

New Phytologist Supporting Information

Article title: **Rye B chromosomes encode a functional Argonaute-like protein with *in vitro* slicer activities similar to its A chromosome paralog**

Authors: Wei Ma, Tobias Sebastian Gabriel, Mihaela Maria Martis, Torsten Gursinsky, Veit Schubert, Jan Vrána, Jaroslav Doležal, Heidrun Grundlach, Lothar Altschmied, Uwe Scholz, Axel Himmelbach, Sven-Erik Behrens, Ali Mohammad Banaei-Moghaddam and Andreas Houben

Article acceptance date: 18 June 2016

The following Supporting Information is available for this article:

Fig. S1 Phylogenetic analysis of kinesin KIF (a), Shortage in chiasmata SHOC1 (b) and Argonaute AGO (c).

Fig. S2 Gene structure model for three selected genes based on *in silico* identification.

Fig. S3 Workflow representation of how to distinguish A- and B- located sequences.

Fig. S4 B-located ScKIF4A-like fragments where subjected to structural modifications.

Fig. S5 Chromosomal location of ScKIF4A is outside of the non-disjunction controlling region.

Fig. S6 CAPS analysis for A- and B-derived ScKIF4A transcripts.

Fig. S7 CAPS analysis for A- and B-derived ScSHOC1 transcripts.

Fig. S8 Gene structure model of A- and B-derived ScAGO4B transcripts.

Fig. S9 Domain structure of ScAGO4B protein.

Table S1 List of primer sequences for PCR, RT-PCR, FISH probe preparations, *in vitro* transcription and quantitative RT-PCR.

Table S2 List of sequence identifiers and description of KIFs, SHOC1 and AGOs sequences used for phylogenetic tree construction.

Table S3 Accession numbers of *ScKIF4A*, *ScSHOC1* and *ScAGO4B* genes.

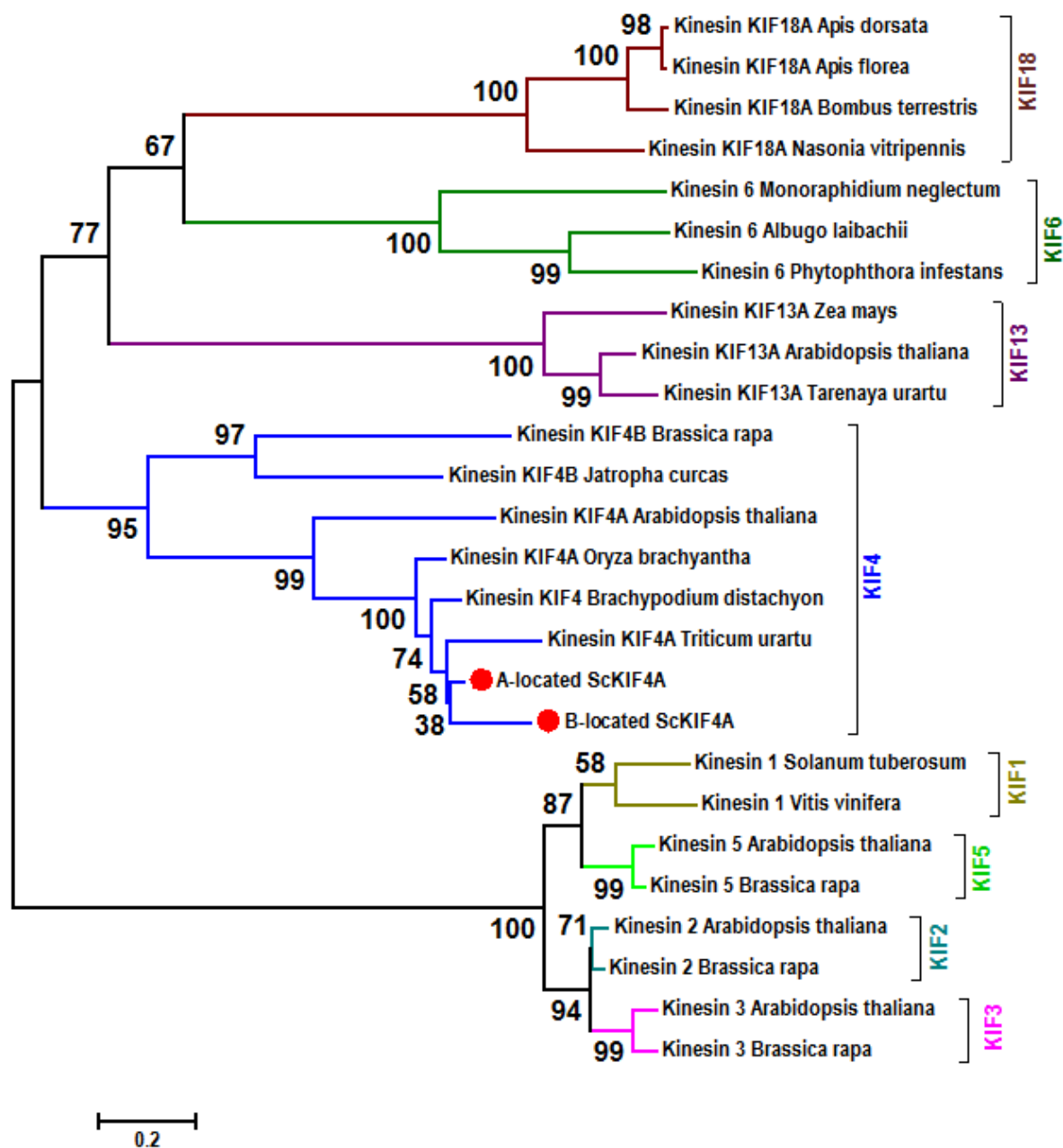
Table S4 Categories of B-located transcripts.

Tables S5 and S6 The functional annotation and classification of B chromosome-located gene candidates in vegetative (root and leaf) tissue (**Table S5**) and (anther) tissue (**Table S6**).

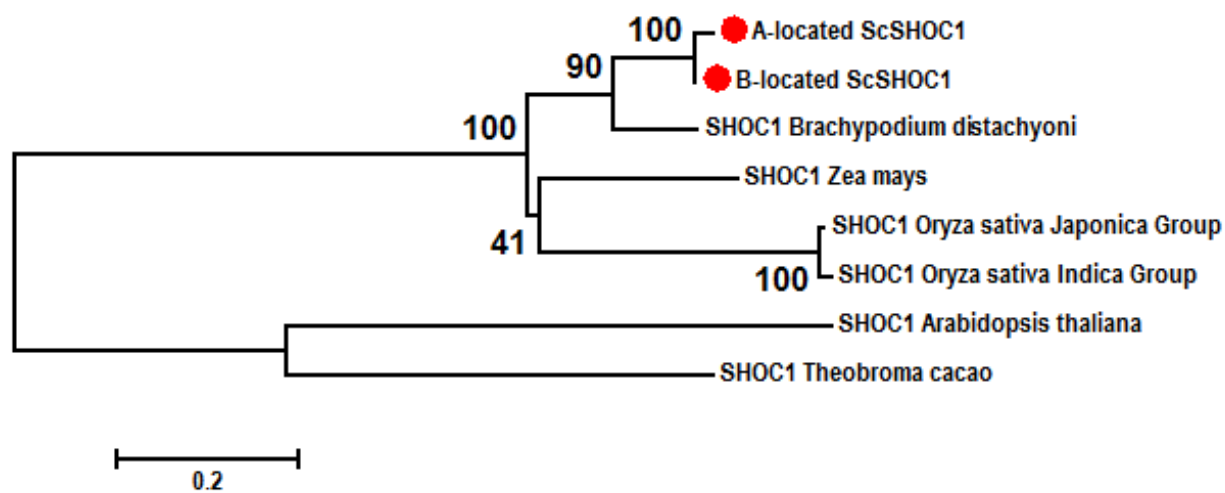
Methods S1 Material and Methods.

