



# Deciphering the Diploid Ancestral Genome of the Mesohehexaploid *Brassica rapa*

**Feng Cheng**

**chengfeng@caas.cn**

**Institute of Vegetables and Flowers  
Chinese Academy of Agricultural Sciences**



<http://brassicadb.org>

**31 Mar, 2014**

# Contents

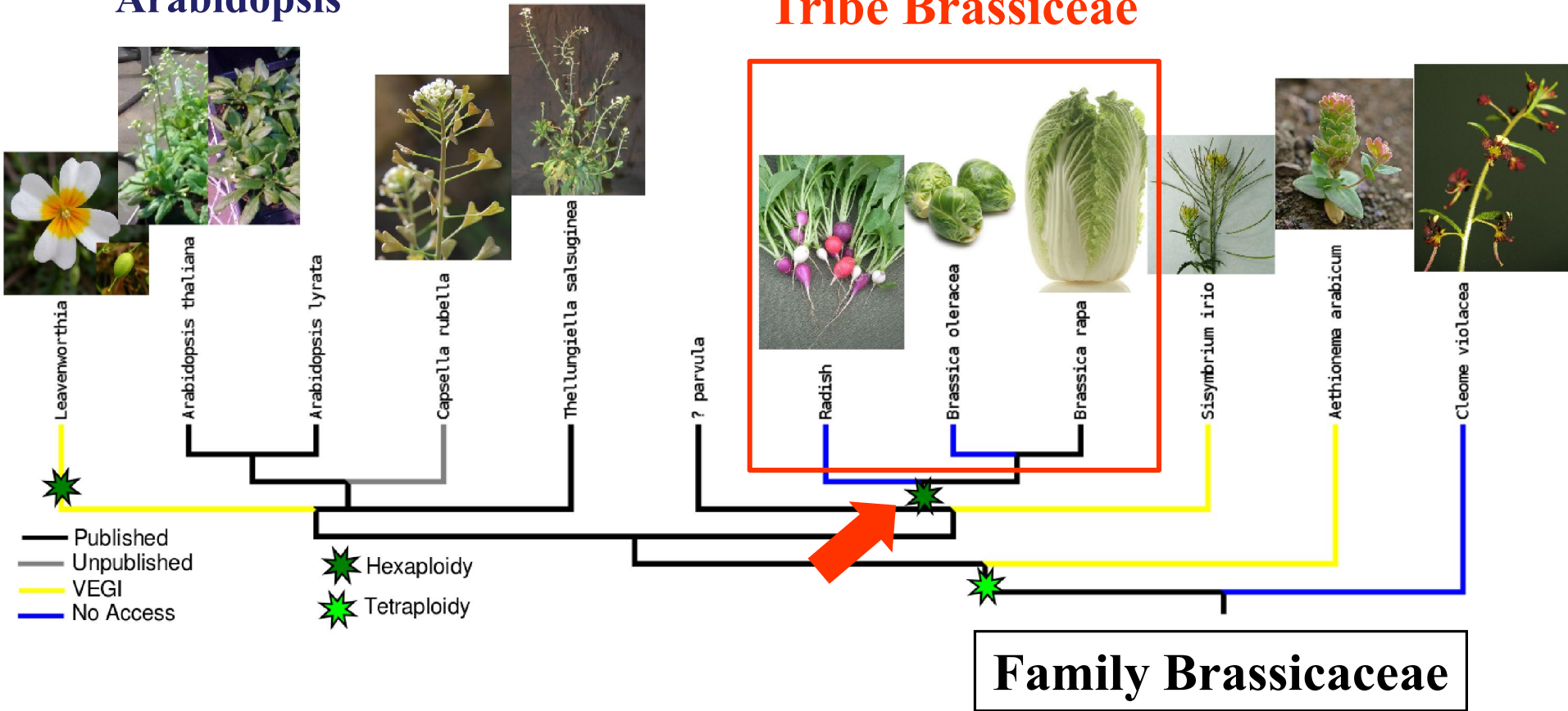
- **Inferring the diploid ancestor of *B. rapa***
- **Paleocentromere evolution in *B. rapa***
- **Common ancestor of Brassiceae**



