New Phytologist

Research

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

Wei Ma[1], Tobias Sebastian Gabriel[1], Mihaela Maria Martis[2,3], Torsten Gursinsky[4], Veit Schubert[1], Jan Vrána[5], Jaroslav Doležel[5], Heidrun Grundlach[2], Lothar Altschmied[1], Uwe Scholz[1], Axel Himmelbach[1], Sven-Erik Behrens[4], Ali Mohammad Banaei-Moghaddam[6] and Andreas Houben[1]

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstrasse 3, 06466 Stadt Seeland, Germany; [2]Institute of Bioinformatics and Systems Biology/Munich Information Center for Protein Sequences, Helmholtz Center Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany; [3]National Bioinformatics Infrastructure Sweden, Department of Clinical and Experimental Medicine, Linköping University, SE-558185 Linköping, Sweden; [4]Institute of Biochemistry and Biotechnology, Section Microbial Biotechnology, Faculty of Life Sciences, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Str. 3, 06120 Halle/Saale, Germany; [5]Institute of Experimental Botany, Center of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, 78371 Olomouc, Czech Republic; [6]Department of Biochemistry, Institute of Biochemistry and Biophysics (IBB), University of Tehran, PO Box 13145-1384 Tehran, Iran

Authors for correspondence:
Ali Mohammad Banaei-Moghaddam
Tel: +9821 61113385
Email: am_banaei@ut.ac.ir

Andreas Houben
Tel: +49 0 39482 5486
Email: houben@ipk-gatersleben.de

Received: *4 March 2016*
Accepted: *18 June 2016*

New Phytologist (2016)
doi: 10.1111/nph.14110

Key words: Argonaute, B chromosomes, B-located genes, gene erosion. gene expression, pseudogenization, *Secale cereale*.

Summary

- B chromosomes (Bs) are supernumerary, dispensable parts of the nuclear genome, which appear in many different species of eukaryote. So far, Bs have been considered to be genetically inert elements without any functional genes.
- Our comparative transcriptome analysis and the detection of active RNA polymerase II (RNAPII) in the proximity of B chromatin demonstrate that the Bs of rye (*Secale cereale*) contribute to the transcriptome. In total, 1954 and 1218 B-derived transcripts with an open reading frame were expressed in generative and vegetative tissues, respectively. In addition to B-derived transposable element transcripts, a high percentage of short transcripts without detectable similarity to known proteins and gene fragments from A chromosomes (As) were found, suggesting an ongoing gene erosion process.
- *In vitro* analysis of the A- and B-encoded AGO4B protein variants demonstrated that both possess RNA slicer activity. These data demonstrate unambiguously the presence of a functional *AGO4B* gene on Bs and that these Bs carry both functional protein coding genes and pseudogene copies.
- Thus, B-encoded genes may provide an additional level of gene control and complexity in combination with their related A-located genes. Hence, physiological effects, associated with the presence of Bs, may partly be explained by the activity of B-located (pseudo)genes.

Introduction

In addition to the standard set of A chromosomes (As), in many eukaryotes, so-called supernumerary B chromosomes (Bs) have been found causing a numerical chromosome variation. Bs occur in a wide range of taxa from fungi to plants and animals. However, the maximum number of Bs tolerated by individuals varies among different species. Despite the diversity of Bs, they share some common features. They are dispensable and not essential for growth and development of the organisms; they do not pair or recombine with standard As at meiosis and do not follow Mendelian inheritance (Jones & Rees, 1982). Generally, it is assumed that Bs are derived from standard and/or sex chromosomes, either from the same or related species. The available experimental data support both scenarios, suggesting that the way in which Bs form differs between different species and B types (Jones & Houben, 2003). Although unbalanced numbers of As, as in the case of aneuploidy, often cause severely distorted phenotypes (Siegel & Amon, 2012), in many species, the presence of Bs is associated with only mild or non-obvious phenotypes if their number is low. This suggests that Bs are depleted of functional genes. Conversely, excessive numbers of Bs can cause phenotypic effects and may reduce the fertility and fitness of the host. The contributions of single Bs to these phenotypes are usually cumulative, with the severity of effects increasing with the number of Bs (reviewed in Jones & Rees, 1982; Jones, 1995; Bougourd & Jones, 1997; Carlson, 2009). In rye, a large number of Bs depress both viability and fertility (reviewed in Jones & Puertas, 1993). For example, the seed setting is decreased (Müntzing, 1963), heading time is delayed, plant height is reduced (Kishikawa, 1965) and germination is delayed in the presence of Bs (Moss, 1966). In cichlids, Bs may play a role in sex determination (Yoshida *et al.*, 2011)

and, in the fungus *Nectria haematococca*, they may be responsible for antibiotic resistance and pathogenicity (Coleman *et al.*, 2009).

Regardless of the conceivable intra- or interspecies origin of Bs, A- or sex chromosome-derived sequences (including genic sequences) are potential components of any B, for example, the inactive ribosomal RNA (rRNA) genes of the daisy *Brachycome dichromosomatica* (Donald *et al.*, 1997), the Chinese raccoon dog *Nyctereutes procyonoides* (Szczerbal & Switonski, 2003) and the African cichlid fish *Haplochromis obliquidens* (Poletto *et al.*, 2010). Active B-derived rRNA has been demonstrated in *Crepis capillaris* (Leach *et al.*, 2005) and the grasshopper *Eyprepocnemis plorans* (Ruiz-Estévez *et al.*, 2012, 2014).