Fig. S1 Phylogenetic analysis of kinesin KIF (a), Shortage in chiasmata SHOC1 (b) and Argonaute AGO (c). Sequence comparison performed for kinesin KIF (a), Shortage in chiasmata SHOC1 (b) and Argonaute AGO (c) from different species. After aligning the sequences with Clustal W, phylogenetic trees were calculated using the boot-strapped neighbor-joining algorithm in MEGA 6.06 with 1000 trials (<http://www.megasoftware.net/>). Bootstrap values are indicated as percentages of the 1000 trials at their respective node. For convenience rye A- and B-derived sequences are indicated with red circles before the name.

(a)



(b)



(c)

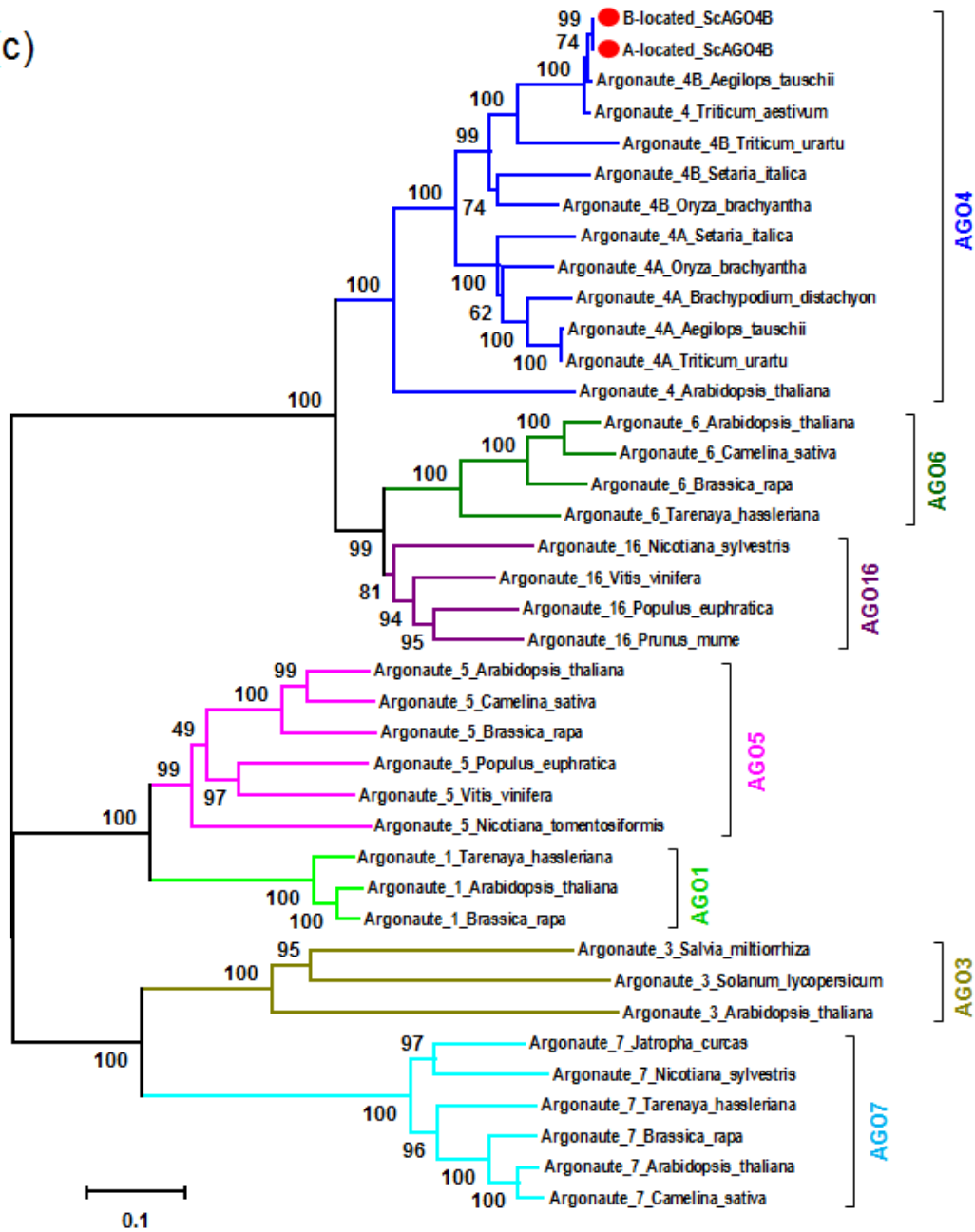


Fig. S2 Gene structure model for three selected genes based on *in silico* identification. The black and orange boxes correspond to the A- located introns and individual exons, respectively. Grey boxes with numbers represent the B-located genomic contigs. The regions of missing B-located genomic sequences are indicated in black vertical lines. The similarity (%) between A-located sequences and B-located counterparts, positions of start/stop codon, primer sites and CAPS markers are indicated.