# Genome sequencing of Brassicaceae

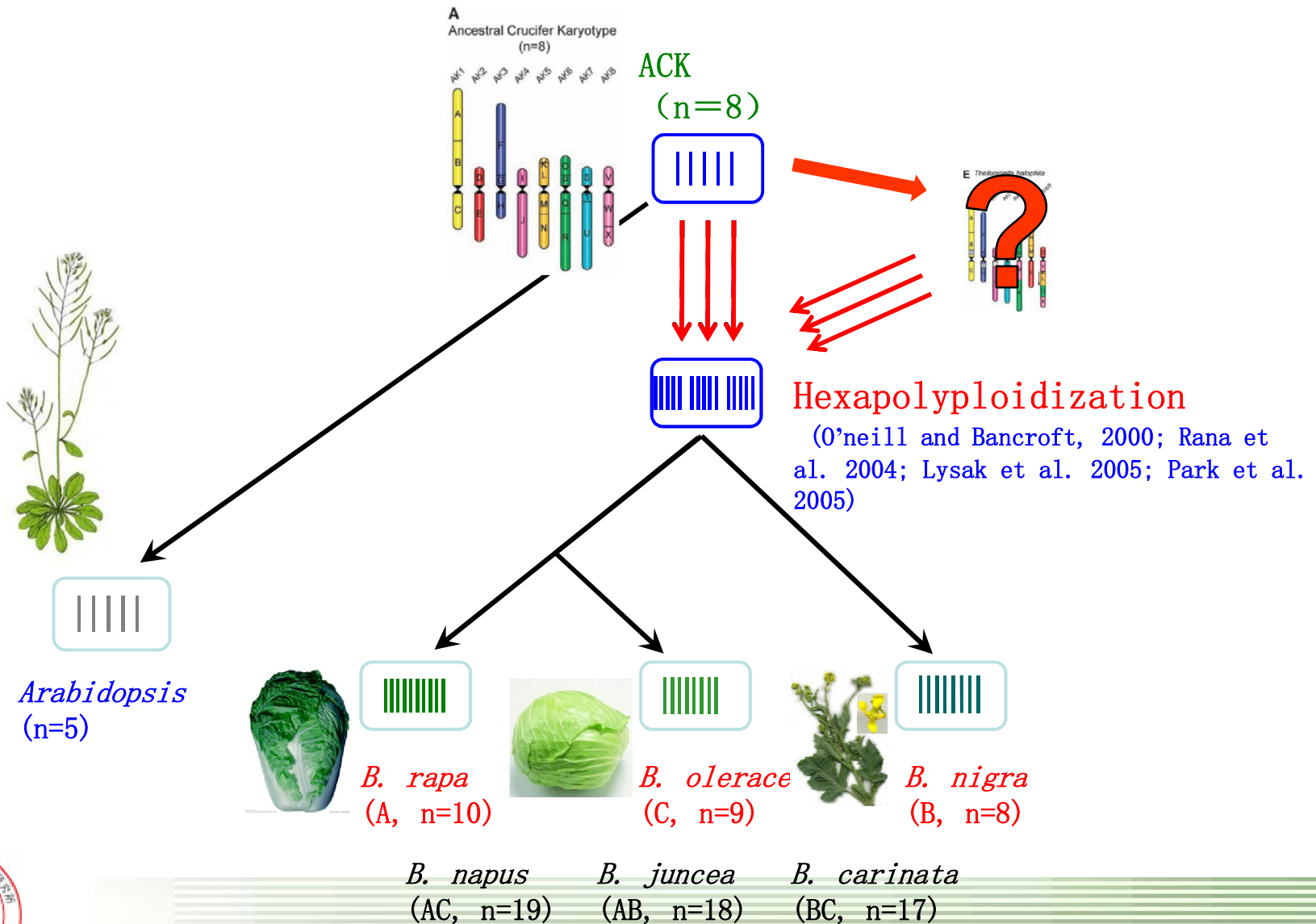
## Arabidopsis

## Tribe Brassiceae



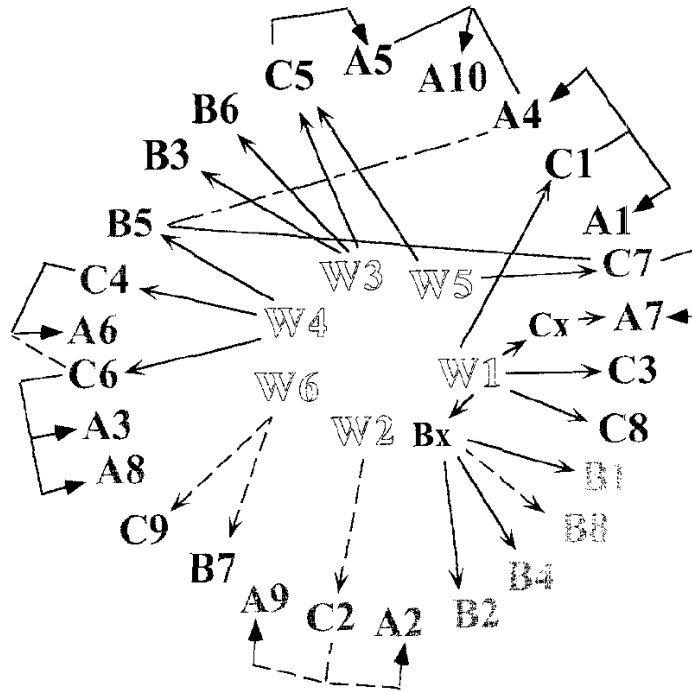
From James, Berkeley

# Evolution of *Brassica* species



# Argument of *Brassica* ancestor

1. Number of chromosomes ( $n=6/7/8$ ) ?



2. Triplicate from a common diploid ancestor?

3. Both whole genome and segmental duplication?

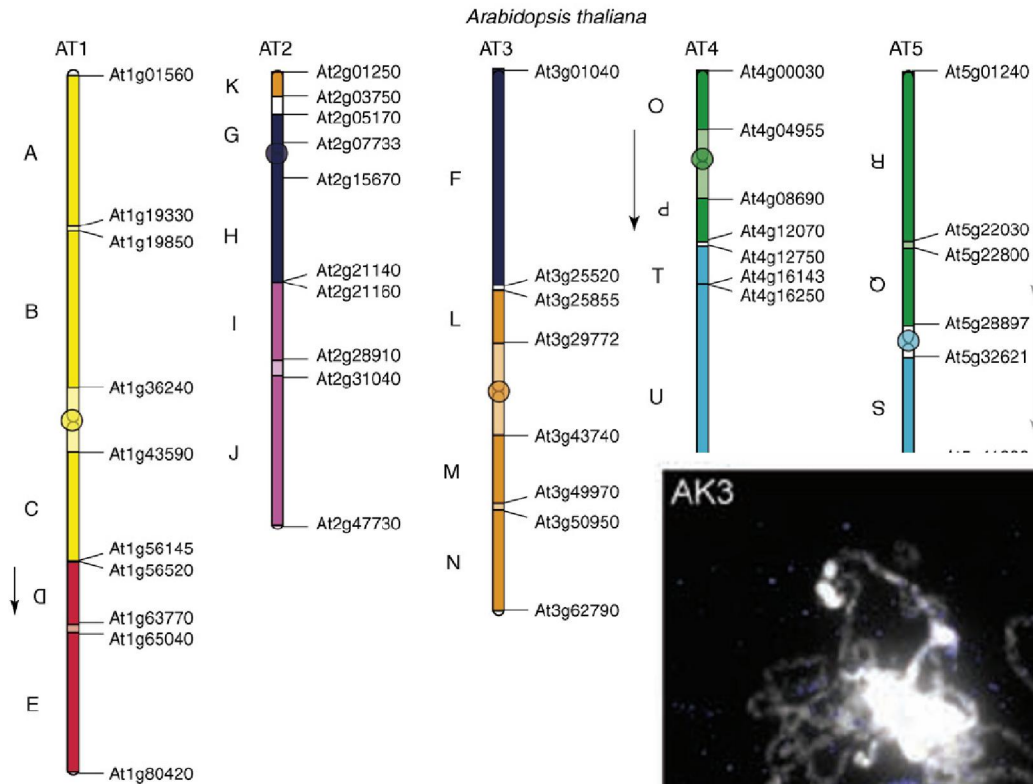


# Inferring the diploid ancestor of *B. rapa*



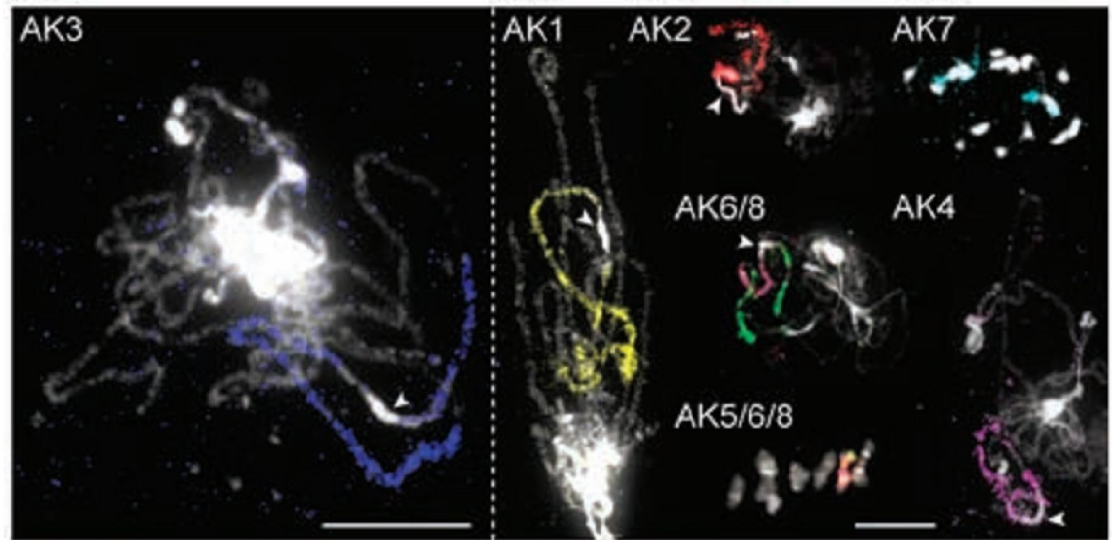


# Genomic structure of Brassicaceae



**CCP, comparative chromosome painting: FISH based karyotype investigation and chromosome comparison among species.**

*A. thaliana*,  
*A. lyrata*, *C. rubella*,  
*B. napus*



*Isobel Parkin, 2005, Genetics*  
*Terezie and Lysak, 2008, Plant cell*  
*Schranz, 2009, Trends in Plant Science*



# Intervals of 24 blocks in Arabidopsis

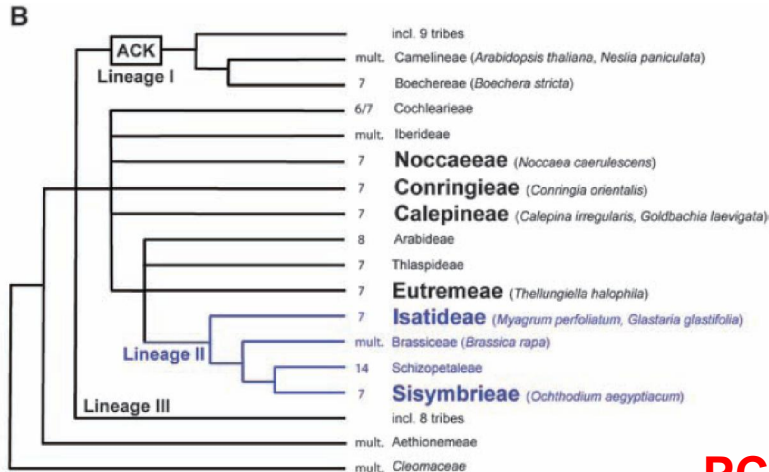
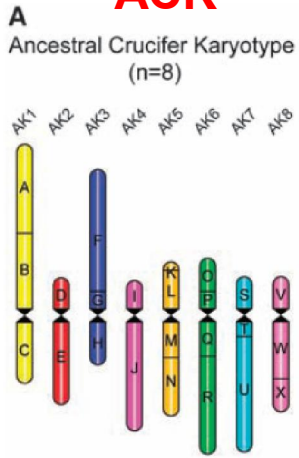
Block	~Interval	At LG <sup>a</sup>	At order <sup>b</sup>	At orient. <sup>c</sup>	At LG <sup>d</sup>	At order <sup>e</sup>	At orient. <sup>f</sup>	Brassica block <sup>g</sup>	Frequency in Br <sup>h</sup>
A	At1g01560–At1g19330	1	1	+	1	1	+	C1A	4
B	At1g19850–At1g36240	1	2	+	1	2	+	C1B	5
C	At1g43590–At1g56145	1	3	+	1	3	+	C1C	3
D	At1g63770–At1g56520	2	4	+	1	4	–	C1D	1
E	At1g65040–At1g80420	2	5	+	1	5	+	C1E	2
F	At3g01040–At3g25520	3	6	+	3	11	+	C3A	3
G	At2g05170–At2g07733	3	7	+	2	7	+	C2A	1
H	At2g15670–At2g21140	3	8	+	2	8	+	C2A	1
I	At2g21160–At2g28910	4	9	+	2	9	+	C2B	4
J	At2g31040–At2g47730	4	10	+	2	10	+	C2C	3
K	At2g01250–At2g03750	5	11	+	2	6	+	C2A	2
L	At3g25855–At3g29772	5	12	+	3	12	+	C3B	2
M	At3g43740–At3g49970	5	13	+	3	13	+	C3C	0
N	At3g50950–At3g62790	5	14	+	3	14	+	C3D	3
O	At4g00030–At4g04955	6	15	+	4	15	+	C4A	2
P	At4g12070–At4g08690	6	16	+	4	16	–	C4A	2
Q	At5g28897–At5g22800	6	17	+	5	20	–	C5B	3
R	At5g22030–At5g01240	6	18	+	5	19	–	C5A	3
S	At5g41900–At5g32621	7	19	+	5	21	–	C5C	3
T	At4g12750–At4g16143	7	20	+	4	17	+	C4B'	4
U	At4g16250–At4g38770	7	21	+	4	18	+	C4B	2
V	At5g48520–At5g42970	8	22	+	5	22	–	C5D	3
W	At5g49430–At5g60390	8	23	+	5	23	+	C5E	3
X	At5g60550–At5g67385	8	24	+	5	24	+	C5F	2