More recently, sequencing of flow-sorted (Martis *et al.*, 2012) or microdissected (Silva *et al.*, 2014; Valente *et al.*, 2014) Bs has allowed the identification of thousands of B-located genic sequences. In the cichlid *Astatotilapia latifasciata*, the majority of identified B genes are fragmented. However, some genes on Bs are largely intact and the expression of three has been confirmed *in silico* (Valente *et al.*, 2014). Because Bs are dispensable, it is expected that they are prone to the accumulation of mutations and, consequently, they undergo pseudogenization (Green, 1990; Makunin *et al.*, 2014; Banaei-Moghaddam *et al.*, 2015). In rye, at least 11 subregions of 15 B-located pseudogene-like fragments are transcriptionally active. Some have the ability to modulate the expression of their counterparts on As in a tissue- and genotype-dependent manner (Banaei-Moghaddam *et al.*, 2013). A comparable *in trans* regulation of A-located genes by B-encoded genic sequences has been demonstrated for maize (Huang *et al.*, 2016). However, a detailed comparative analysis of A- and B-located genes, with regard to their completeness and functionality, is still missing.

Here, we describe the first functional protein coding gene originating from B. The identification of active RNA polymerase II (RNAPII) in the proximity of B interphase chromatin, and comparative transcriptome analysis, revealed that rye Bs contribute to the transcriptome. We compared the rye B-derived Sc*KIF4A*, Sc*SHOC1* and Sc*AGO4B* genes with their ancestral A-located counterparts and confirmed an A-chromosomal origin and subsequent pseudogenization of B-located genes. The *in vitro* analysis of A- and B-encoded ScAGO4B proteins showed similar RNA-mediated slicing activity for both variants. Thus, rye Bs carry functional protein coding genes.

Materials and Methods

Plant material and cultivation

Seeds of a rye (*Secale cereale* L.) self-fertile inbred line 7415 with and without Bs (Jimenez *et al.*, 1994), hexaploid wheat (*Triticum aestivum* L.) 'Lindström' with and without added standard rye Bs (Lindström, 1965) and rye JBK lines with truncated Bs (lacking the terminal part of the long arm) (Ribeiro *et al.*, 2004) were germinated on wet filter papers. Plant cultivation conditions are described in Supporting Information Methods S1.

Probe preparation, indirect immunostaining and fluorescence *in situ* hybridization (FISH)

Cloned Sc*KIF4A*, Sc*AGO4B*-specific fragments derived from rye 0B DNA and the cloned Sc*SHOC1*-specific fragment derived from rye 0B anther cDNA were used as gene-specific probes (Table S1). D1100 (Sandery *et al.*, 1990), Sc55c1 (Martis *et al.*, 2012) and Revolver (Carchilan *et al.*, 2009) were used as B-specific probes. All probes were labeled by nick translation with ChromaTide Texas Red-12-dUTP or Alexa Fluor 488-5-dUTP (Molecular Probes; http://www.invitrogen.com). The following primary and secondary antibodies were used: anti-RNA polymerase II CTD phospho (Ser2) (RNAPIISer2ph) with goat anti-rat Alexa488 (1 : 200; Jackson ImmunoResearch, West Grove, PA, USA). The non-phosphorylated RNAPII was detected with mouse monoclonal antibody (1 : 300, ab817; Abcam Cambridge, UK) and goat anti-mouse Cy5 (1 : 300; Jackson ImmunoResearch).

Fluorescence *in situ* hybridization and indirect immunostaining were performed as described previously (Houben *et al.*, 2007; Ma *et al.*, 2010). To achieve a lateral optical resolution of *c.* 120 nm (super-resolution, obtained with a 488-nm laser), we applied structured illumination microscopy (SIM) using a C-Apo 63×/1.2 W Korr objective of an Elyra (Jena, Germany) PS.1 microscope system and the software ZEN (Carl Zeiss GmbH). Images were captured separately for each fluorochrome using the 561-, 488- and 405-nm laser lines for excitation and appropriate emission filters.

Rye B sequence identification

The DNA of sorted rye Bs, first described in Martis *et al.* (2012), was further amplified by the Illustra GenomiPhi V2 DNA Amplification Kit (http://www.gelifesciences.com), and 38 μg DNA in total were obtained. A standard Illumina (San Diego, CA, USA) TruSeq DNA library was prepared and sequenced using the Illumina Hiseq2000 device (24-fold coverage). Furthermore, three Illumina Nextera Mate-Pair libraries (span sizes: 9, 7 and 5 kb; two-fold to six-fold coverage) were constructed and sequenced (Illumina MiSeq) as described previously (Beier *et al.*, 2015). In addition, a low-coverage PACBIO reads library (0.3-fold) was used. A first attempt to assemble the rye B genomic data by SOAPDENOVO2 (Luo *et al.*, 2012) led to a highly fragmented assembly with a low L50 of 251 base pairs and very low sequence amounts, which was not used further on. The insufficient assembly metrics are probably caused by the overall high repetitivity of B interspersed with only few and relatively short low copy genic regions. To overcome the assembly difficulties, a template-guided and gene-centric approach was developed. For that purpose, gene-containing contigs of a novel whole genome sequencing (WGS) rye genome assembly (early access provided by E. Bauer, Technical University of Munich, Germany) were used as bait to specifically extract genic reads from the paired end and mate pair sets of the flow-sorted B sequences. The reads were aligned to the rye WGS contigs via BWA (Li & Durbin, 2009), and only reads with unambiguous matches were selected. Using SOAPDENOVO2, new, gene-enriched B contigs were generated from the mapped

reads and further scaffolded with the PacBio reads by Bambus (Pop *et al.*, 2004). Gaps within scaffolds were closed with the help of the paired end library and GapFiller (Nadalin *et al.*, 2012).

RNA-sequencing (RNA-seq) and *in silico* analysis of B transcripts

Rye Illumina paired-end RNA-seq reads obtained from generative (anther) and vegetative (root and leaf) tissues with (4B) and without (0B) Bs of several plants were assembled to transcripts using the Bowtie/TopHat (Trapnell *et al.*, 2009) and Cufflinks (Roberts *et al.*, 2011) pipelines. As genomic reference, a subset of the WGS rye genome assembly (contigs that contain genes) and the template-guided, gene-focused rye B assembly were used as a combined pool. Read alignments with >1 mismatch and multiple matches to the reference sequences were discarded. This restrictive approach is necessary to unambiguously distinguish between A- and B-derived transcripts. However, at the same time, it also removes all genic sequences which are still identical between the main A genome and Bs, so that the resulting numbers of B-derived transcripts should be seen as a lower boundary.