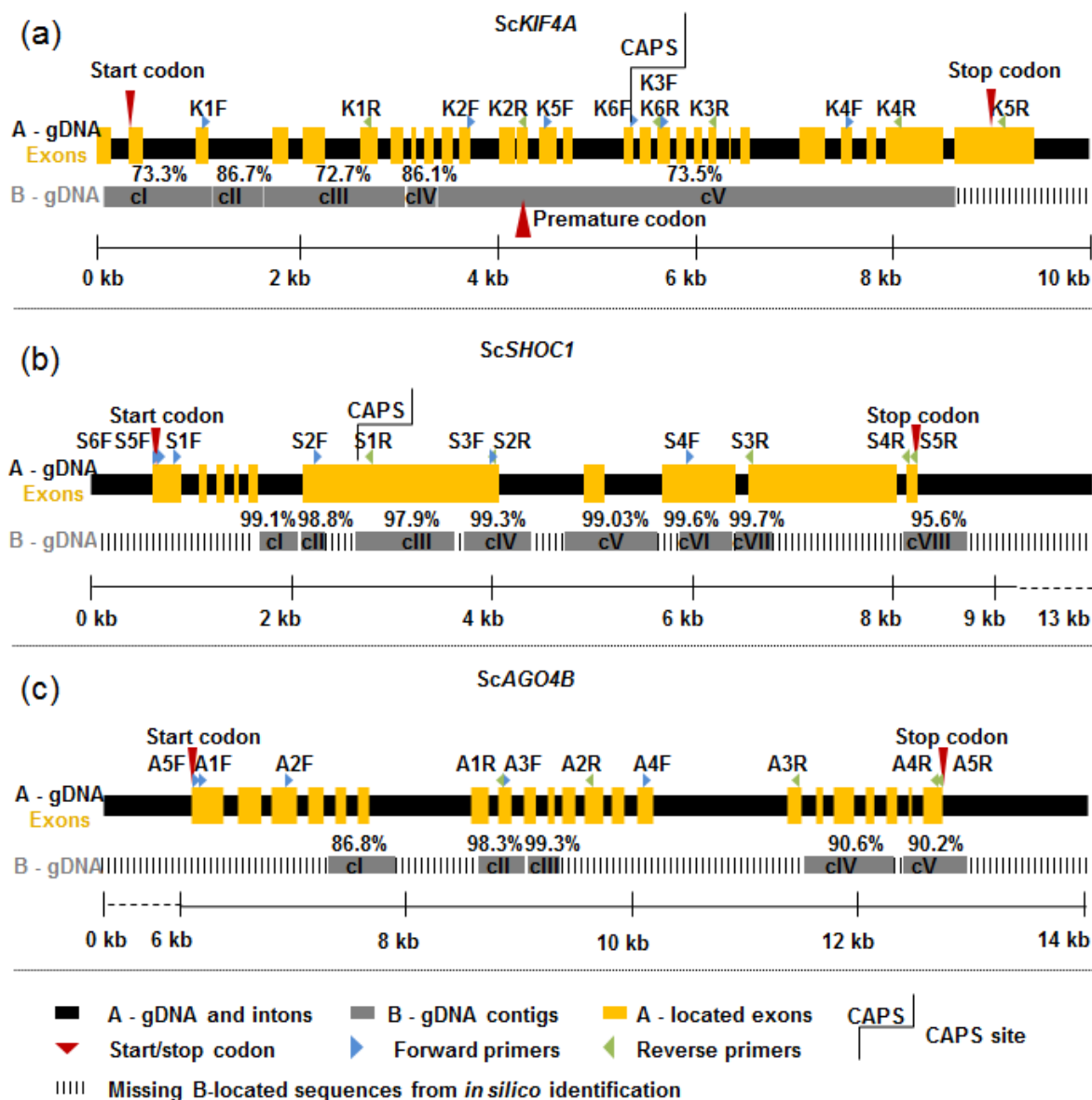


Fig. S3 Workflow representation of how to distinguish A- and B- located sequences. Three criteria are categorized according to PCR and gel electrophoresis (Step 1), B-located sequences for category (b) and (c) are able to be distinguished due to fragment size differences from A-located counterparts. For category (a), B-located sequences are able to be distinguished only after sequencing analysis (Step 2). Alignments of informative sequenced *ScKIF4A* clones derived from gDNA of rye and wheat with Bs (+B) as well as without Bs (0B) (Step 2). A- and B-located *ScKIF4A* sequences are able to be distinguished by polymorphic sites (CAPS marker position used to distinguish A- and B- located *ScKIF4A* are indicated).

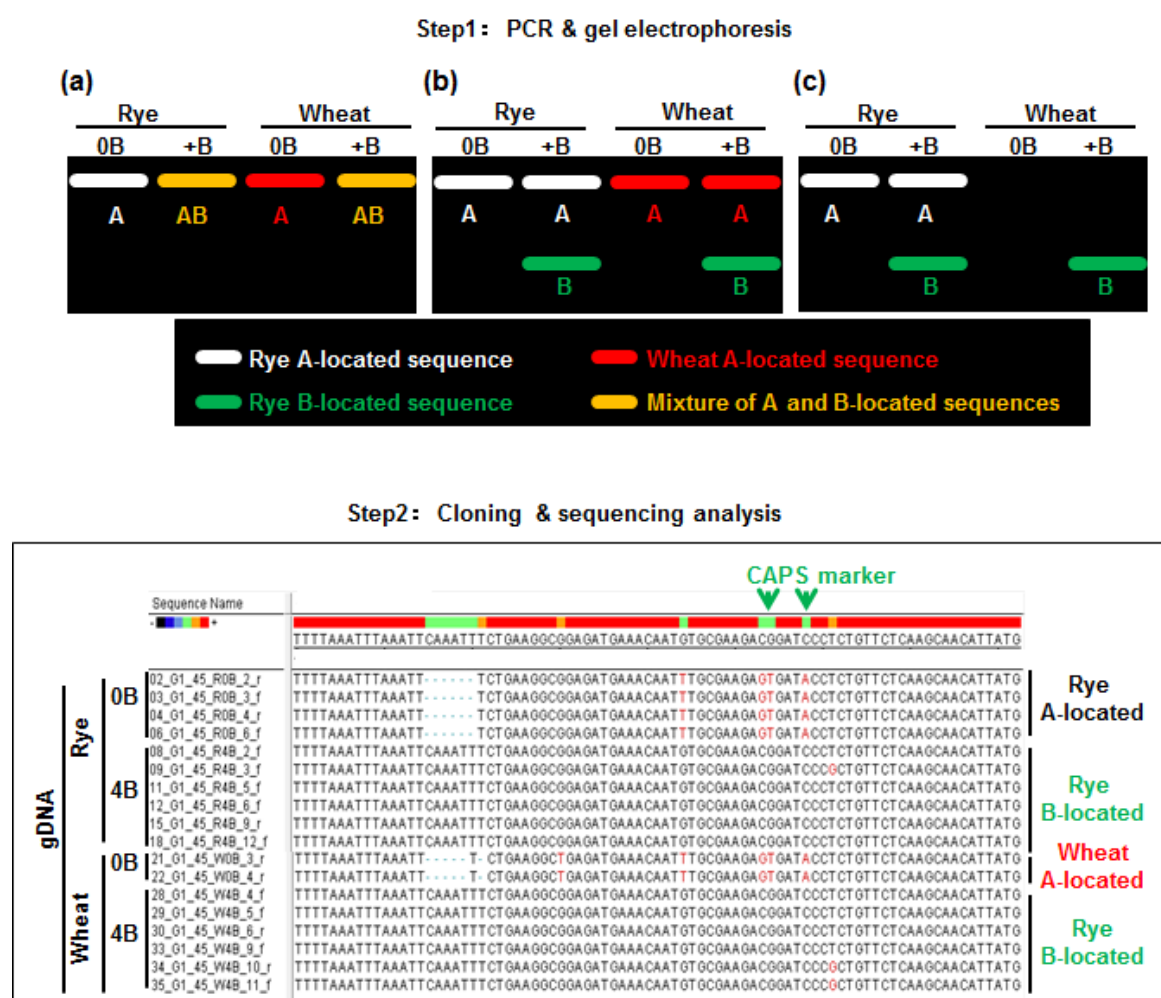


Fig. S4 B-located *ScKIF4A*-like fragments were subjected to structural modifications. PCR performed on genomic DNA of rye and wheat without and with B chromosomes. Schemata show the polymorphic sites between two types of rye B-located *ScKIF4A*-like fragments and their A-located counterpart according to the sequencing results.

