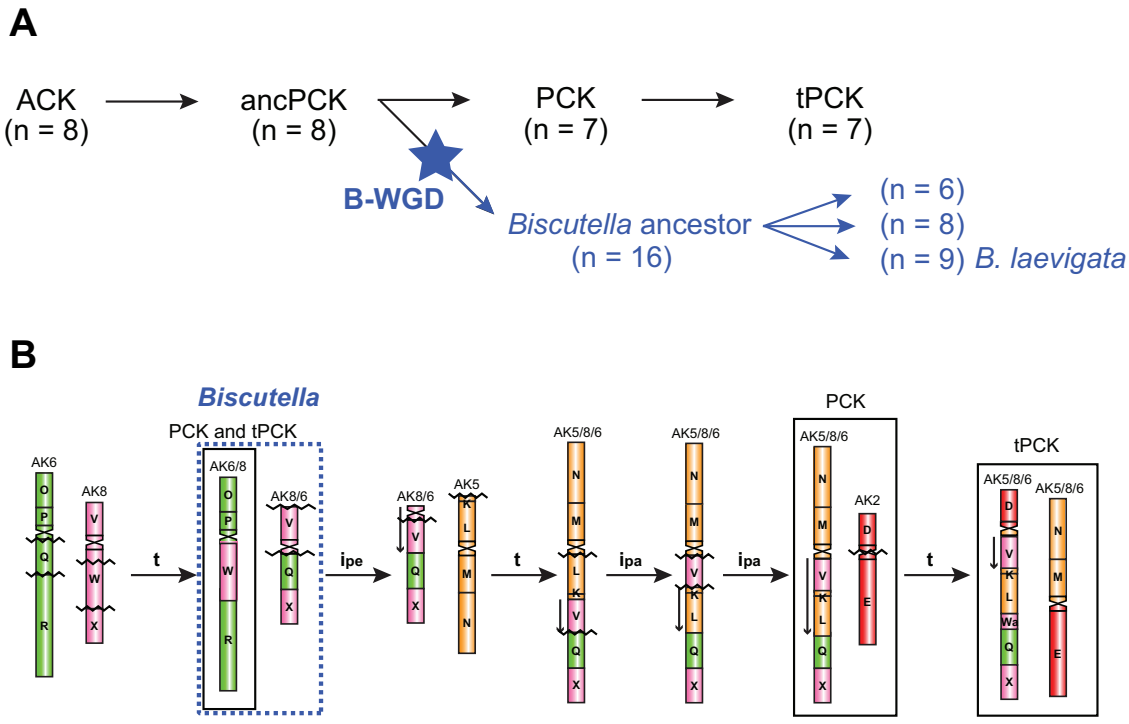
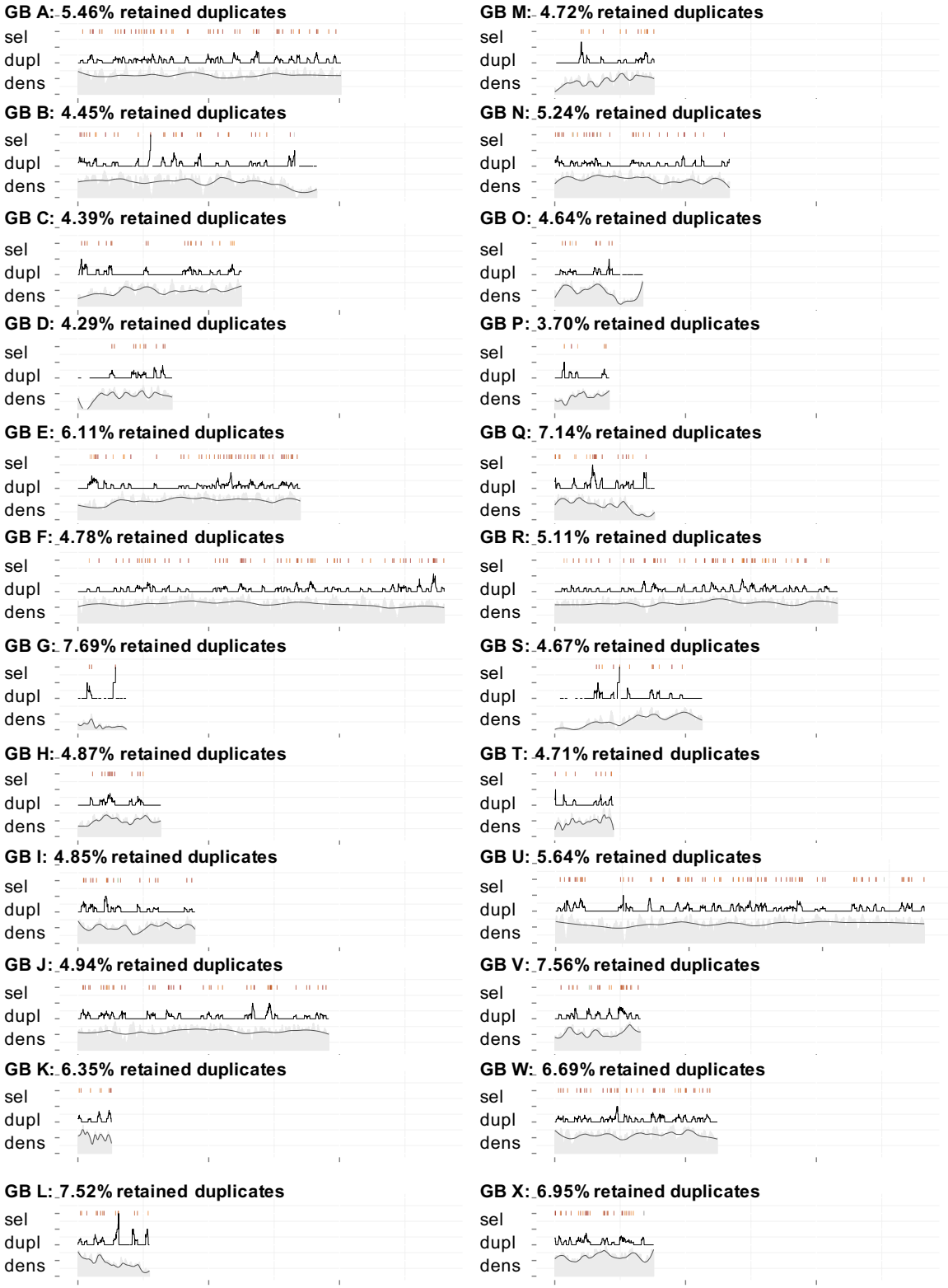
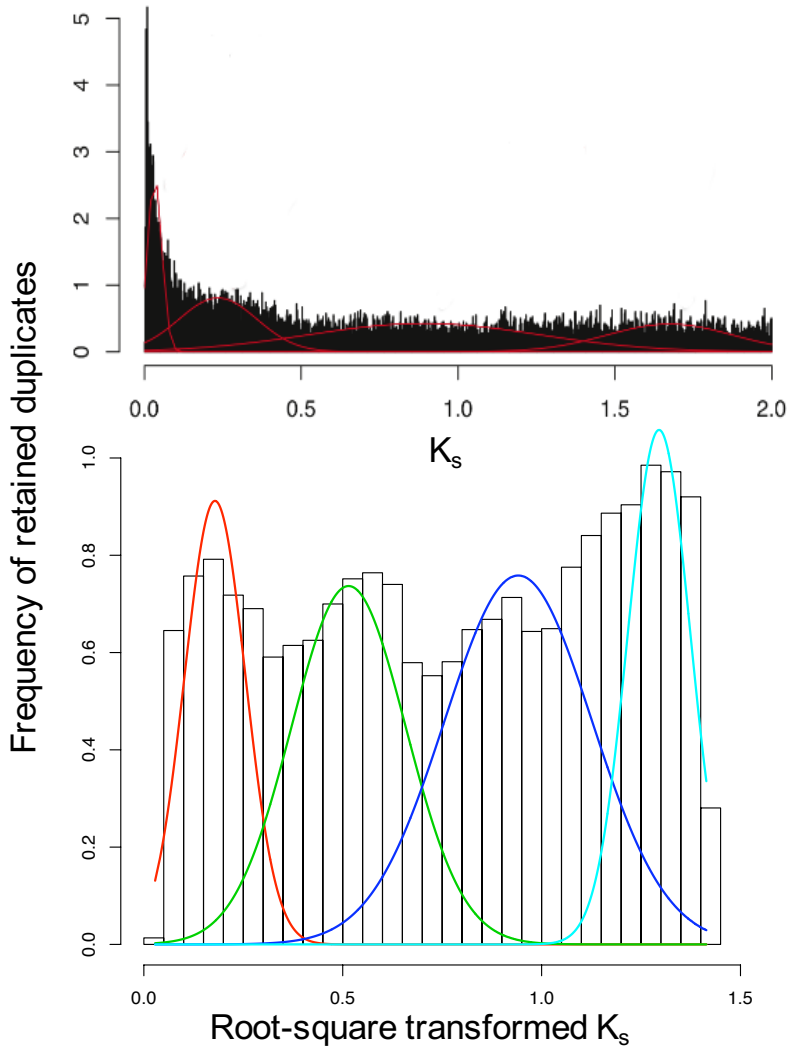