The assembled transcript abundances (FPKM, fragments per kilobase of transcript per million mapped fragments) in the different tissues, as well as the confidence intervals for each FPKM, were estimated using Cufflinks (Roberts *et al.*, 2011). The assembled Cufflinks transcripts were processed with Cuffdiff to measure the abundance of transcripts and to test for differential transcript expression. The fold change ($\log_2$)-filtered differentially expressed genes between 0B and 4B rye plants for both individual tissues (vegetative and anthers) were visualized using R v.3.0.2 (R Core Team, 2014) (volcano plots) with a 5% false discovery rate (FDR). Furthermore, the obtained assembled Cufflinks transcripts were scanned for coding regions by TransDecoder (http://transdecoder.sourceforge.net), and filtered for a minimum length open reading frame of 180 base pairs (60 amino acids). In the case of splice variants, the longest transcript per gene locus was used as representative.

The genes and gene-like sequences of the four transcript set a-0B, a-4B, v-0B and v-4B thus identified were functionally characterized by the AHRD pipeline (Tomato Genome Consortium, 2012), which combines search results against protein databases with PFAM and Interpro scans into human readable description and gene ontology (GO) annotations. To obtain a better overview of relevant functions, the gene candidates were classified by an elaborated keyword search of their description line, together with associated Pfam and Interpro descriptions, into the following four categories: transposons (e.g. reverse transcriptase, transposase); potential transposons (e.g. zinc finger, aspartic protease); unknown; and canonical genes, with specified functions.

The B-located gene candidates classified as canonical genes were tested for specific GO term associations via the GOstats package (v.2.32.0) (Falcon & Gentleman, 2007) against a universal gene set containing all rye A and rye B genic sequences. To judge whether or not the observed differences between the subset of rye B canonical genes and the reference set are significant, a Fisher's exact test was performed. The identified over-represented GO terms of B gene candidates in both anthers and root/leaf tissues were visualized using the R package.

Genomic DNA and RNA extraction, PCR and reverse transcription-polymerase chain reaction (RT-PCR)

Genomic DNA was extracted from leaves by a DNAeasy plant mini kit (Qiagen). Total RNA was isolated from anthers by the TRIzol method (Life Technologies Inc., Gathersburg, MD, USA). RNA samples were treated with DNA-free DNase (Ambion TURBO DNase; Invitrogen) before cDNA synthesis. The absence of DNA was confirmed by PCR with *Bilby*-specific primers (Table S1). All cDNAs (20 µl) were generated from 1 µg DNase I-treated RNA, using the Reverse Aid H Minus First Strand cDNA Synthesis Kit (Fermentas, Hanover, MD, USA). Details of the PCR, quantitative PCR and RT-PCR experiments are described in Methods S1 and Table S1.

Sequence analysis

PCR products were purified using an Invisorb Spin DNA Extraction Kit (Stratec Molecular, Berlin, Germany) and subsequently cloned with the help of the StrataClone PCR cloning kit (Stratagene, La Jolla, CA, USA). Recombinant colonies were identified by colony PCR using an M13 primer pair (Table S1). DNA fragments were sequenced by the facility of the IPK (Gatersleben, Germany). Sequences were analyzed by Sequencher 5.2.4 (Gene Codes Corp., Ann Arbor, MI, USA.), assembled using Seqman pro 12.0.0 (Dnastar Inc., Madison, WI, USA) and processed by EditSeq and MegAlign Lasergene 8 (Dnastar).

Molecular phylogenetic analyses

Reference IDs for the phylogenetic analysis of kinesins (KIFs), shortage in chiasmata 1 (SHOC1) and Argonaute (AGO) sequences used in this study are given in Table S2. After aligning the sequences with ClustalW, phylogenetic trees were calculated using the boot-strapped neighbor-joining algorithm in Mega 6.06 (http://www.megasoftware.net/) with 1000 trials.

Genotyping of Sc*KIF4A* and Sc*SHOC1* fragments by cleaved amplified polymorphic sequences (CAPS)

Primers K7F and K6R (Table S1) were used to amplify A- and B-derived Sc*KIF4A* fragments. Primers S2F and S1R (Table S1) were used to amplify A- and B-derived Sc*SHOC1* fragments. Then, the PCR products were checked for non-specific amplification on a 1% agarose gel. The *Bam*HI enzyme for Sc*KIF4A* and the *Nsi*I enzyme for Sc*SHOC1* were used for genotyping; 20 µl of the restriction enzyme digestion reaction included: 10 µl PCR mixture, 2 µl 10× buffer and 10 units of *Bam*HI or *Nsi*I (Thermo Fischer Scientific, Waltham, MA, USA). The reaction was incubated for 5 h at 37°C. The digestion products were checked after 2% agarose gel electrophoresis.

Target cleavage *in vitro* assay

Details of the target cleavage assay, including *in vitro* transcription, generation of small interfering RNAs (siRNAs), and the cell culture and preparation of cytoplasmic BY-2 cell extract are described in Methods S1.

Accession numbers

RNA-seq and B sequence reads are deposited in EMBL/ENA (accession no. PRJEB12520). The GenBank accession numbers for each sequence of Sc*KIF4A*, Sc*SHOC1* and Sc*AGO4B* are listed in Table S3.

Results

Active RNAPII enzymes are closely associated with rye B chromatin

Most eukaryotic genes are transcribed by RNAPII. Therefore, the co-localization of transcripts and RNAPII is a sensitive prove for active transcription (Komarnitsky *et al.*, 2000; Sims *et al.*, 2004). We combined immunostaining with FISH to test the potential transcriptional activities of rye Bs. Immunolabeling visualized active RNAPII enzymes phosphorylated at serine 2 (RNAPIISer2ph) in interphase nuclei of rye and wheat containing two Bs. To identify the Bs in both species, a subsequent FISH was performed, applying the sub-terminal B-specific repeat D1100 (Sandery *et al.*, 1990) and the rye-specific repeat *Revolver* (Carchilan *et al.*, 2009) as markers, respectively. By super-resolution microscopy (SIM), the close association of active RNAPII and B chromatin was proven in both species, and its distribution did not differ between A and B chromatin. Although heterochromatic regions (strongly stained by 4′,6-diamidino-2-phenylindole (DAPI)) of rye were devoid of active RNAPII, the D1100 repeats and *Revolver*-positive chromatin were intermingled by RNAPIISer2ph-specific immunosignals (Fig. 1a,b). It is obvious that the distribution of active RNAPII enzymes indicating transcriptional activity is similar at A and B euchromatin in interphase nuclei.