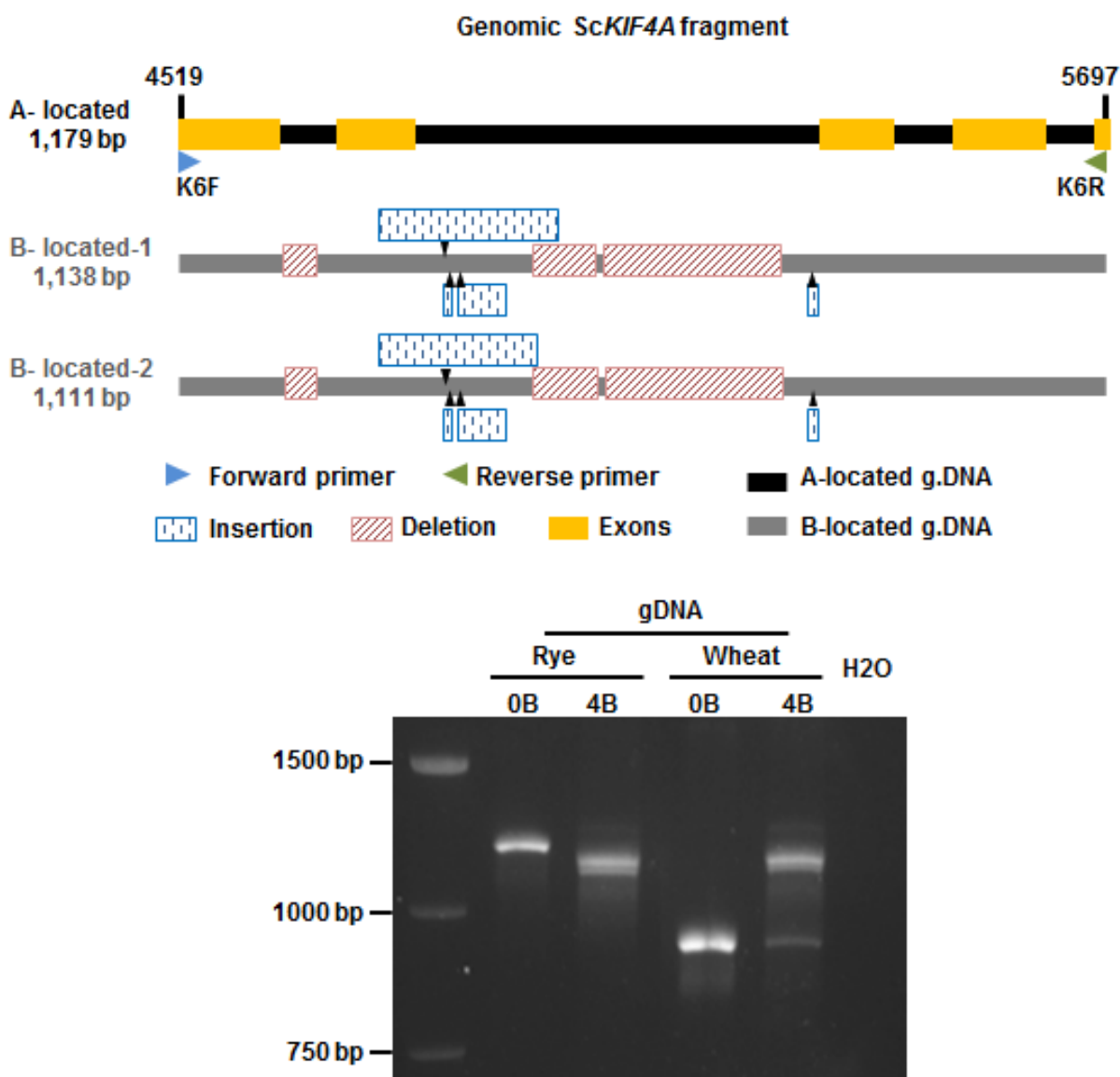


Fig. S5 Chromosomal location of *ScKIF4A* is outside of the non-disjunction controlling region. Meiotic metaphase I cell of rye with one deficient B chromosome after FISH with labeled *ScKIF4A* (in red). FISH with the B-specific *Sc55c1* repeat (in green) allowed marking the border of the non-disjunction controlling region of rye Bs. Chromosomes are stained by DAPI (in blue). Arrowheads, signals from Bs; arrows, A-localized FISH signals. The inset shows the signals on further enlarged A and B chromosomes. Bar, 5 μ m.

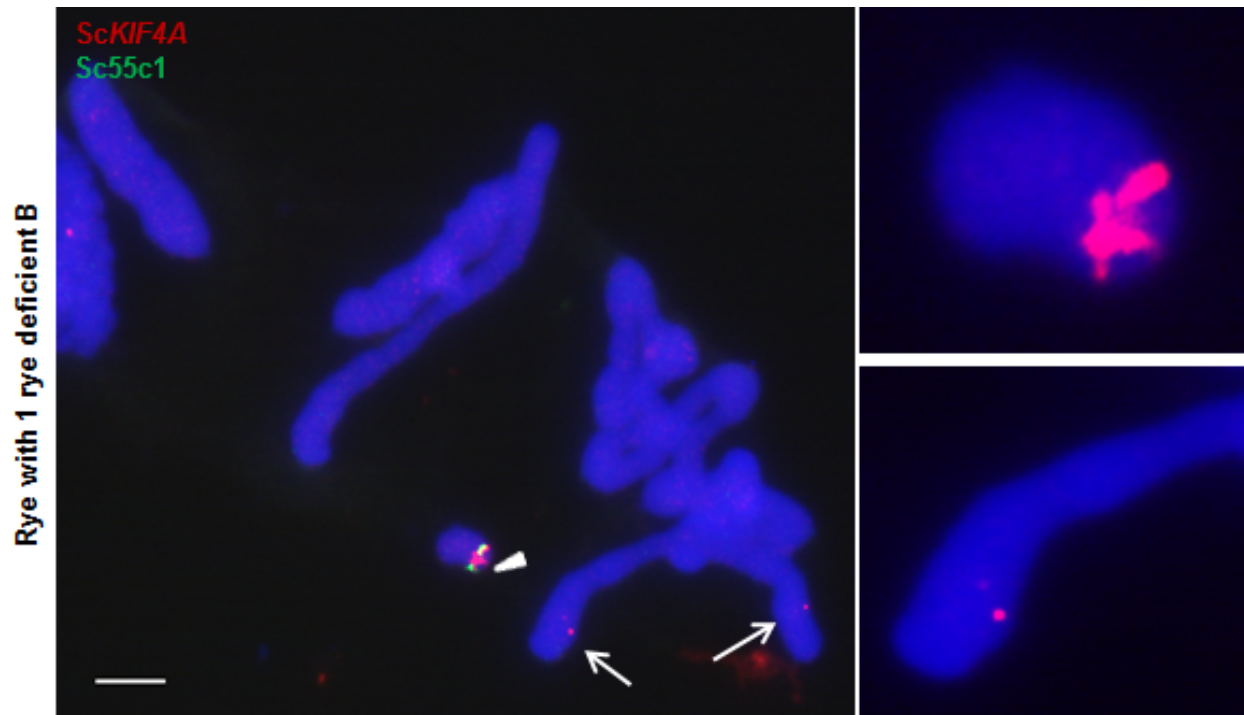


Fig. S6 CAPS analysis for A- and B-derived *ScKIF4A* transcripts. (a) Schemata represent the sequence comparison between two rye B- derived *ScKIF4A* fragments and their A- derived counterpart. RT-PCR (primers K6F and K6R) was performed using anther cDNAs from rye with and without Bs. A major contribution of the B chromosome is found in 2 and 3B plants. (b) Schemata represent the CAPS site position for B-located *ScKIF4A* transcripts. Two SNP sites between rye B-located *ScKIF4A*-like fragments and its A-located counterpart were found at A-located genomic position 5358 (T to G) and 5362 (A to C) (Fig. S3). RT-PCR (primers K7F and K6R) using rye and wheat cDNA from anthers of 0B and +B carriers was performed following by CAPS analysis with *Bam*HI enzyme. Extra bands are only found in rye and wheat possessing Bs. Three biological replicates were used for each experiment. *GAPDH*-specific primers were used to quantify the amount of cDNA.