# Main diploid karyotypes in Brassicaceae

## ACK



**ACK: Ancestral Crucifer Karyotype, n=8**

**PCK: Proto-Calepineae Karyotype, n=7**

**tPCK: translocated PCK, n=7**

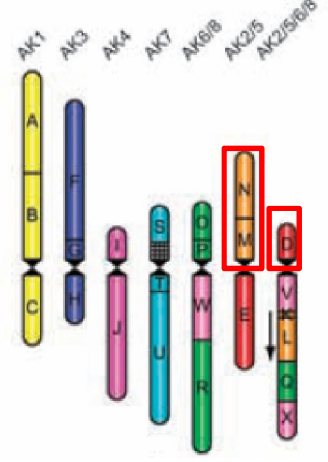
## PCK

Proto-Calepineae Karyotype (n=7)



## tPCK

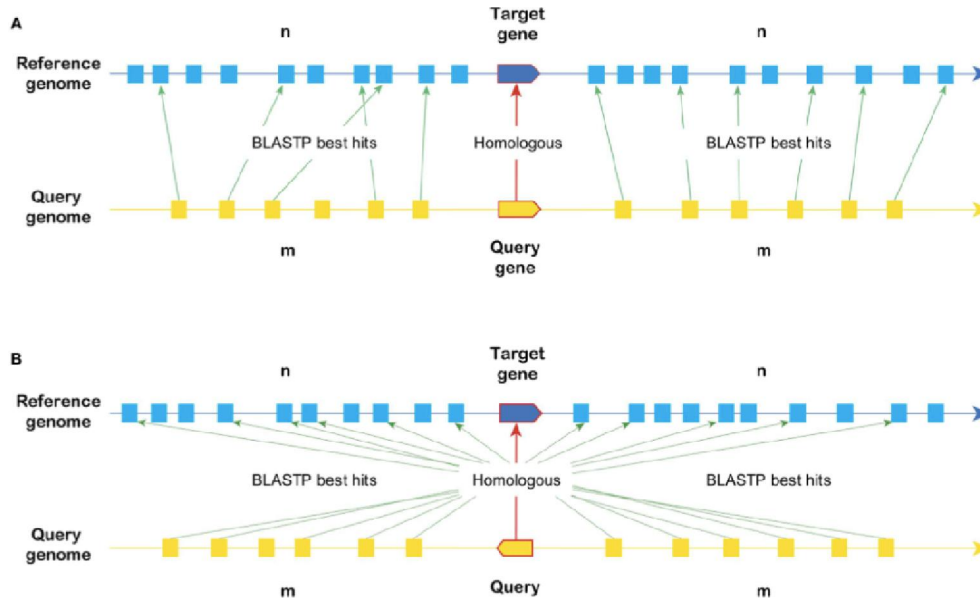
Isatideae, Sisymbrieae, Eutremeae



Terezie and Lysak, 2008, Plant cell



# At-Br syntenic gene determination

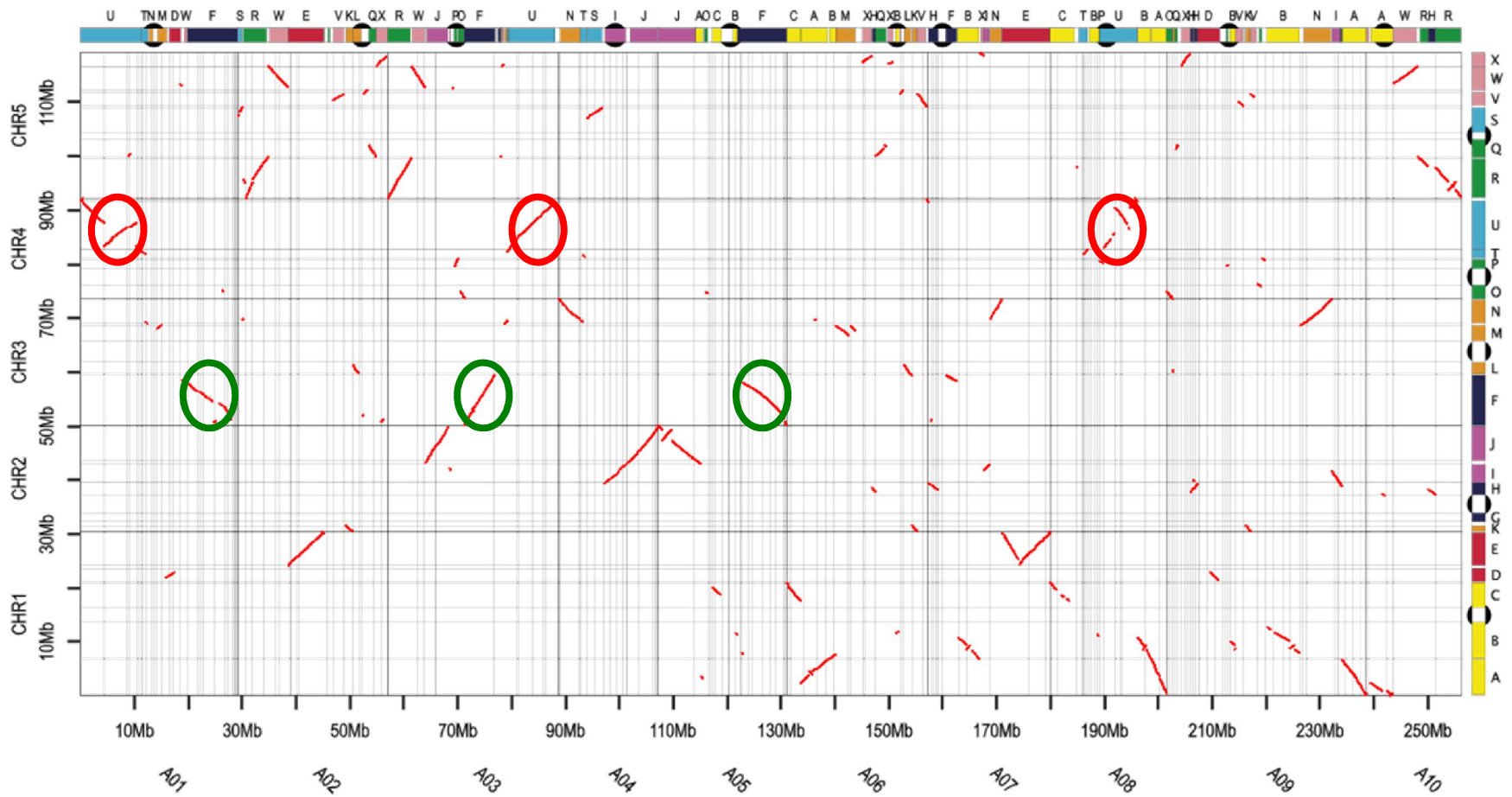


## SynOrths

- #At genes: 24,730; #Br genes: 38,945
- **18,388** At genes syntenic to **30,926** Br genes
- Among Br genes: **10,147** as one copy, **7,067** as two copies, **2,215** as three copies

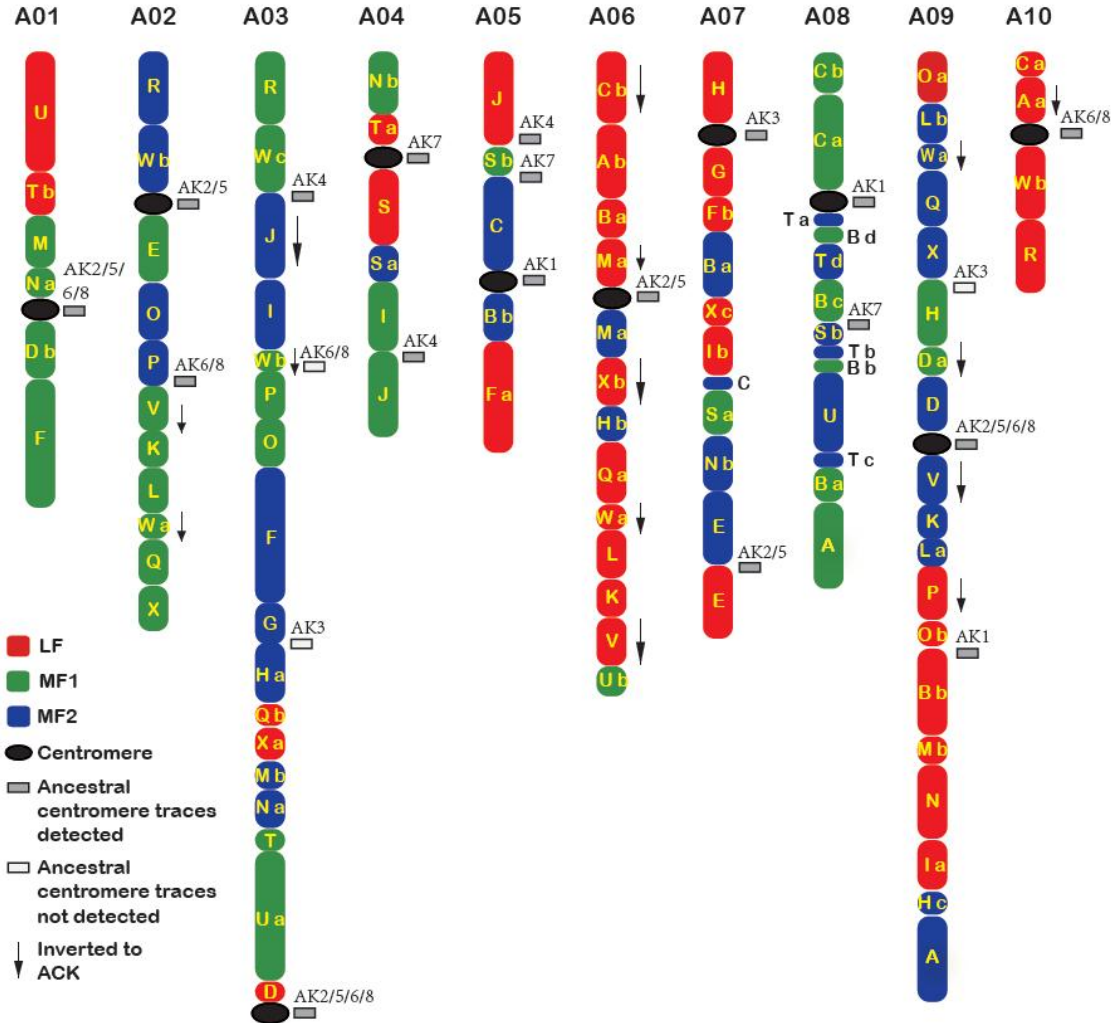


# Plot of syntenic genes



# Triplicated 24 blocks in *B. rapa*

*Brassica rapa*  
(n=10)