In addition, in the same experiment, we applied antibodies against inactive RNAPII which is not phosphorylated. Interestingly, this enzyme modification is similarly distributed within the interphase nuclei around the rye and wheat B chromatins, as found for the active RNAPII (Fig. 1a,b). In rye with 2Bs, even an accumulation of inactive RNAPII at B chromatin was proven (Fig. 1a). Obviously, RNAPII molecules seem to be present permanently within euchromatin, where they can become activated when transcription is required.

Comparative *in silico* identification reveals a high number of B-associated transcripts

To determine to what extent Bs influence transcription levels and to obtain an overview of the amounts and types of B-located gene-like sequences which are still transcribed, we compared RNA-seq data from rye plants with and without Bs, additionally separated into anthers and vegetative tissues (roots and leaves). The four sets of transcript reads (a-0B, a-4B, v-0B, v-4B) were mapped with high stringency to a genomic sequence pool, consisting of rye A WGS contigs and B-specific assemblies. The resulting annotations of A- and B-located gene models were then used for further comparative analyses with respect to transcriptional differences and functional biases.

For the identification of significantly differentially expressed genes, only transcripts that exhibited an FPKM change of two-fold or more and $P < 0.05$ were considered (Fig. 2a). Using this threshold, 916 genes showed significant expression differences between 0B and 4B plants in vegetative and 1569 in generative tissues. The observed expression changes can either be attributed to a direct contribution from B-located genic sequences to the transcript population or to a modulation of A-located genes as a consequence of B presence. The latter seems to be more frequent as, for both tissues, the most prominent effect of Bs was the down-regulation of genes with a proportion of 92%. Less than 10% of the differentially expressed genes are upregulated under the influence of Bs.

For the functional annotation, only assembled transcripts with an open reading frame of ≥ 180 bp were considered. In total, 1954 and 1218 B-located and expressed genic sequences could be clearly identified in the respective generative and vegetative 4B tissues (Tables S4–S6). The largest part of the B gene-like sequences belongs to uncharacterized, often short, gene fragments, without detectable similarity to functionally annotated proteins; in anthers, they add up to 54% and, in vegetative tissues, up to 41% of all transcripts, suggesting an ongoing gene erosion process of B-located genic sequences, as postulated by Banaei-Moghaddam *et al.* (2013). The second largest part of B transcripts stems from active transposable elements. With a proportion of 43%, anthers of B-possessing plants showed an elevated transposon expression compared with the 25% detected in vegetative tissues. This finding is in line with the known increase in transposon transcription in the germ line of plants (Martinez & Slotkin, 2012). Nevertheless, 16% (308) of the B-located anther and 20% (247) of the vegetative tissue transcripts showed similarities to canonical protein coding genes which carry a proper functional annotation. A GO enrichment analysis for these genes revealed 28 over-represented categories (Fig. 2b).

The strongest enrichments were found for photosynthetic and respiratory processes, which confirms the proposed ongoing flow of plastid DNA fragments from chloroplasts and mitochondria into Bs (Martis *et al.*, 2012; Ruban *et al.*, 2014). The over-represented 'negative' regulatory functions, such as 'gene silencing by RNA', 'negative regulation of gene expression' and 'negative regulation of metabolic process', provide good hints of the possible causes of the detrimental effects of too many Bs and fit with the above-described high proportion of reduced transcription under the influence of Bs. Very interesting enriched GO terms with respect to B propagation and drive are 'microtubule-based movement', 'mitosis' and 'cell division'. Overall, it is intriguing that the 'junkyard' Bs produce at least 300 active transcripts from different functional domains. The following main parts of this article describe three key player genes, chosen from