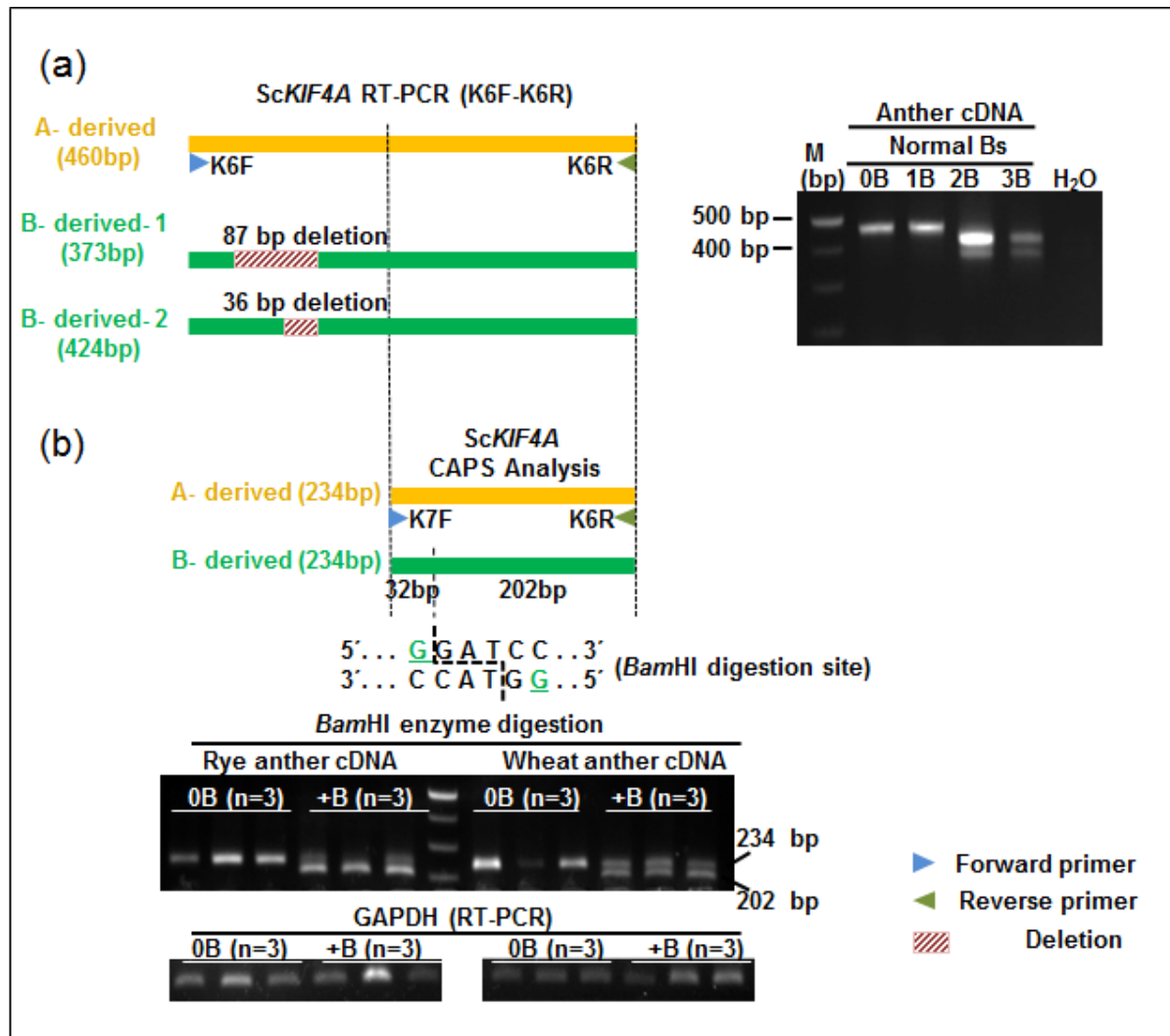


Fig. S7 CAPS analysis for A- and B-derived *ScSHOC1* transcripts. Sequencing (a) and schemata (b) represent one SNP site between rye B-located *ScSHOC1*-like fragment and its A-located counterpart. This SNP site locates in the exonic region at A-located genomic position 2,707 (G to T). RT-PCR using rye and wheat cDNA from anthers of 0B and +B carriers was performed and followed by CAPS analysis with *NsiI* enzyme. Extra bands are only found in rye and wheat possessing Bs (c). Three biological replicates were used for each experiment. *GAPDH*-specific primers were used to quantify the amount of cDNA.

