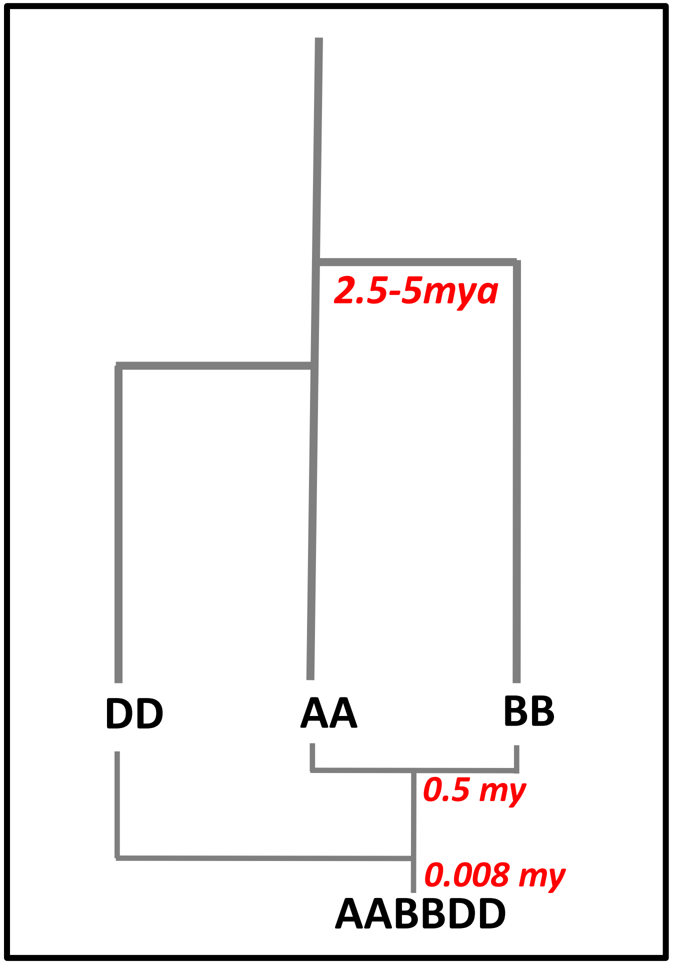

DISTRIBUTION OF COPY NUMBER RATIO BRACHY VS WHEAT

mean copy number 2.7

Final Analyses



Gene loss and amplification in polyploid



Assigning sub assemblies to genome by reference to AA and DD genome sequences and phylogenomics

ALIGNMENT OF WHEAT ASSEMBLIES ACCORDING TO CONSERVED GENE ORDER

Aim:

The alignment to orthologs generates an ancestral Triticeae chromosomal order, and the alignment should use available wheat genetic maps

Method:

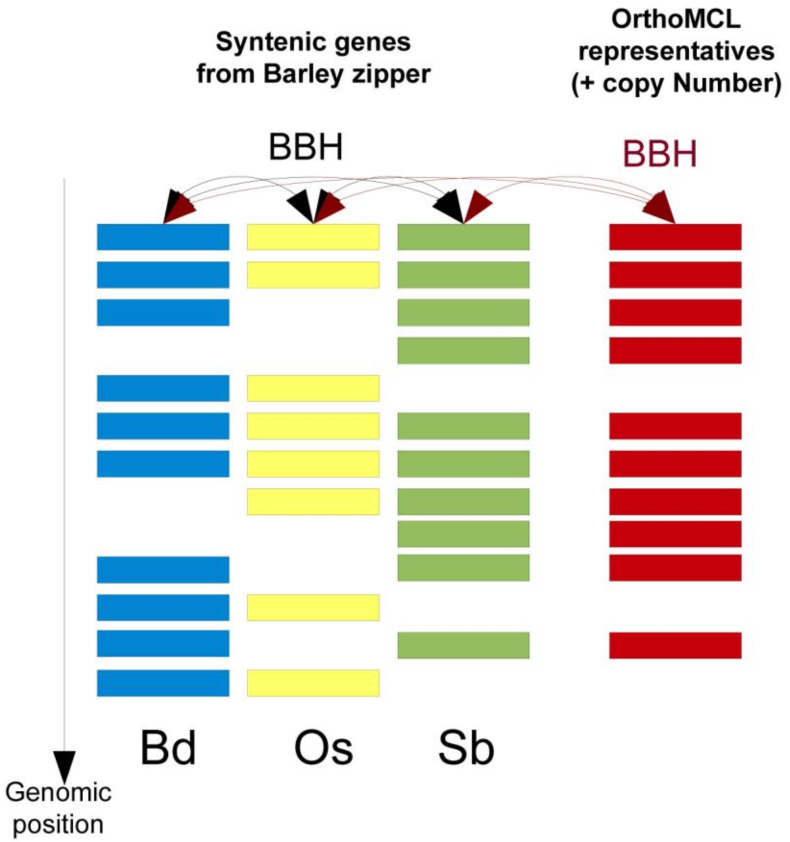
1 Define syntenic blocks of Bd/Os/Sb from Barley genome zipper results

2 Extract genes out of the syntenic blocks

3 Order these genes according to Bd genome position

4 Anchor Os/Sb genes to arranged bd genes

5 Anchor Representatives to the arranged bd/os/sb genes



PRINCIPAL INVESTIGATORS

- Neil Hall, Anthony Hall (Liverpool)
- Dick McCombie (CSHL)
- Keith Edwards (Bristol)
- Paul Kersey, Ewan Birney (EBI)
- Klaus Mayer (H-Z Munich)
- Mike Bevan, Cristobal Uauy (JIC)

CSH Cold Spring Harbor Laboratory

HelmholtzZentrum münchen
German Research Center for Environmental Health

EMBL-EBI

UNIVERSITY OF
LIVERPOOL

John Innes Centre

University of
BRISTOL

bbsrc
biotechnology and biological sciences
research council

NSF

Bundesministerium
für Bildung
und Forschung

SEVENTH FRAMEWORK
PROGRAMME