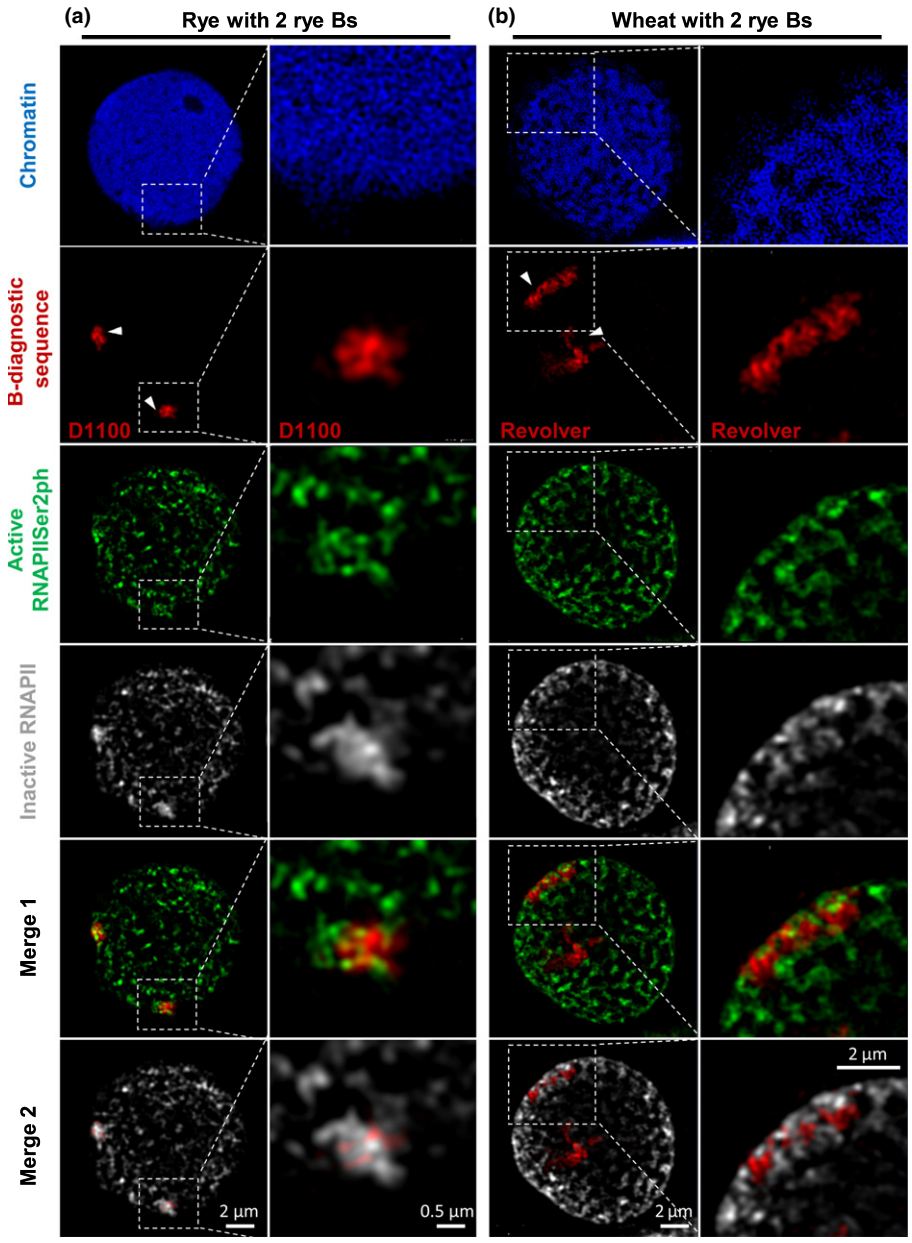

Fig. 1 The distribution of active and inactive RNA polymerase II (RNAPII) in rye and wheat nuclei with two B chromosomes (Bs) was identified by structured illumination microscopy (SIM). Immunostaining of RNAPIISer2ph (active) and fluorescence *in situ* hybridization (FISH) with the B-specific repeat D1100 or Revolver to identify rye B chromatin show the presence of active RNAPII at rye B chromatin (Merge 1). Inactive RNAPII also co-localizes with B chromatin and, in rye, it is even amplified (Merge 2). The right panels show the regions of interest (rectangle) magnified. (a) Interphase nuclei of rye possessing two Bs. (b) Interphase nuclei of a wheat–rye two B addition line.

the active B-located gene candidates, in more detail and with additional experimental proof for function.

The B-located genes Sc*KIF4A*, Sc*SHOC1* and Sc*AGO4B* are transcribed

Based on our RNA-seq dataset, three putative chromatin or epigenetically related rye B-derived genic sequences were selected from the generative transcript data for further analysis. A phylogenetic analysis revealed their high similarity to the kinesin family member 4 gene (*KIF4A*), shortage in chiasmata gene (*SHOC1*) and Argonaute family member 4 gene (AGO4B) (Fig. S1a–c). KIF4 kinesins are key players in several crucial cellular processes, including chromatid motility, chromosome condensation (Zhu & Jiang, 2005) and intracellular transport (Sekine *et al.*, 1994). SHOC1 is required for the promotion or stabilization of single-strand invasion and the formation of class I cross-overs during meiotic recombination processes (Macaisne *et al.*, 2008). AGO4 is involved in the regulation of gene expression (see the Discussion section). As a general property, this protein has a 'slicer activity', that is, it is capable of catalyzing sRNA-directed endonucleolytic degradation of targeted RNA transcripts (Zilberman *et al.*, 2003; Qi *et al.*, 2006).

The corresponding A- and B-located genomic sequences of each of the transcripts were identified by BLAST analysis of a database containing rye genomic 0B and sequences of flow-sorted Bs described in the European Nucleotide Archive (accession nos. ERP001061, PRJEB12520). The gene structure with putative intronic and exonic regions was predicted by FGENESH (http://linux1.softberry.com/berry.phtml) based on the A-located genes