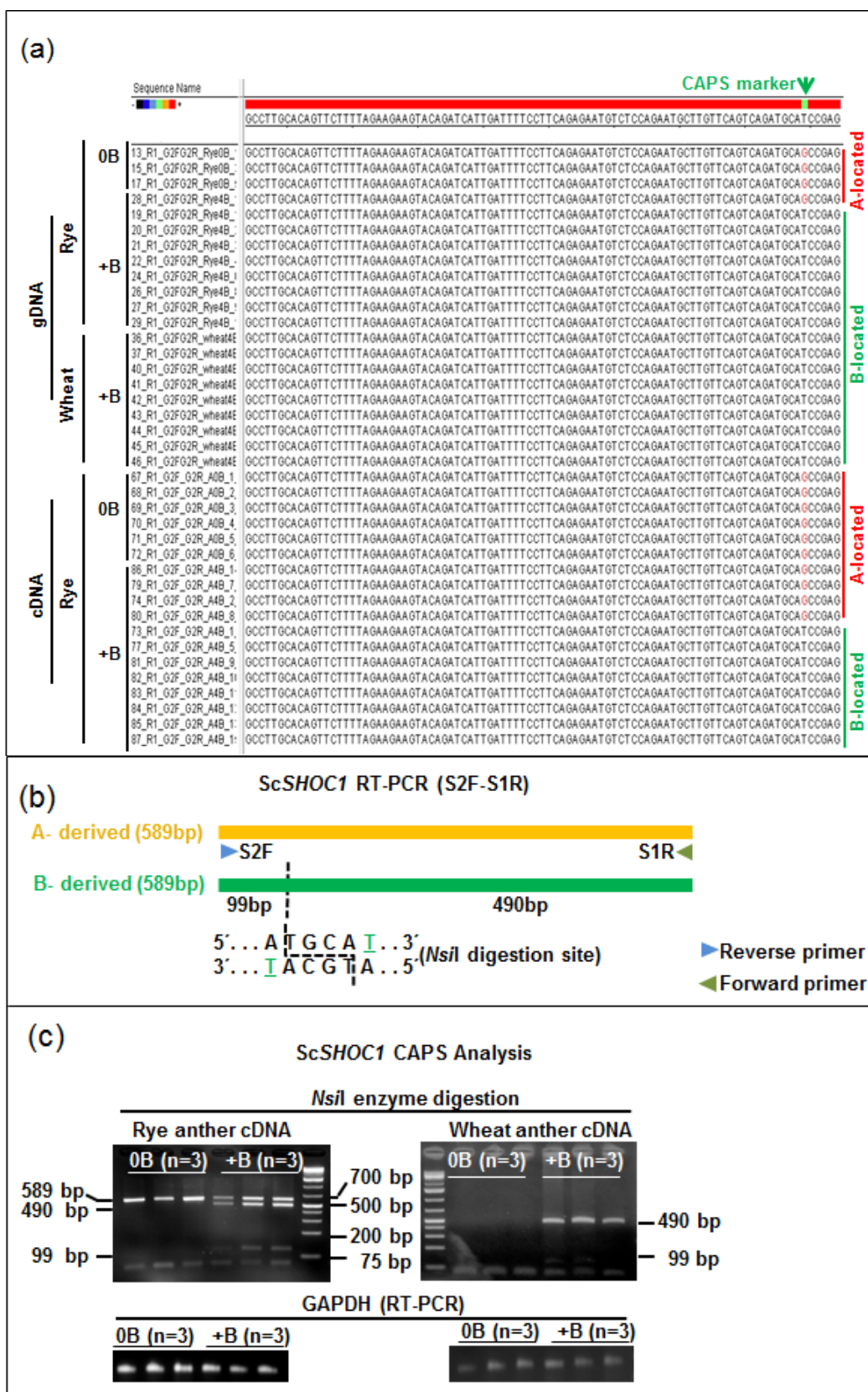


Fig. S8 Gene structure model of A- and B-derived *ScAGO4B* transcripts. (a) Schemata represent the sequence comparison between A- and B-derived *ScAGO4B* transcripts. Percentage of similarity, position of sequence polymorphisms, primers and start/stop codons are indicated. (b) Alignments of informative sequenced *ScAGO4B* clones derived from anther cDNA of rye and wheat with Bs (+B) as well as without Bs (0B). A- and B-derived *ScAGO4B* sequences could be distinguished by polymorphic sites.

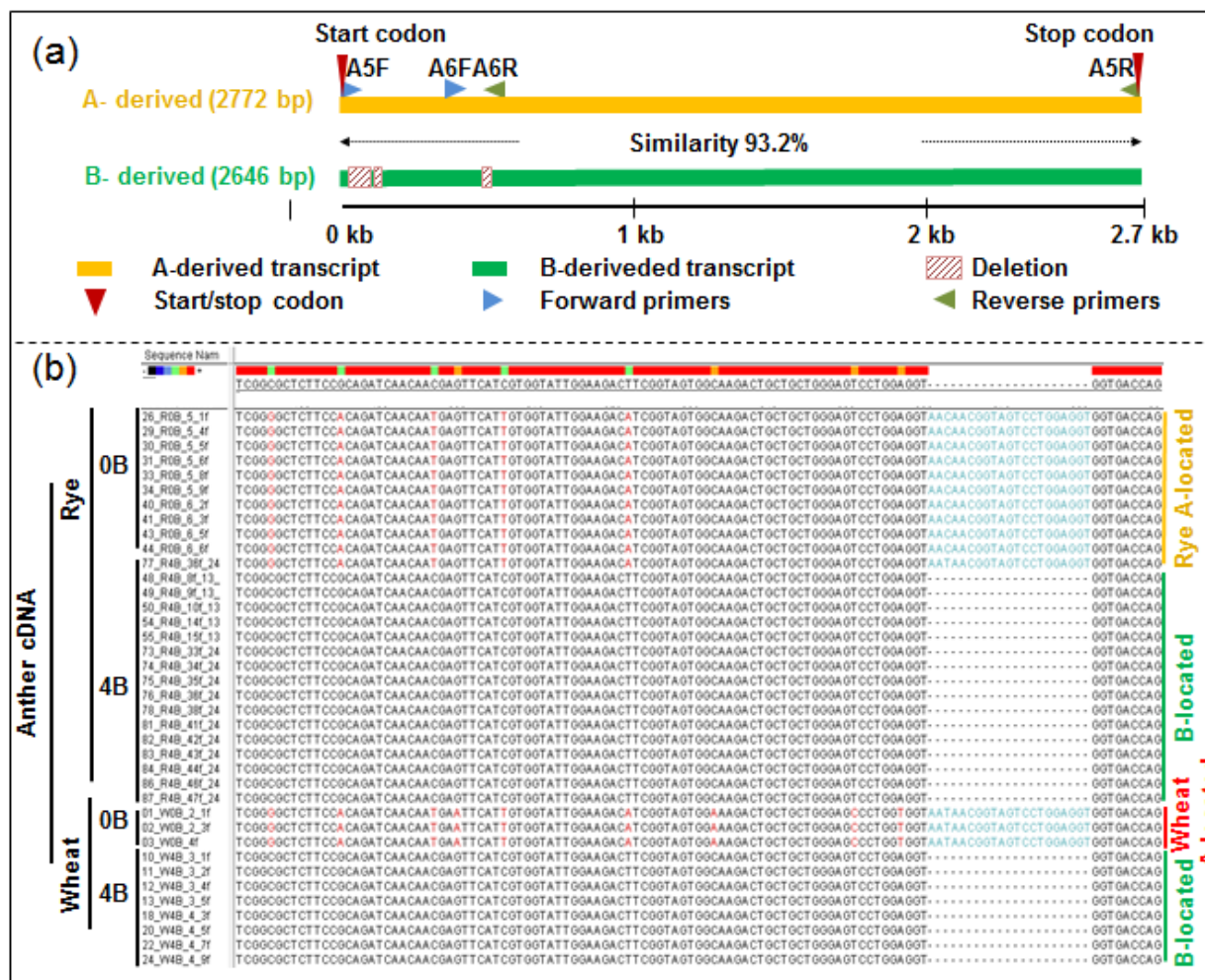


Fig. S9 Domain structure of ScAGO4B protein. Argonaute proteins are classified by the PAZ and PIWI domains. The PAZ domain contains a nucleic acid binding interface. A 5'RNA binding site exists and an active site (DDH) reside in the PIWI domain. The polymorphic sites of B-located ScAGO4B are indicated.