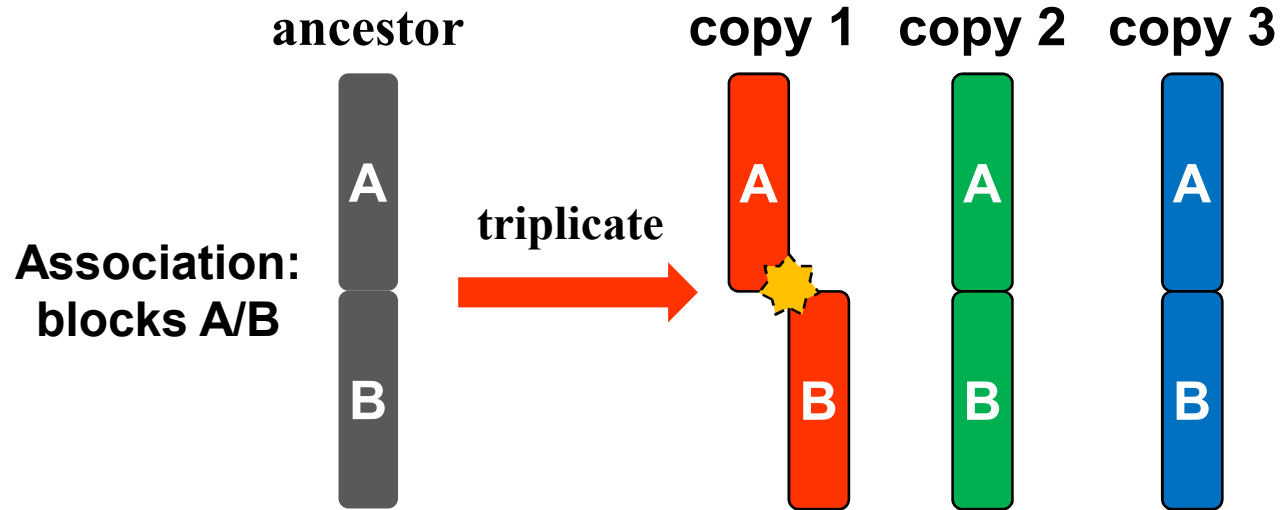
71 blocks detected,  
( $24 \times 3 - 1$ ), one copy of  
G lost after WGT

No segmental  
duplication observed

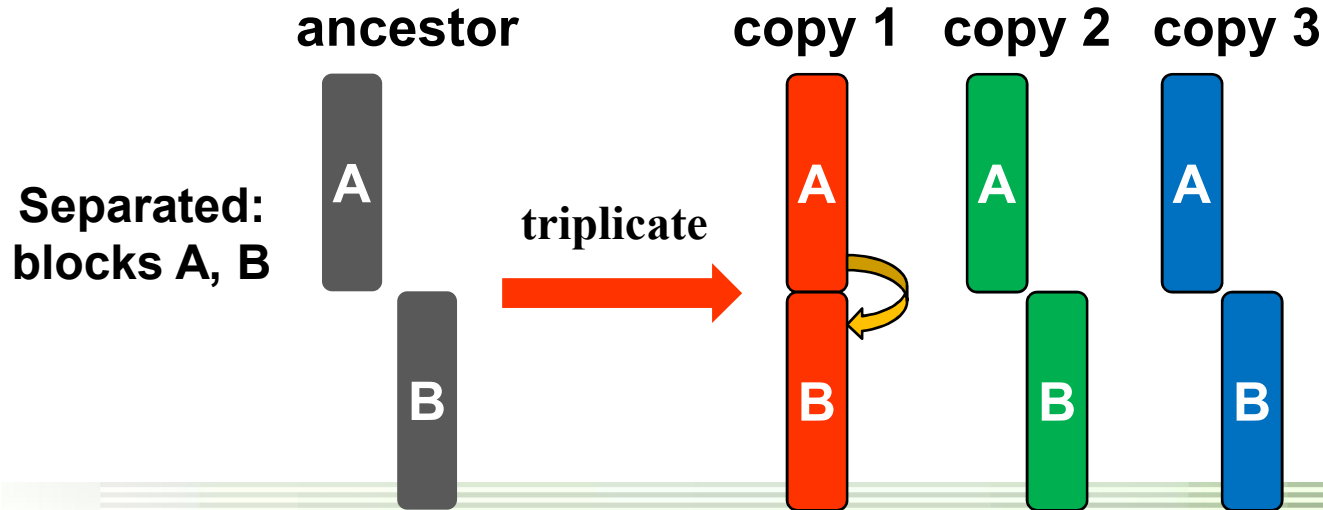


# Reshuffling of blocks in *B. rapa*

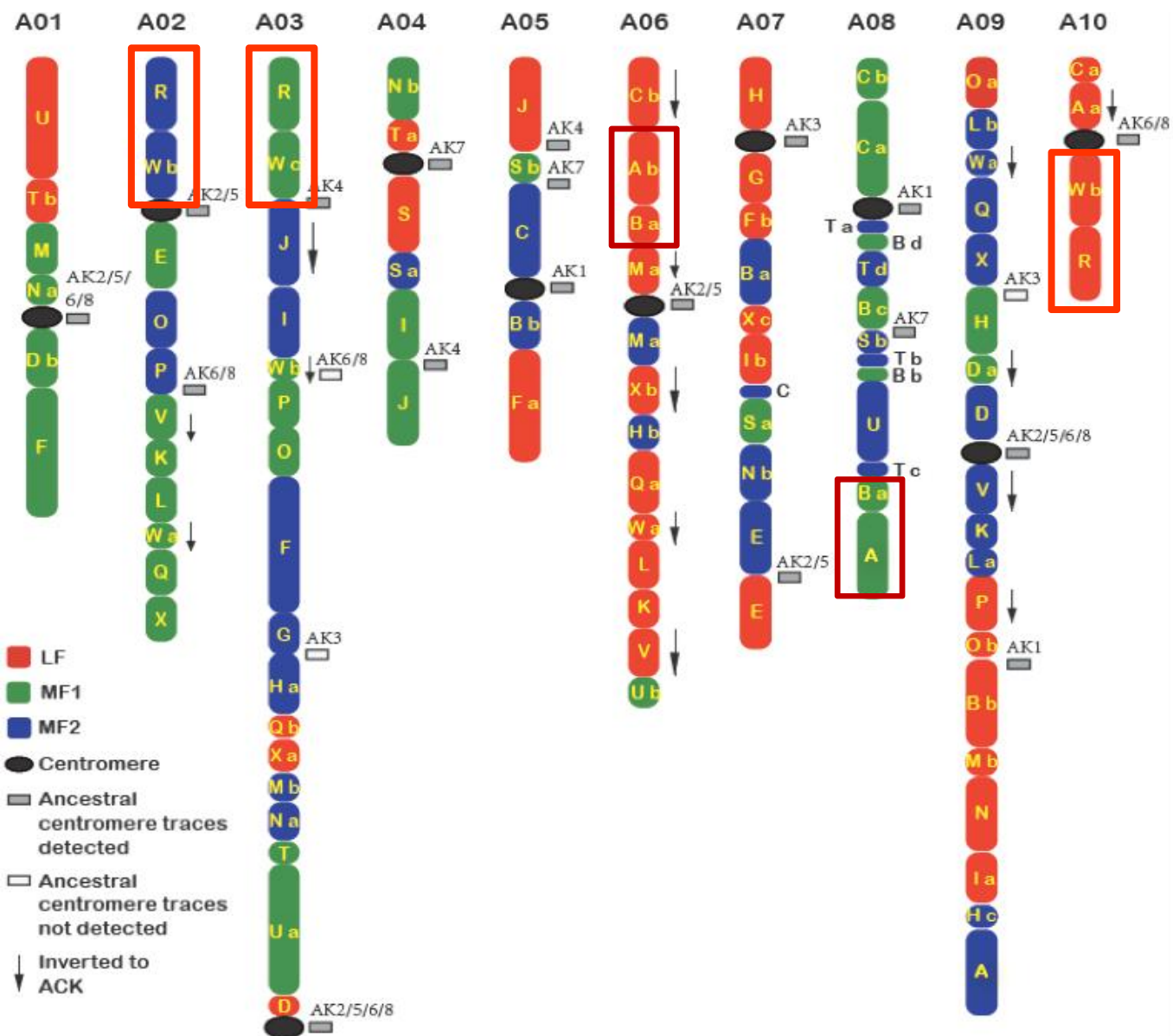
1)



2)