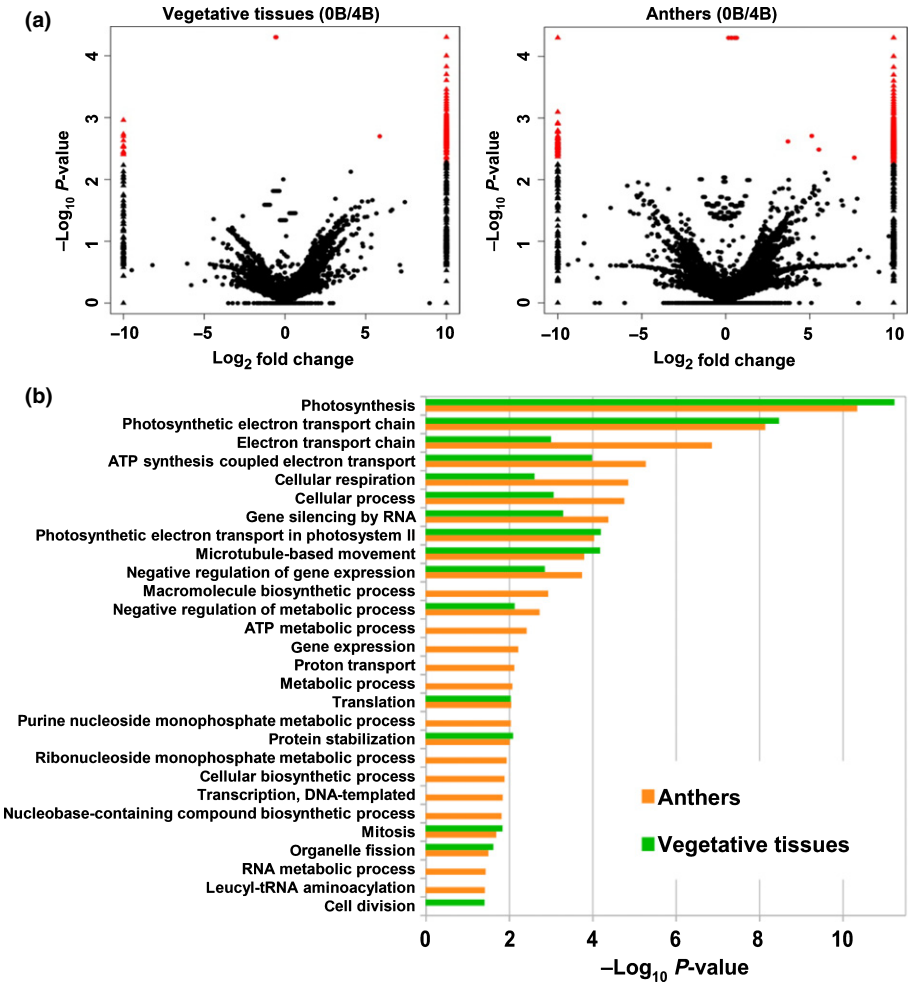

Fig. 2 Differential expression between 0B and 4B rye plants and gene ontology (GO) enrichment of B-located transcripts. (a) The volcano plots reveal differences in gene expression between rye 0B and 4B in the vegetative (root and leaves) and generative (anther) samples. The statistical evidence as negative $\log_{10}$ of the *P* value is plotted on the *y*-axis against the $\log_2$ fold change of expression on the *x*-axis. In each graph, every point represents an individual transcript. Transcripts with low *P* values (< 0.05) and high fold change appear towards the top of the plot far to either the left- or the right-hand side (red dots/triangles) and are highly significantly differentially expressed. Upregulated genes are displayed by dots and/or triangles with positive *x*-values (right-hand side), whereas downregulated genes are shown by negative *x*-values (left-hand side). Black dots represent genes that were not found to differ significantly between 0B and 4B generative samples. In the vegetative tissues, 916 genes were found to display a significant differential expression between 0B and 4B; < 10% were upregulated under the influence of 4Bs. For 1569 genes, the amount was almost doubled in anthers, but still maintained a nearly identical downregulated proportion of 92% (1450 cases). (b) The bar chart highlights significantly enriched GO categories of the still transcribed portion of B-located gene candidates identified in anthers (orange) and root/leaves (green) in comparison with a combined reference set of all rye A and B genes for the respective tissues. To determine whether the observed differences are significant, Fisher's exact test was performed using the GOstats software package with a *P* value cut-off of 0.05.

(Fig. S2). All GenBank accession numbers corresponding to the different nucleotide sequences are available in Table S3. Primers were designed for both A- and B-located sequences in the putative exonic regions based on the gene models (Fig. S2; Table S1). Next, the *in silico*-identified sequences were sequence confirmed after PCR and RT-PCR using genomic DNA, anther cDNA of rye and wheat with (+B) as well as without (0B) Bs as templates. Full-length transcripts of *ScSHOC1* and *ScAGO4B* derived from As or Bs were obtained by RT-PCR using the primer pairs S6F/S5R and A5F/A5R, respectively.

To distinguish between A- and B-derived sequences, the following criteria were used (Fig. S3): the same amplicon size exists, but B-specific single-nucleotide polymorphisms (SNPs) are detectable after sequencing; additional differently sized amplicons are formed in wheat and rye with Bs; and amplicons are only formed in wheat carrying Bs with the same size as the additional differently sized amplicons which are formed in rye with Bs.

Different states of pseudogenization were observed after comparison between A- and B-located genomic contig sequences (Fig. S2). The A- and B-located Sc*KIF4A* sequences revealed an overall 78.5% similarity over the entire length of 10 kb (Fig. S2a). However, a 1.1-kb-long subregion of Sc*KIF4A* showed only 53.6% similarity, and a 224-bp deletion and a 201-bp-long insertion, as well as some SNPs, were observed exclusively in the B-located sequence (Fig. S4). Notably, at least two variants of the B-located Sc*KIF4A* were found, which are characterized by a few polymorphic sites (Figs S3, S4). The *c.* 13-kb-long A-located Sc*SHOC1* sequence exhibited 98.6% (Fig. S2b)

and the *c.* 8-kb-long A-located Sc*AGO4B* sequence exhibited 93% similarity to the corresponding B-located sequences at the genomic level (Fig. S2c). Hence, we conclude that the B-located genic sequences undergo pseudogenization.

To test whether the B-originating transcripts are potentially functional, *in silico* translation (http://web.expasy.org/translate/) was performed based on the B-derived mRNAs, which were identified by searching the RNA-seq reads and B-located genomic contigs or by sequencing of RT-PCR products. The predicted amino acid sequences were analyzed to identify potential stop codons. One of the nucleotide changes (C > T) was found in the B-located Sc*KIF4A* genomic contig cV (position 872), which resulted in a premature stop codon (Fig. S2a). B-encoded Sc*SHOC1* and Sc*AGO4B* genes did not reveal any premature stop codon.

Amplification increases the copy number of the B-located genic sequences

To verify the A and B localization of the candidate genes, FISH was performed on rye mitotic and meiotic metaphase cells containing Bs. After FISH with a Sc*KIF4A*-specific probe, all mitotic Bs displayed a strong hybridization signal at the long chromosome arm near the D1100-positive region, but outside the nondisjunction controlling region. By contrast, two pairs of As showed only weak Sc*KIF4A* hybridization signals (Figs 3a, S5). Similarly, Sc*SHOC1* gave strong hybridization signals on all Bs, but only faint signals were observed on two A pairs at metaphase I of meiosis (Fig. 3b). Sc*AGO4B* also displayed a comparable situation, that is, the meiotic Bs showed very intense Sc*AGO4B* signals at an interstitial position, whereas a single A pair displayed a rather weak hybridization signal (Fig. 3c). The stronger hybridization signals on Bs suggest the amplification of the B-encoded Sc*KIF4A*, Sc*SHOC1* and Sc*AGO4B* sequences.