Supplemental Figure 1: Tentative scenario of genome evolution in *Biscutella* from the ancestral crucifer karyotype (ACK). (A) The ancestral proto-Calepineae karyotype (PCK; i.e. ancPCK; n = 8) presumably diverged from ACK through rearrangements involving chromosomes AK6 and AK8. This n = 8 genome has undergone the BI-m-WGD event leading to an ancestral mesotetraploid *Biscutella* genome (n = 16), altered by genome fractionation and descending dysploidies towards the modern diploidized genomes of *Biscutella* species with n = 9, 8, and 6. In the extended lineage II, the ancPCK has undergone chromosome rearrangements leading to the origin of PCK and tPCK genomes (n = 7). (B) Chromosome rearrangements underlying the origin of ancestral genomes of ancPCK/Biscutella (n = 8), PCK (n = 7) and tPCK (n = 7) from the older ACK genome (n = 8). *Biscutella* chromosome Bv5 (O1/P1/W1/R1) and several ACK-like genomic block associations link the modern *B. laevigata* genome with both the ACK (n = 8) and the ancestor of the PCK (n = 7). A reciprocal translocation between ACK-specific chromosomes AK6 and AK8 led to the origin of AK6/8 chromosome present in *Biscutella* as well as in PCK and tPCK. Subsequent rearrangements involving chromosomes AK8/6, AK5 and AK2 led to the further differentiation of PCK and tPCK genome, respectively. t: translocation, ipe: pericentric inversion, ipa: paracentric inversion.