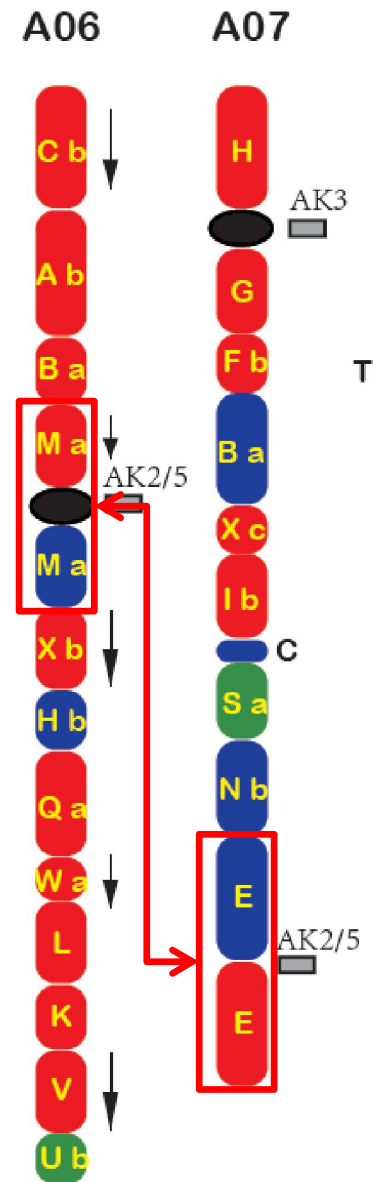
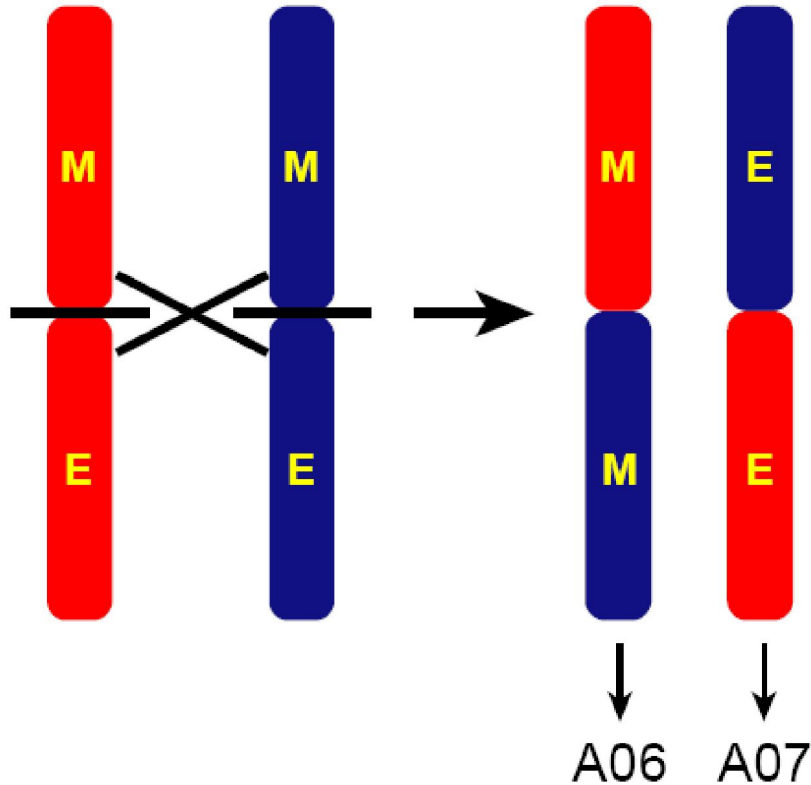
# Conserved block associations

ACK (n=8)				PCK (n=7)				tPCK (n=7)			
Chr <sup>a</sup>	GB asso	# <sup>c</sup>	<i>B. rapa</i> Chr	Chr	GB assoc.	#	<i>B. rapa</i> Chr	Chr	GB assoc.	#	<i>B. rapa</i> Chr
AK1	A-B	2	A06, A08	1	A-B	2	A06, A08	1	A-B	2	A06, A08
	B-C	2	A05, A08		B-C	2	A05, A08		B-C	2	A05, A08
AK2	D-E	0	-	2	D-E	0	-	2	N-M	3	A01, A03, A09
AK3	F-G	2	A03, A07	3	F-G	2	A03, A07	3	M-E	0	-
	G-H	3	A03, A07, A09		G-H	3	A03, A07, A09		F-G	2	A03, A07
AK4	I-J	2	A03, A04	4	I-J	2	A03, A04	3	G-H	3	A03, A07, A09
AK5	K-L	3	A02, A06, A09	5	S-T	2	A04, A08	4	I-J	2	A03, A04
	L-M	0	-		T-U	3	A01, A03, A08	5	S-T	2	A04, A08
	M-N	3	A01, A03, A09	O-P	3	A02, A03, A09	6	T-U	3	A01, A03, A08	
AK6	O-P	3	A02, A03, A09	6	P-W	1	A03	6	O-P	3	A02, A03, A09
	P-Q	0	-		W-R	3	A02, A03, A10		P-W	1	A03
	Q-R	0	-	N-M	3	A01, A03, A09	7	W-R	3	A02, A03, A10	
AK7	S-T	2	A04, A08	7	M-V	0	-	7	D-V	1	A09
	T-U	3	A01, A03, A08		V-K	3	A02, A06, A09		V-K	3	A02, A06, A09
AK8	V-W	0	-	7	K-L	3	A02, A06, A09	7	K-L	3	A02, A06, A09
	W-X	0	-		L-Q	3	A02, A06, A09		L-Q	3	A02, A06, A09
					Q-X	3	A02, A03, A09		Q-X	3	A02, A03, A09

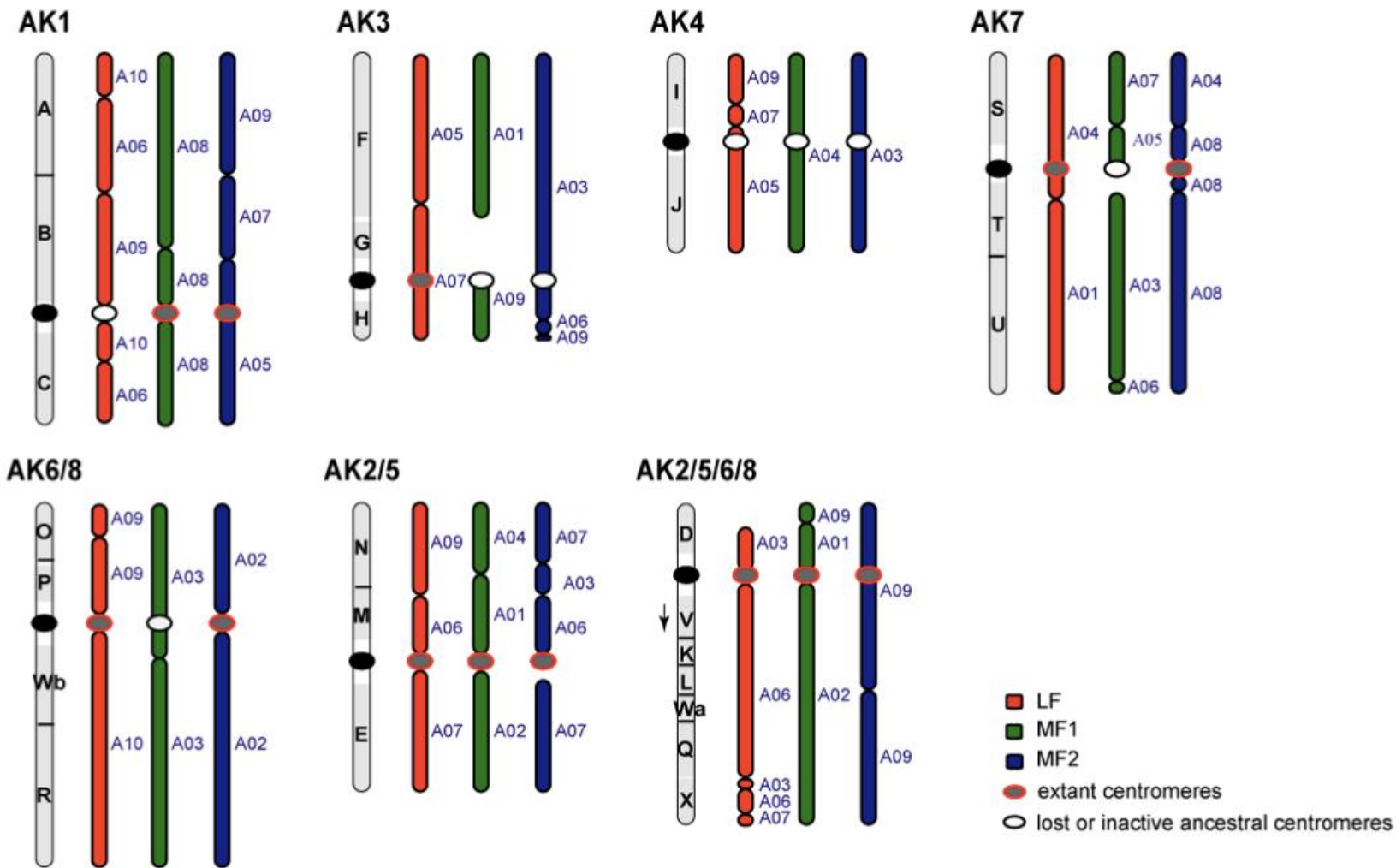
- 6 block associations in ACK not detected;
- 2 block associations (D/E, M/V) in PCK not detected;
- 1 block association (M/E) in tPCK not detected.