The number of Bs affects the gene expression pattern

The total transcriptional activity of A- and B-encoded genes was tested in anthers of rye carrying different numbers of Bs. This was performed by quantitative RT-PCR using primers which did not discriminate between A- and B-derived transcripts. To further determine the relative contribution of A- and B-derived mRNAs in rye anthers carrying different numbers of Bs, the A or B origin of the transcript was identified.

An increased expression of total Sc*KIF4A* was found in 2B plants, whereas reduced and more similar expression patterns were revealed in plants with 0B, 1B, 3Bs or 4Bs (Fig. 4a). To determine the relative contribution of A- and B-derived Sc*KIF4A* transcripts, cDNAs of rye anthers with different number of Bs were used to perform RT-PCR with the primers K7F and K6R. A CAPS analysis was performed after PCR of the cloned RT-PCR products using the primer pair K7F/K6R (Fig. S6a,b). Interestingly, in plants carrying more than 2Bs, all Sc*KIF4A* transcripts derived exclusively from Bs (Figs 4b, S6a).

Sc*SHOC1* revealed a similar dynamic expression pattern. An increased expression was found in 2B and 4B plants (Fig. 4c). To determine the relative contribution of B-derived Sc*SHOC1* transcripts, cDNAs of rye anthers with different numbers of Bs were used to perform RT-PCR with the primer pair S2F/S2R. A CAPS analysis of the cloned RT-PCR products was subsequently performed to determine the relative contribution of A- and B-derived Sc*SHOC1* using the primer pair S2F/S1R (Fig. S7). The analysis revealed that the amounts of B-derived Sc*SHOC1* transcripts increased almost linearly from 1B to 3B, but were reduced in plants with 4Bs (Fig. 4d). In 3B plants, 86% of Sc*SHOC1* transcripts were encoded by Bs.

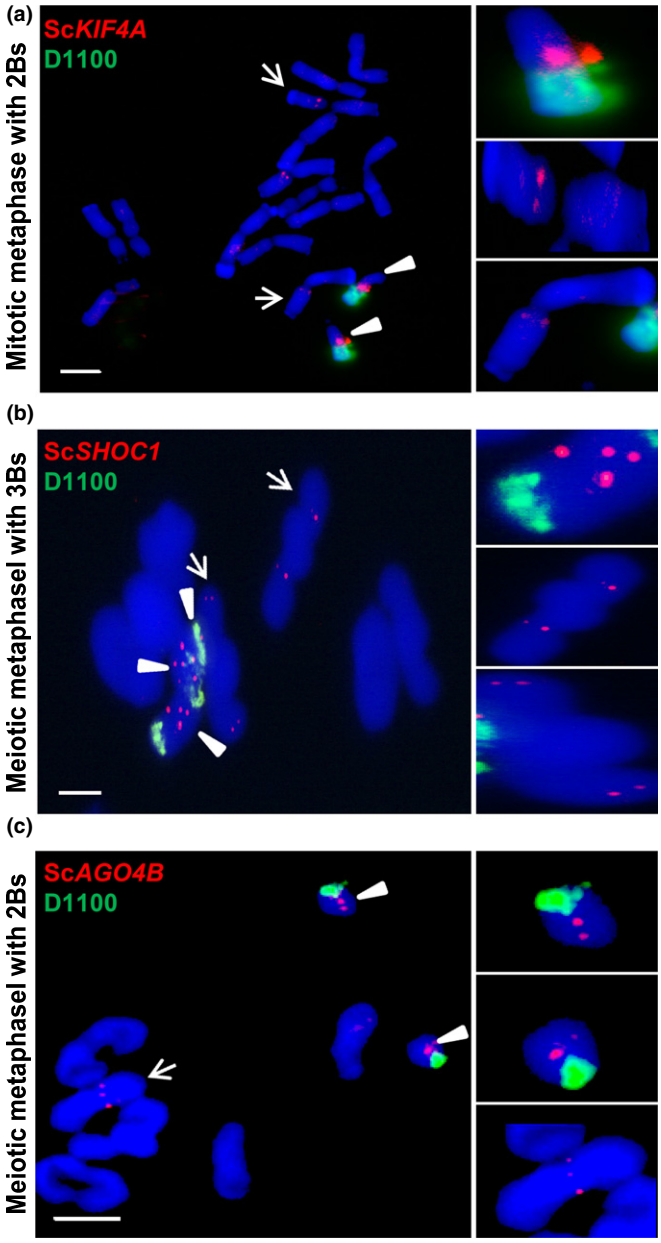

Fig. 3 Chromosomal locations of Sc*KIF4A*, Sc*SHOC1* and Sc*AGO4B* by fluorescence *in situ* hybridization (FISH). Mitotic metaphase or meiotic metaphase I cells of rye with B chromosomes (Bs) after FISH with labeled Sc*KIF4A* (a), Sc*SHOC1* (b) and Sc*AGO4B* (c) (in red). FISH with the B-specific D1100 repeat (in green) allowed the identification of Bs. Chromosomes are stained by 4′,6-diamidino-2-phenylindole (DAPI) (in blue). Arrowheads, signals from Bs; arrows, A-localized FISH signals. The inset shows the signals on further enlarged A and B chromosomes. Bars, 5 μm.