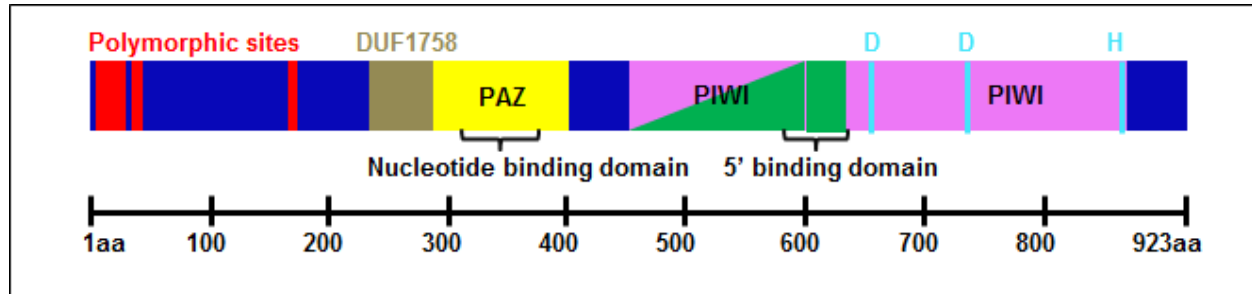


Table S1 List of primer sequences for PCR, RT-PCR, FISH probe preparations, in vitro transcription and quantitative RT-PCR.

(a) List of primer sequences for PCR, RT-PCR and FISH probe preparations for *ScKIF4A*, *ScSHOC1* and *ScAGO4B* genes

Primer Name	Primer seq (5'-3')	T _m (°C)	FISH probe length (bp)
K1F	TTGAGCAATGTGTGTATCCA	53.2	
K1R	AGTTGACTTCTCAGATGCAG	55.3	
K1F	TTGAGCAATGTGTGTATCCA	53.2	3280 bp
K2R	TCATCCCACGATTTTCCATT	53.2	2494 bp
K2F	TGTCTTGTGAACAGTTAGCA	53.2	
K3R	TCGAGCAAATCTTTCAGTCT	53.2	
K3F	GATGCTCACCAACAGTTGCT	57.3	2442 bp
K4R	CGCCGCCGTGTATTATTGAT	57.3	1603 bp
K4F	AAGAAAAGGTGGTTGCGCTT	55.3	
K5R	CGGATGCCACAGAAAACACA	57.3	
K5F	GCTGAAGTAACACGGCAGAA	57.3	
K6R	TCTTGAGCTGAGAGACCTGC	59.4	
K6F	AGGAGGCGGAGATGAAACAA	57.3	
S1F	CCATATCCTCCGCGCTATCC	61.4	
S1R	ATCGTCAACCAGCACCAACT	57.3	
S2F	AGGAGGATCTTTTGTCCGCA	57.3	
S2R	CTCCCAGCACCCAAGTTGTA	59.4	
S3F	TACAACTTGGGTGCTGGGAG	59.4	
S3R	TTACAGCAGAAAGGGAGCGA	57.3	
S4F	CTGAACAGCGGCACATAGAG	59.4	
S4R	CTTCCCGTGTGCTGCAAATA	57.3	
S5F	GACTACTTCTCTCCGGCGTC	61.4	5022 bp
S4R	CTTCCCGTGTGCTGCAAATA	57.3	
S6F	ATGCGGACTCGCTTCCTC	59.89	
S5R	CGTACCTTCTTCCCGGTATC	58.92	
A1F	GGTGCCCATCATAGCAGAAG	59.4	2685 bp
A1R	AGAGCCTTTGTGTATCTTTGCAG	58.9	2786 bp
A2F	AGTCCTGGAGGTAACAACGGT	59.8	
A2R	GGGACTCTCTTCGATTACAGTATGAG	59.7	
A3F	TCGAGAAGTCAAGGCAGAAG	57.3	2587 bp
A3R	TAAACAGGGAAGCCATCATC	55.3	

A4F	TGGA CTCAACACACTGCTTC	57.3	2704 bp
A4R	GAACATGGAGCTCCTCACTTTCTC	62.7	
A5F	GCTCTAGAGCCATGGACCCGCATGATGGGGAG	70.2	
A5R	TCCCCCGGGGAATCAGCAGAAGAACATAGAGCTC	68.3	
A6F	GCCTCTTTACTATCGGGGCT	54.0	
A6R	GTCTCCTCACCTCTTCTGG	56.0	

(b) List of primer sequences for *GAPDH*, *M13* and *Bilby*

Primer Name	Primer seq (5'-3')	Tm (°C)	Note
GAPDH-F	CAATGATAGCTGCACCACCAACTG		Banaei-Moghaddam <i>et al.</i> (2013)
GAPDH-R	CTAGCTGCCCTTCCACCTCTCCA		
M13-F	TTGTAAAACGACGGCCAGTG	57.3	
M13-R	GGAAACAGCTATGACCATG	54.5	
Bilby-F	TTTGCGACAATGACTCAAGC		Francki (2001)
Bilby-R	TGTAGCTCATCGTGGAGTCG		

(c) List of primer sequences for *in vitro* transcription

Primer Name	Primer seq (5'-3')	Note
GFP-F	CGtaatacgactcactatagAGAATCGAGTTAAAAGGTATT G	lowercase letters: the promoter for T7 RNA polymerase
GFP-R	ATTTgcggccgcAGAATCGAGTTAAAAGGTATTG	lowercase letters: <i>NotI</i> site

(d) List of primer sequences for quantitative RT-PCR

Primer Name	Primer seq (5'-3')
ScKIF4A-F	TCTGCCGTGTTACTTCAGCC
ScKIF4A-R	ACGCAATGGAAAATCGTGGG
ScSHOC1-F	CGAGGGATGGAGCCTCTAAG
ScSHOC1-R	AGTTCCTCCTCTGGCTTTCC
ScAGO4B-F	GTGACCAGAAGAGGGTGAGG
ScAGO4B-R	CGGAGTGCTGCCTGAGTATG

Table S2 List of sequence identifiers and description of KIFs, SHOC1 and AGOs sequences used for phylogenetic tree construction (see separate Excel spreadsheet).

Table S3 Accession numbers of ScKIF4A, ScSHOC1 and ScAGO4B genes.

Sequence content	Gene	Genbank No.	Notes	Fig.
A – located gDNA	<i>ScKIF4A</i>	KT750012	Complete	S2a
	<i>ScSHOC1</i>	KT750013	Complete	S2b
	<i>ScAGO4B</i>	KT750014	Complete	S2c
B – located gDNA	<i>ScKIF4A</i>	KT946783	Partial (variant 1)	S4
	<i>ScKIF4A</i>	KT946784	Partial (variant 2)	S4
A – derived mRNA	<i>ScKIF4A</i>	KT956052	Complete	S2a
	<i>ScSHOC1</i>	KT750017	Complete	S2b
	<i>ScAGO4B</i>	KT750015	Complete	S2c, S8
B – derived mRNA	<i>ScKIF4A</i>	KT750020	Partial (variant 1)	S6a
	<i>ScKIF4A</i>	KT750021	Partial (variant 2)	S6a
	<i>ScSHOC1</i>	KT750018	Complete	
	<i>ScAGO4B</i>	KT750016	Complete	S8
B-located contig cI	<i>ScKIF4A</i>	KT713602	Partial	S2a
B-located contig cII		KT713603	Partial	
B-located contig cIII		KT713604	Partial	
B-located contig cIV		KT713605	Partial	
B-located contig cV		KT713606	Partial	
B-located contig cI	<i>ScSHOC1</i>	KT713607	Partial	S2b
B-located contig cII		KT713608	Partial	
B-located contig cIII		KT713609	Partial	
B-located contig cIV		KT713610	Partial	
B-located contig cV		KT713611	Partial	
B-located contig cVI		KT713612	Partial	
B-located contig cVII		KT713613	Partial	
B-located contig cVIII		KT713614	Partial	
B-located contig cI	<i>ScAGO4B</i>	KT713615	Partial	S2c
B-located contig cII		KT713616	Partial	
B-located contig cIII		KT713617	Partial	
B-located contig cIV		KT713618	Partial	
B-located contig cV		KT713619	Partial	

Table S4 Categories of B-located transcripts.