# Homologous reshuffling destroyed M/E association



# Reconstruction of three ancestral subgenomes in *B. rapa*



LF in red, MF1 in green, and MF2 in blue



# Paleocentromere evolution in *B. rapa*





# Traces of paleocentromeres in *B. rapa*

1) 10 centromeres of *B. rapa* **Vs.** 21 centromeres of tPCK;

2) Traces of 21 paleocentromeres (tPCK) in *B. rapa*.

**Repeat sequences of centromere for *Brassica* species:**

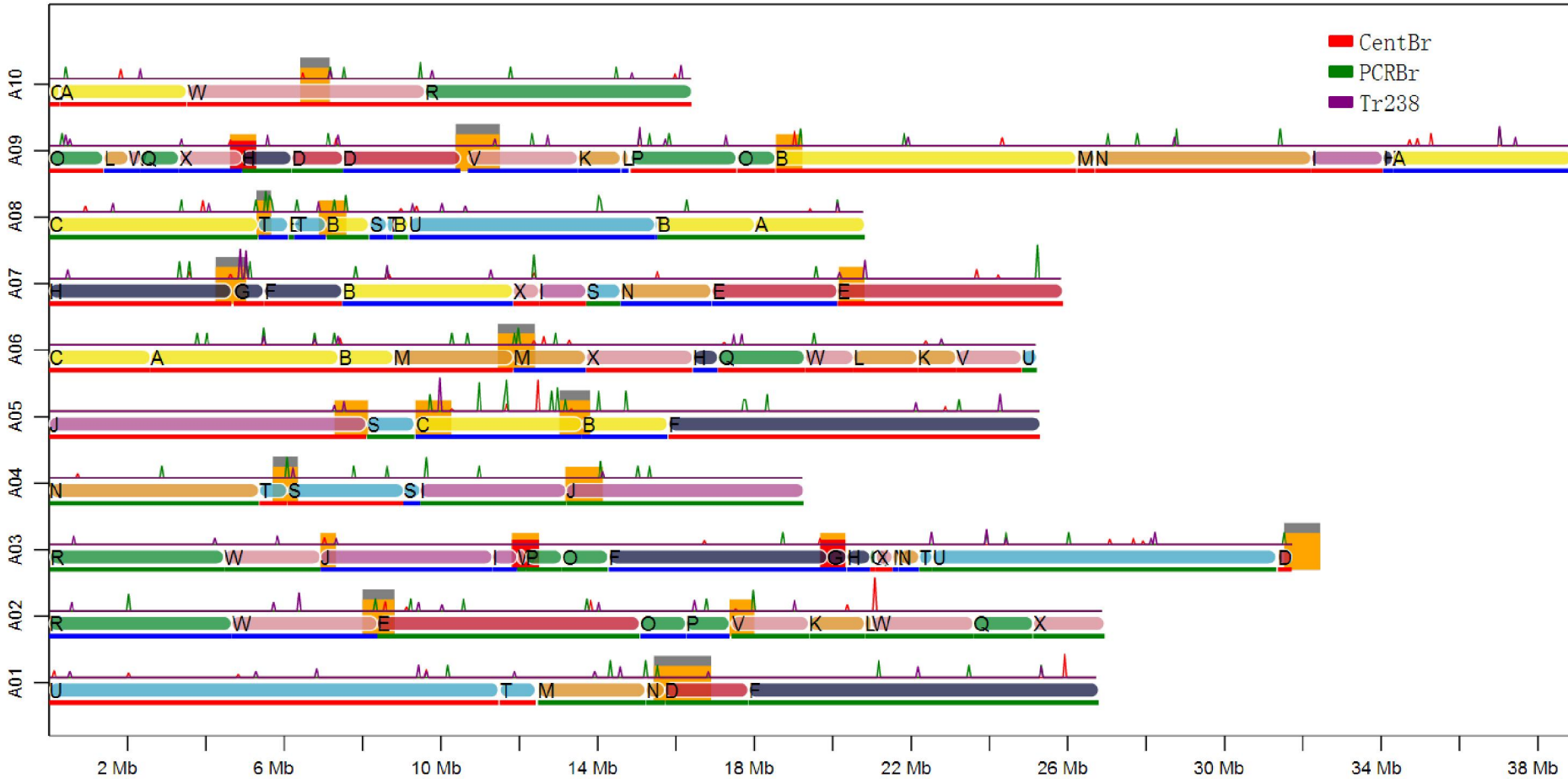
CentBr (Lim et al., 2005; Koo et al., 2011)

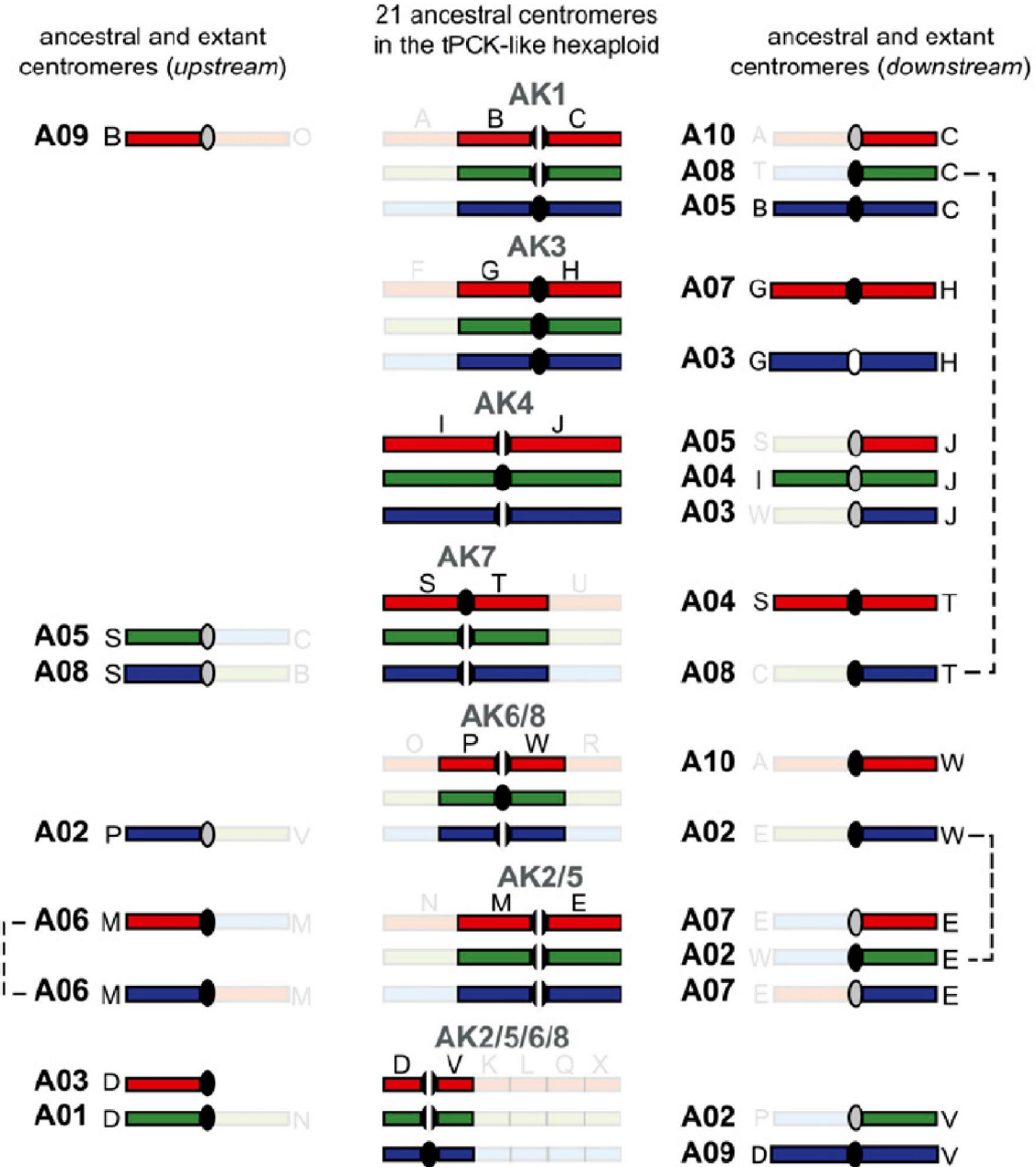
PCRBr, TR238 (Lim et al., 2007)