Supplemental Figure 2: Transcriptome data of *Biscutella leavigata* along each genomic block (GB) A to X. The density of transcripts (dens) and of retained duplicates (dupl) was computed in sliding windows of 40 Arabidopsis genes. Evidence of soft (light red) to strong (dark red) purifying selection (sel) based on the K_a/K_s ratio is indicated for each retained duplicate. % retained duplicates is the proportion of unigenes present in the transcriptome that showed evidence of Bl-m-duplicates.



WGD event	**n**	**m**	**α**	**β**
Mean K_s	0.032 ± 0.001	0.235 ± 0.003	0.886 ± 0.010	1.678 ± 0.006
Standard deviation	0.023 ± 0.001	0.121 ± 0.003	0.366 ± 0.010	0.198 ± 0.004

Supplemental Figure 3: Distribution of synonymous substitutions (K_s and root-square transformed K_s) between pairs of duplicated transcripts in Buckler Mustards *(Biscutella laevigata* subsp. *laevigata*; 2n = 4x = 36). The curves represent a mixture of normal distributions fitted to the overall K_s distribution of transcripts larger than 450 bp with $0 \leq K_s \leq 2$. Mixture models on both root-squared K_s and K_s yielded similar results. The table (Table 1) presents means and standard errors of significant peaks based on K_s after 100 bootstrap computations.

Supplemental Table 1: Sequenced cDNA libraries of *Biscutella laevigata* subsp *leavigata*, using Roche 454 GS FLX Titanium and SOLiD 5500 xl

Library	Organs	Individuals#	Number of reads
454 (standardized)	Leaves & roots	Ab	270,469
SOLiD1 (quantitative)	Leaves & roots	Ca	17,123,098
SOLiD2 (quantitative)	Leaves & roots	Bj	20,662,107
SOLiD3 (quantitative)	Leaves	Ab	16,621,814
SOLiD4 (quantitative)	Leaves	Ab	14,131,769
SOLiD5 (quantitative)	Roots	Bd	10,182,227
SOLiD6 (quantitative)	Roots	Bd	35,266,604

#Selected from Parisod & Christin, 2008, New Phytologist 178: 436–447

Supplemental Table 2: Iterative hybrid assembly of cDNA libraries detailed in Table S1 to infer the *Biscutella laevigata* subsp *laevigata* transcriptome, with summary statistics of resulting contigs.