	Vegetative tissues	Anther
Unknown	54.4% (662)	41.2% (805)
Transposons	25.4% (309)	43.0% (841)
Canonical genes	20.3% (247)	15.8% (308)
Sum	1,218	1,954

Tables S5 and S6 The functional annotation and classification of B chromosome-located gene candidates in vegetative (root and leaf) tissue (Table S5) and (anther) tissue (Table S6) (see separate Excel spreadsheet).

Methods S1

Plant cultivation

After germination, root tips were collected to investigate the type and number of Bs in each plant individual and the seedlings were transferred to soil and cultivated for 4 wk under short-day conditions (8 h light/16 h dark, 20°C/18°C). Then, vernalization (10 h light/14 h dark, 4°C) followed for at least 4 wk. Finally, the plants were grown under long-day conditions (16 h light, 22°C day/16°C night), and all experimental materials from different tissues were collected during this period.

Genomic DNA and RNA extraction, PCR and RT-PCR

25 µl PCR or RT-PCR reaction mixtures contained: 100 ng genomic DNA or 1 µl cDNA, 10 µM of each forward and reverse primers (Table S1), 5 mM of each deoxynucleotide triphosphates, 2.5 µl 10xPCR reaction buffer and 1 unit of *Taq* polymerase (Qiagen, Hilden, Germany). The cycling protocol was: 94°C for 3 min, 35 cycles at (94°C for 40s, annealing temperature (Table S1) for 40 s, 1 min kb⁻¹ elongation at 72°C), 72°C final elongation for 10 min. 25 cycles PCR were run with *GAPDH*-specific primers to quantify the abundance of transcripts.

qRT-PCR was performed using the SYBR Green Master (Applied Biosystems) on the 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). 10 µl of PCR mixture contained 1 µl of cDNA template, 5 µl of 2× Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA), and 0.33 mM of the forward and reverse primers (Table S1) for each gene. The amplification conditions were one cycle of 10 min at 95°C, 40 cycles of two consecutive steps of 15 s at 95°C, and 60 s at 60°C as a standard dissociation protocol. *GAPDH*-specific primers (Table S1) were used as endogenous control.

Target cleavage *in vitro* assay

In vitro transcription

To generate plasmids for *in vitro* transcription of A- and B-derived *ScAGO4B*-like transcripts, the complete open reading frames were amplified by A5F and A5R primers (Table S1) using rye

anther 0B and 4B cDNAs, then inserted via *Xba*I/*Sma*I restriction sites into a modified pSP64-poly(A) vector (Promega, Fitchburg WI, USA) that contained an additional *Swa*I restriction site downstream of the poly(A) sequence. Transcription and further treatment of the transcript were performed using standard procedures. Firefly luciferase RNA was generated by SP6 RNA polymerase (Thermo Scientific, Waltham, MS, USA) from the *Xho*I-linearized plasmid pSP-luc(+) (Promega, Fitchburg WI, USA). The mRNAs encoding the AGO4-like proteins were synthesized in the presence of the monomethylated cap analog m⁷GP₃G (Jena Biosciences, Jena, Germany) from *Swa*I-linearized plasmids using SP6 RNA polymerase.

To generate a target for the GFP-specific siRNA, a 432 bp sequence was amplified by PCR from plasmid pGFP-C1 using primers GFP-F and GFP-R (Table S1). Transcription of the GFP target RNA was performed from this PCR fragment by T7 polymerase in the presence of 0.5 $\mu\text{Ci } \mu\text{l}^{-1}$ [α -³²P]CTP using standard conditions.

siRNAs

The sequences of the 24 nt variant of GFP-specific siRNA were 5'-aaguucauccaugccauguguaau-3' (guide strand) and 5'-uacacauggcauggaagaacuuua-3' (passenger strand). To produce siRNA duplexes, the single-stranded RNA oligonucleotides (Biomers, Ulm) were heated for 1 min at 90°C in annealing buffer (30 mM HEPES-KOH, pH 7.4, 100 mM potassium acetate, 2 mM magnesium acetate) and hybridized for 60 min at 37°C.

Cell culture and preparation of cytoplasmic BY-2 cell extract

Nicotiana tabacum BY-2 cells were cultured as described (Gursinsky *et al.*, 2009) at 23°C in Murashige-Skoog liquid medium. Evacuolated BY-2 protoplasts for the preparation of cytoplasmic extract (BYL) were obtained by percoll gradient centrifugation (Komoda *et al.*, 2004; Gursinsky *et al.*, 2009).

Target cleavage assay

In vitro translation of *AGO4B*-like mRNAs was performed in 50% (v/v) BYL at previously described conditions (Schuck *et al.*, 2013). 1.5 µg of the mRNA were translated in a 20 µl reaction for 60 min at 25°C in the presence of 50 nM siRNA. Then, the same amount of siRNA was added again and the reaction continued for 90 min. Afterwards, 2 µg of firefly luciferase (competitor) mRNA and the ³²P-labeled GFP target RNA (50 fmol) were added and the cleavage reaction was carried out for 15 min at 25°C. Total RNA was isolated from the reaction by treatment with 20 µg Proteinase K in the presence of 0.5% SDS for 30 min at 37°C, followed by extraction with chloroform and ethanol precipitation. ³²P-labeled products were separated on 5% Tris-borate polyacrylamide gels containing 8 M urea and visualized by phosphor-imaging.

References

- Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R, Houben A. 2013.** Formation and expression of pseudogenes on the B chromosome of rye. *The Plant Cell* **25**: 2536–2544.
- Francki MG. 2001.** Identification of *Bilby*, a diverged centromeric Ty1-*cop* retrotransposon family from cereal rye (*Secale cereale* L.). *Genome* **44**: 266–274.
- Gursinsky T, Schulz B, Behrens SE. 2009.** Replication of tomato bushy stunt virus RNA in a plant in vitro system. *Virology* **390**: 250–260.
- Komoda K, Naito S, Ishikawa M. 2004.** Replication of plant RNA virus genomes in a cell-free extract of evacuated plant protoplasts. *Proceedings of the National Academy of Sciences, USA* **101**: 1863–1867.
- Schuck J, Gursinsky T, Pantaleo V, Burguán J, Behrens SE. 2013.** AGO/RISC-mediated antiviral RNA silencing in a plant in vitro system. *Nucleic Acids Research* **41**: 5090–5103.