Nucmer **alignment**



# Traces of paleocentromeres

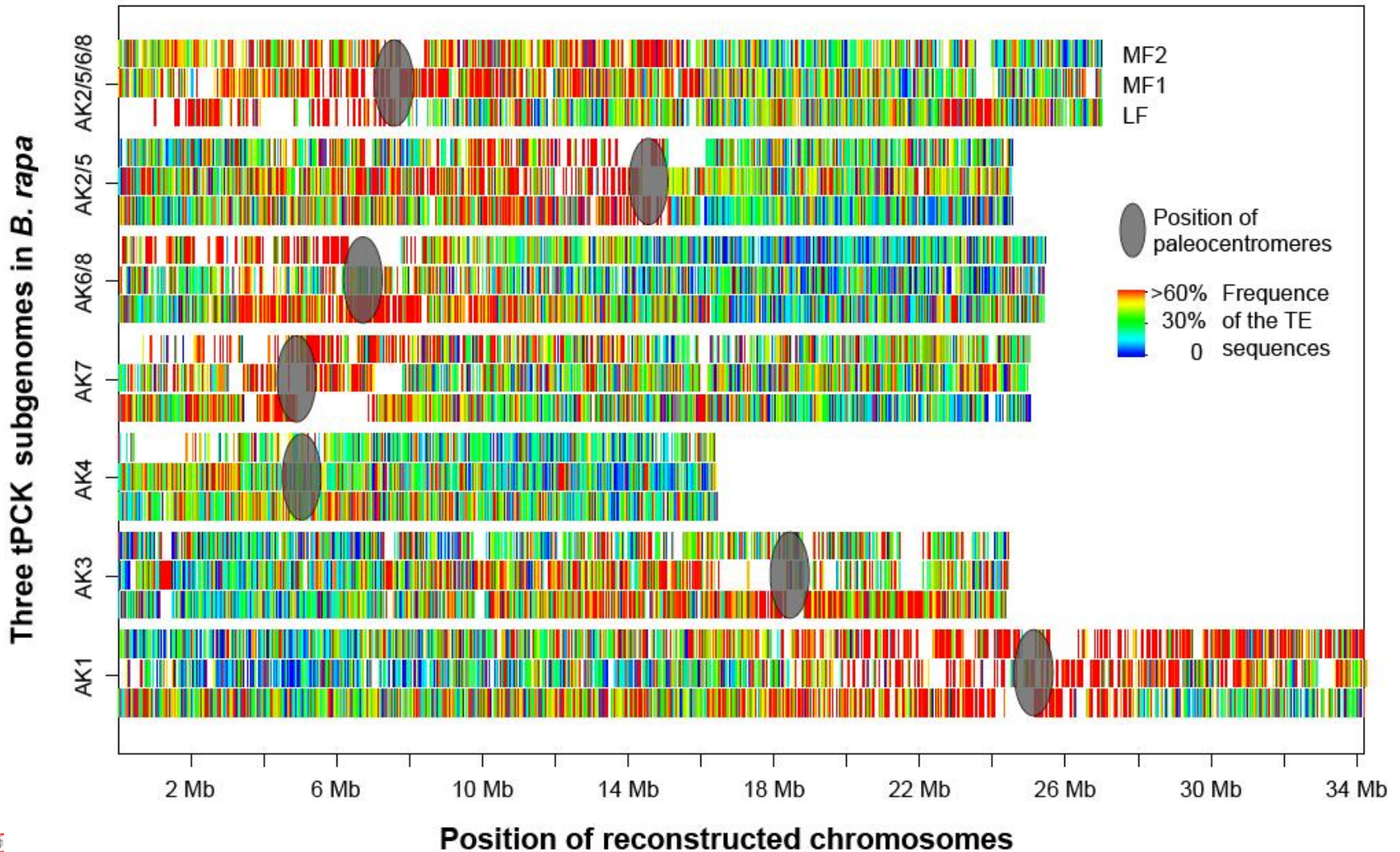




## 21 paleocentromeres of tPCK:

- 10: centromeres of *B. rapa*;
- 8: inactivated traces;
- 3: totally lost without traces.

# TEs support the regions of paleocentromeres

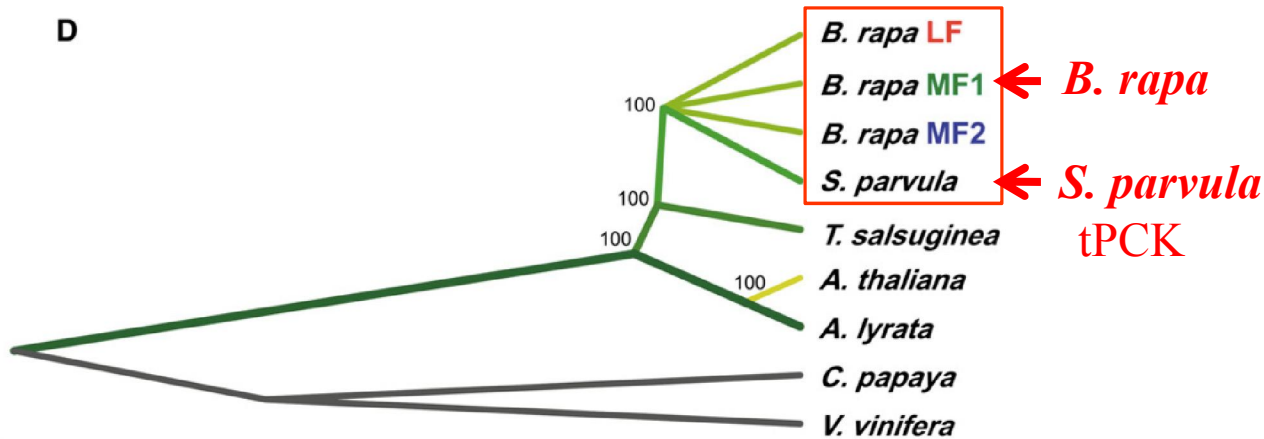
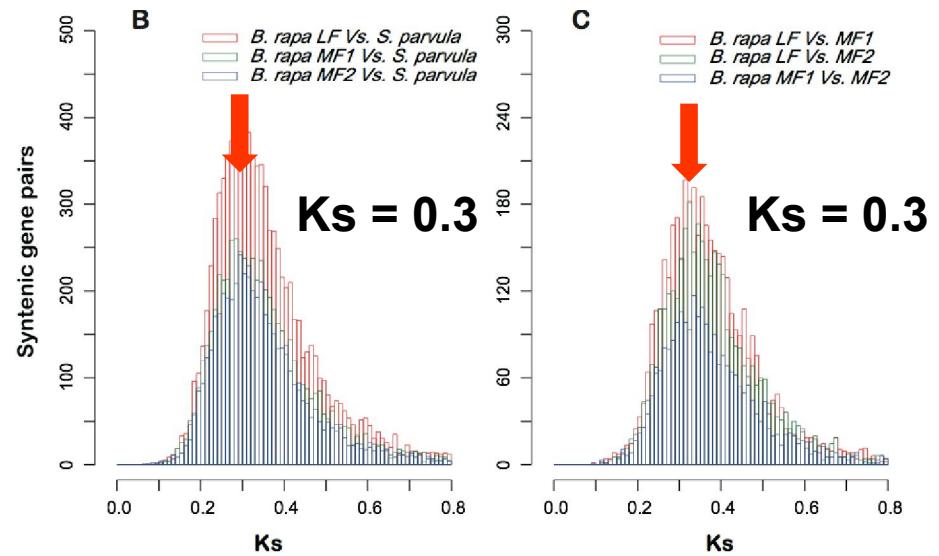
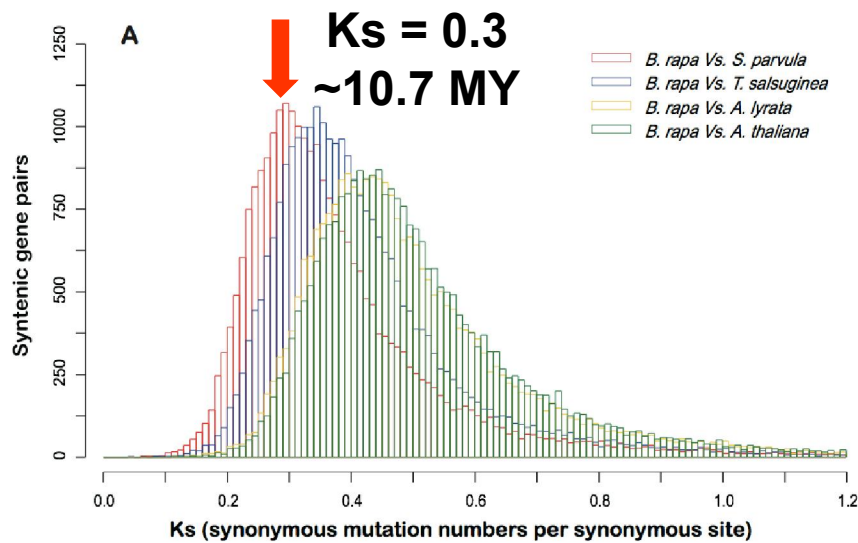