Assembly step	Total number of contigs	Number of large contigs[a]	Largest contig (bp)[b]	Coverage long reads[b,c]	Coverage short reads[b,c]
1	404,105	15,772	3,692	4.75	24.64
2	402,172	20,036	5,402	4.31	25.29
3	333,770	16,874	4,118	4.96	10.45
4	347,677	22,214	6,236	4.06	18.03
5	313,668	22,711	4,305	4.06	12.18
6	1,134,879	24,462	5,282	3.71	15.83

[a]Larger than 1000 bp

[b]Among contigs assembled during the corresponding step

[c]Mean coverage of long and short reads produced by Roche 454 and SOLiD, respectively

Supplemental Table 3: Overall retention bias of *Biscutella laevigata* transcripts highlighted by a selection of significantly (α= 0.05) over-represented GO categories among unigenes that did not group in gene families (with negative odds-ratios) vs that presented several transcripts (with positive odds-ratios) in gene families.

GO-slim	Log(odds ratio)
DNA recombination	-2.31
DNA dependent DNA replication	-2.22
DNA repair	-0.79
response to DNA damage stimulus	-0.75
DNA metabolic process	-0.6
response to light stimulus	0.53
response to temperature stimulus	0.64
response to UV	0.81
response to osmotic stress	0.9
response to metal ion	0.95
defense response to fungus	0.95
response to salt stress	0.96
response to high light intensity	0.98
response to UVB	1.1
regulation of photosynthesis	1.41
regulation of photosynthesis light reaction	1.66
photoinhibition	2.22
postreplication repair	2.44

Supplemental Table 4: Genomic blocks of the ancestral crucifer karyotype and corresponding *Arabidopsis thaliana* border BAC clones

Ancestral chromosome (ACK)	*Arabidopsis thaliana* chromosome	Genomic block	Block borders			
			BAC clone	GenBank Accession	BAC clone	GenBank Accession
AK1	AT1	A	T25K16	AC007323	- F18O14	AC025808
	AT1	B	F6F9	AC007797	- F12K21	AC023279
	AT1	C	F2J6	AC009526	- T6H22	AC009894
AK2	AT1	D	F13N6	AC058785	- T12P18	AC010852
	AT1	E	T23K8	AC007230	- F23A5	AC011713
AK3	AT3	F	T4P13	AC008261	- MWL2	AB025639
	AT2	G	F16J10	AC007289	- T25N22	AC005693
	AT2	H	T10F5	AC007063	- F5H14	AC006234
AK4	AT2	I	F7O24	AC007142	- T9J22	AC002505
	AT2	J	F18A8	AC003105	- T8I13	AC002337
AK5	AT2	K	F2I9	AC005560	- F3C11	AC007167
	AT3	L	K9I22	AP000599	- T4A2	AP002066
	AT3	M	T10D17	AL353865	- F3A4	AL132978
	AT3	N	F24M12	AL132980	- F16M2	AL138648
AK6	AT4	O	F6N15	AF069299	- T1J1	AF128393
	AT4	P	T3H13	AF128396	- T4C9	AL080318
	AT5	Q	K8E10	AB025618	- T8M17	AF296835
	AT5	R	F7J8	AL137189	- T6G21	AL090689
AK7	AT5	S	F5H8	AB025605	- MPK23	AB020748
	AT4	T	T1P17	AL049730	- F18A5	AL035528
	AT4	U	T6K21	AL021889	- T5J17	AL035708
AK8	AT5	V	MBD2	AB008264	- MGC1	AB028612
	AT5	W	K16F13	AB024025	- K9B18	AB015471
	AT5	X	MUF9	AB011483	- K9I9	AB013390

Supporting Analysis: Phylogenetics of gene pairs under whole genome duplications (WGDs)

Phylogenetic trees for a subset of the analyzed genes confirm that pairs of *Biscutella* transcripts are predominantly derived from the α-WGD and the Bl-m-WGD events.

Duplicated transcripts from *Biscutella laevigata* (Bl) and their orthologs from *Arabidopsis thaliana* (At) and the outgroup *Cleome spinosa* (Cs: GenBank LIBEST_024987) were grouped into gene families and analyzed following the same procedure as detailed in Material and Methods. Among gene families including transcripts from all three species, corresponding ML trees containing Bl-duplicates analyzed here were pruned for duplicates showing a maximum Ks=0.9 to focus on loci investigated here. Bayesian inferences were performed in MRBAYES 3.2.2 (http://mrbayes.sourceforge.net/) on 159 of the resulting families including one transcript from Cs and at least one transcript from At and two from Bl. Using Cs as an outgourp, two independent analyses with four sequentially heated chains (temperature set at 0.02) were run for 10 million generations. The initial 20% of sampled trees were discarded (burn-in) and a tree was sampled every 5000 generations to compute the maximum clade credibility tree. Resulting topologies were sorted, taking uncertainty into account, with the R package APE (http://ape-package.ird.fr).

Under the multispecies coalescent and a rooted species tree topology ((Bl,At),Cs), the coalescence ages for Bl-m-duplicates (Bl-1 and Bl-2) is expectedly shallower than the speciation of At. Gene evolution under recurrent WGDs thus leads to trees including a topology congruent with the one of the species tree. Among collected trees, this topology occured twice when all α- and m-duplicates have been retained (Figure A). Loss of specific duplicates from Bl will always lead to trees including this topology (referred to as Topology A). Gene families showing a discrepant topology were further examined. Only differential retention of α-duplicates between At and Bl lead to discordant topologies such as ((Bl1,At),Bl2),Cs) that match our hypothesis (referred to as Topology B; Figure B). The majority of analyzed gene families (72.3%) presented such topologies congruent with recurrent WGDs. Remaining trees presented variable topologies (an example in Figure C) and revealed genes having unlikely evolved strictly under WGDs. Such cases are collectively referred to as Topology C.

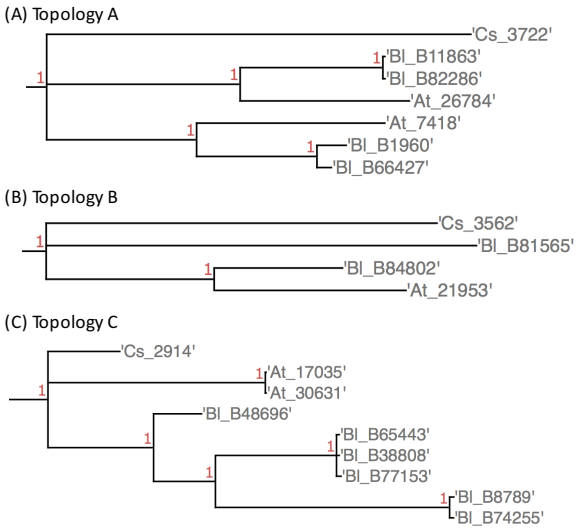


Figure: Examples of gene tree topologies observed among transcripts from *Cleome spinosa* (Cs), *Arabidopsis thaliana* (At) and *Biscutella laevigata* (Bl) grouped into gene families. Following our hypothesis of gene duplication through the α- and, then, the Bl-m-WGD events, coalescence of Bl-m-duplicates is expectedly shallower than the speciation of At. Corresponding gene trees thus includes a topology congruent with the species tree ((Bl,At),CS).

(A) Topologies congruent with the species tree. This particular example shows all duplicates in both At and Bl retained. Loss of a particular At or Bl transcript leads to congruent topologies.

(B) Discrepant topologies matching the species tree. In this case, one of the At-α-duplicates, but not all corresponding Bl-duplicates, may have been lost after the α-WGD event.

(C) Discrepant topologies failing to be reconciled with the species tree, rejecting gene evolution strictly under recurrent WGDs in Bl.

Table: Proportions of the 159 gene families showing phylogenetic trees congruent with the three main topologies presented in the figure and matching the hypothesis of recurrent WGDs

	Number of congruent trees	Proportion	Matching hypothesis
Topology A	86	54.1%	
Topology B	29	18.2 %	72.3%
Topology C	44	27.7%	