# Common ancestor of Brassiceae





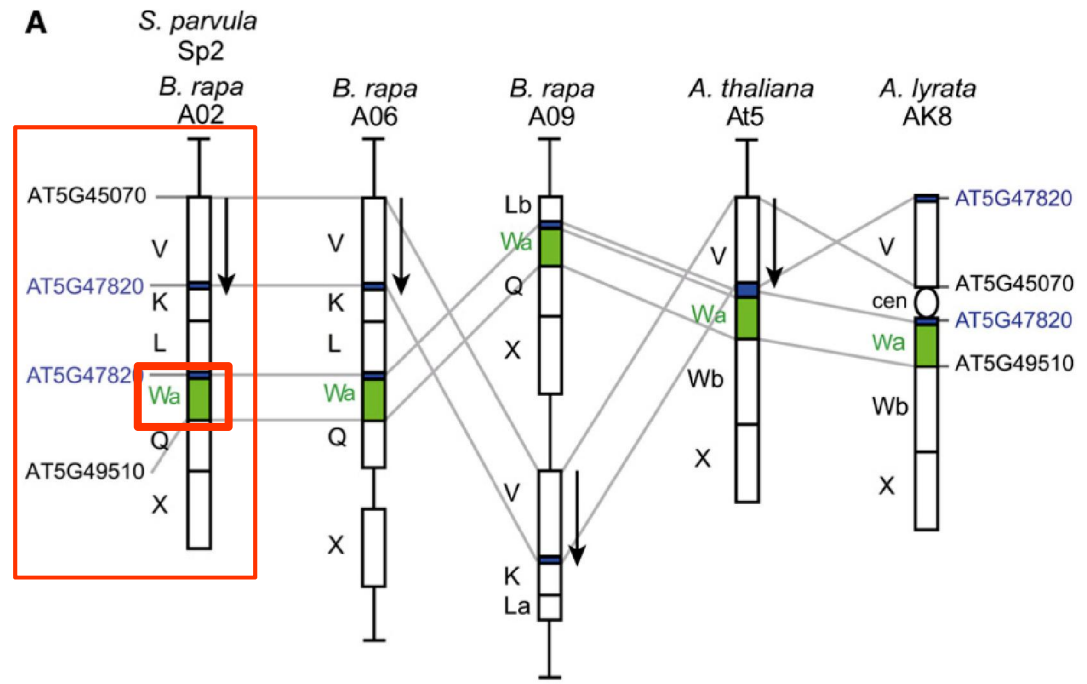
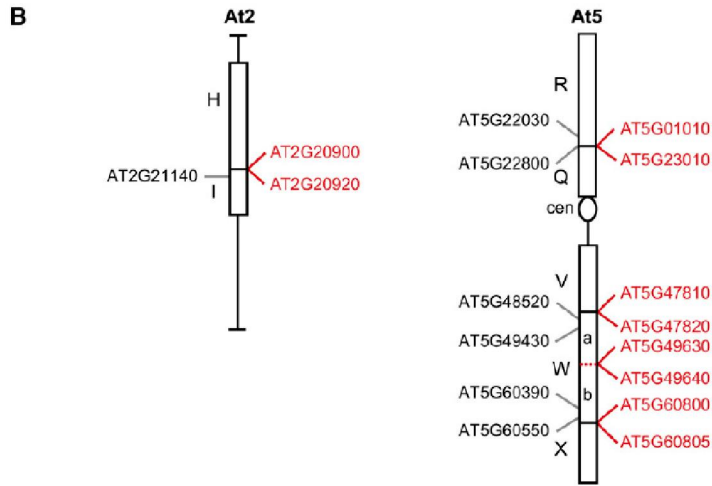
# Close relationship between *S. parvula* and *B. rapa*



591 genes were syntenic among the eight species, which were then used to constructed the NJ phylogeny tree to reveal their evolutionary relationship.



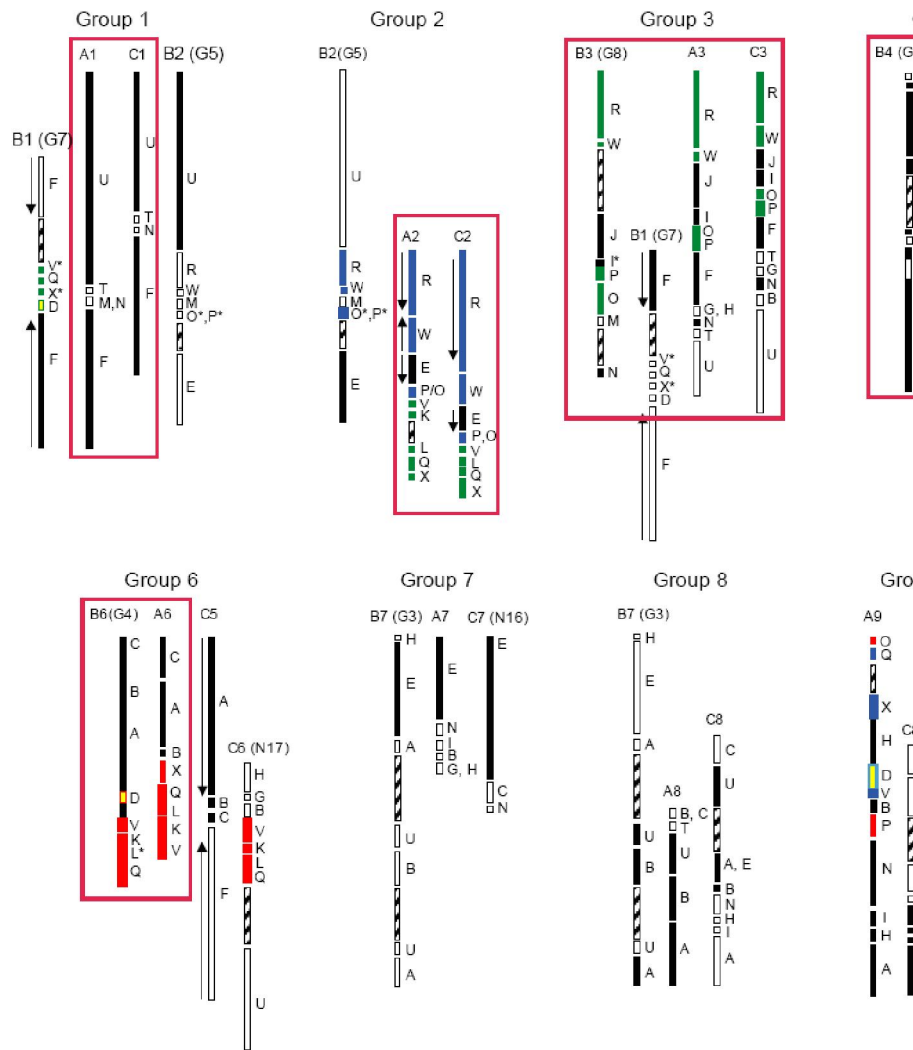
# tPCK specific block association



GB	AK chromosome	Interval	
<u>H</u>	3	AT2G15670	<b>AT2G20900</b>
<u>I</u>	4	<b>AT2G20920</b>	AT2G28910
<u>Q</u>	6	<b>AT5G28885</b>	<b>AT5G23010</b>
<u>R</u>	6	<b>AT5G23000</b>	<b>AT5G01010</b>
<u>V</u>	8	<b>AT5G42130</b>	<b>AT5G47810</b>
<u>W</u>	8	<b>AT5G47820</b>	<b>AT5G60800</b>
<u>X</u>	8	<b>AT5G60805</b>	AT5G67385



# Genetic map—block associations



*Raphanus sativus* (Shirasawa et al., 2011).

LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	LG9
S	S	F	N	E	E	A	O	C
N	F	B	I	W	W	B	J	A
E	H	A	J	P	R	U	W	C
N	F	M	I	O	P	S	R	A
E	D	N	S	M	O	C		C
	X	C	V	R	Q	K		B
	Q	H	L	X	L	V		A
	K	X	X	W	V			
	V	M	U	R	D			
	H	F		T				
	A			U				
	N							
	F							
	M							
	I							
	J							

**Other Brassicaceae species:**  
white mustard *S. alba*,  
*C. amplexicaulis*



# tPCK is the common ancestor of tribe Brassiceae

The Plant Cell, Vol. 25: 1541–1554, May 2013, www.plantcell.org © 2013 American Society of Plant Biologists. All rights reserved.

RESEARCH ARTICLES

The Plant Cell, 2013

## Deciphering the Diploid Ancestral Genome of the Mesohexaploid *Brassica rapa*<sup>CW</sup>



Feng Cheng,<sup>a</sup> Terezie Mandáková,<sup>b</sup> Jian Wu,<sup>a</sup> Qi Xie,<sup>c</sup> Martin A. Lysak,<sup>b,1</sup> and Xiaowu Wang<sup>a,1,2</sup>

<sup>a</sup> Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China

<sup>b</sup> Plant Cytogenomics, Central European Institute of Technology (CEITEC) and Faculty of Science, Masaryk University, CZ-625 00 Brno, Czech Republic

<sup>c</sup> State Key Laboratory of Plant Genomics, National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

The genus *Brassica* includes several important agricultural and horticultural crops. Their current genome structures were shaped by whole-genome triplication followed by extensive diploidization. The availability of several crucifer genome sequences, especially that of Chinese cabbage (*Brassica rapa*), enables study of the evolution of the mesohexaploid *Brassica* genomes from their diploid progenitors. We reconstructed three ancestral subgenomes of *B. rapa* ( $n = 10$ ) by comparing its whole-genome sequence to ancestral and extant Brassicaceae genomes. All three *B. rapa* paleogenomes apparently consisted of seven chromosomes, similar to the ancestral translocation Proto-Calepineae Karyotype (tPCK;  $n = 7$ ), which is the evolutionarily younger variant of the Proto-Calepineae Karyotype ( $n = 7$ ). Based on comparative analysis of genome sequences or linkage maps of *Brassica oleracea*, *Brassica nigra*, radish (*Raphanus sativus*), and other closely related species, we propose a two-step merging of three tPCK-like genomes to form the hexaploid ancestor of the tribe Brassiceae with 42 chromosomes. Subsequent diversification of the Brassiceae was marked by extensive genome reshuffling and chromosome number reduction mediated by translocation events and followed by loss and/or inactivation of centromeres. Furthermore, via interspecies genome comparison, we refined intervals for seven of the genomic blocks of the Ancestral Crucifer Karyotype ( $n = 8$ ), thus revising the key reference genome for evolutionary genomics of crucifers.





# Acknowledgement



**Institute of Vegetables and  
Flowers, Chinese Academy of  
Agricultural Sciences**

**Xiaowu Wang, Jian Wu, Feng Cheng,  
Jianli Liang, Bo Liu, Jishen Liu, Xiaobo  
Wang, Chao Sun  
Rifei Sun, Xixiang Li, Yongchen Du,  
Zhiyuan Fang**



**Plant Cytogenomics, CEITEC,  
Masaryk University, Czech**

**Martin A. Lysak  
Terezie Mandáková**



**中华人民共和国农业部**  
Ministry of Agriculture of the People's Republic of China



**中华人民共和国科学技术部**  
The Ministry of Science and Technology of the People's Republic of China

**Thank you!**