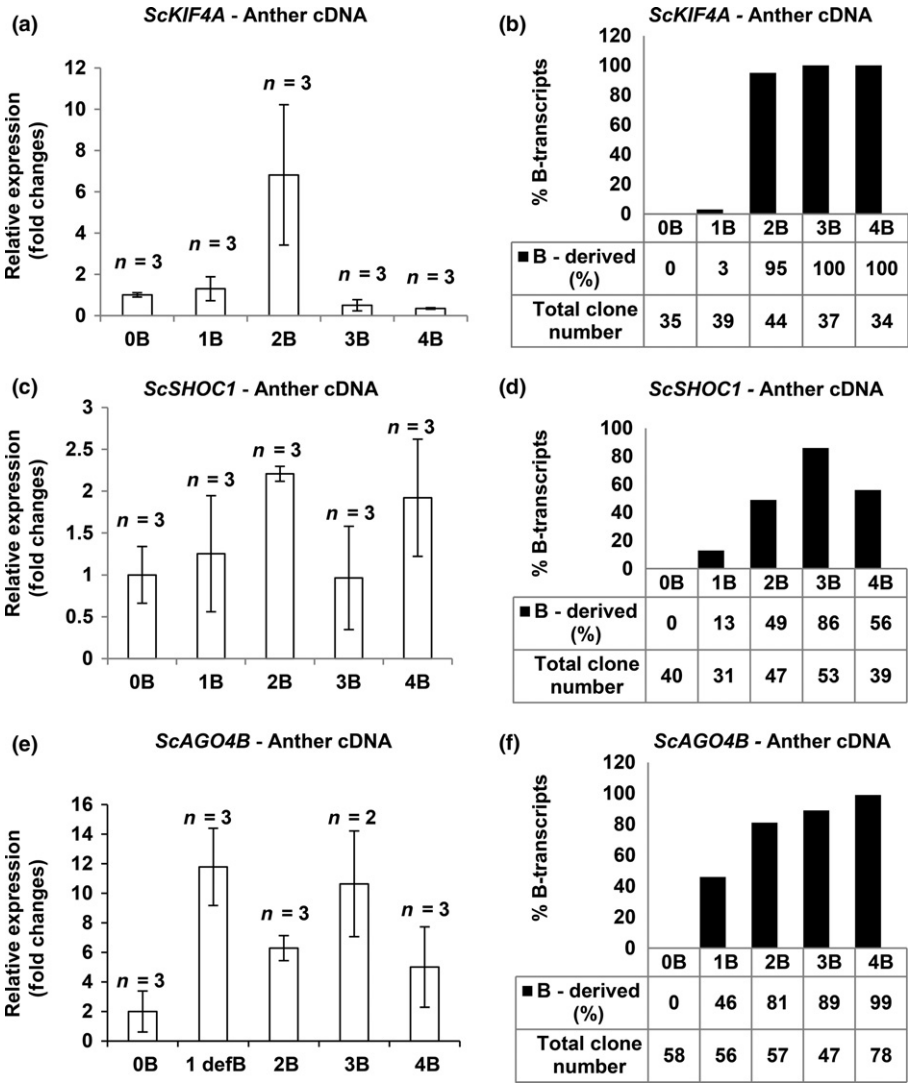

Fig. 4 Quantitative analysis of Sc*KIF4A*, Sc*AGO4B* and Sc*SHOC1* transcripts in the presence and absence of B chromosomes (Bs). The total transcription of Sc*KIF4A* (a), Sc*SHOC1* (c) and Sc*AGO4B* (e) was measured by quantitative reverse transcription-polymerase chain reaction (RT-PCR) in rye anther cDNA containing different numbers of Bs. The number of biological replicates is indicated above the bars. Error bars, ± SD. The contribution of B-derived Sc*KIF4A* (b), Sc*SHOC1* (d) and Sc*AGO4B* (f) transcripts from rye anther cDNA with different numbers of Bs was measured by colony PCR, followed by cleaved amplified polymorphic sequences (CAPS) analysis or nested PCR.

The expression of Sc*AGO4B* varied in the presence of different numbers of Bs in a zig-zag-like pattern (Fig. 4e). The full-length Sc*AGO4B* transcripts were cloned using the primer pair A5F/A5R. To distinguish between A- and B-originating transcripts, a subsequent nested PCR flanking the polymorphic region between A and B transcripts was performed (Figs S2, S8). The PCR results revealed that, with an increasing number of Bs, the amount of B-derived transcripts increased (Fig. 4f). Interestingly, in plants with four additional Bs, 99% of transcripts derived from Bs.

Altogether, we conclude that the total transcriptional activity of Sc*KIF4A*, Sc*SHOC1* and Sc*AGO4B* varies with the presence of different B numbers, whereas the B-derived transcript portion increases gradually and, in general, with increasing numbers of Bs.

A- and B-encoded ScAGO4B-like proteins show similar *in vitro* RNA slicer activities

The protein sequences deduced from the identified A- and B-located Sc*AGO4B* transcripts showed significant homologies to the *Arabidopsis thaliana* (At) *AGO4* gene (Fig. S1c). The A- and B-derived Sc*AGO4B* transcripts revealed a similarity of 93.2% over the entire length. Three deletions (78, 27 and 21 bp in length) and 61 SNPs were observed in the B-derived transcripts (Fig. S8a). This caused a 42-amino-acid deletion (Fig. S9), 38 silent and 23 missense mutations. To test whether these mutations in the B-derived ScAGO4B lead to protein changes, we predicted the conserved domains of the ScAGO4B protein using the National Center for Biotechnology Information (NCBI) protein annotation resource Conserved Domain Database online tool (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). We found that all mutations were located outside of the conserved domains known to be important for the slicer activity of AtAGO4 (Irvine *et al.*, 2006) and were concentrated in the N-terminal, non-conserved region of the protein (Fig. S9).

AGO4 has an endonuclease activity ('slicer' activity), which is directed by a bound, single-stranded sRNA guide strand. The slicer activity accordingly cleaves target RNA molecules that are complementary to the sRNA guide strand (Qi *et al.*, 2006;

Carbonell & Carrington, 2015). To evaluate the functionality of the proteins that are encoded by the A- and B-derived Sc*AGO4B*-like genes, respectively, each was tested for its slicer activity. For this, we applied a cell-free extract (BYL) of tobacco (*Nicotiana tabacum*) BY-2 protoplasts, which enables the reconstitution of functional AGO-containing RNA-induced silencing complexes (AGO/RISC) *in vitro* (Iki *et al.*, 2010; Schuck *et al.*, 2013). That is, the A- and B-derived cDNAs of the Sc*AGO4B* genes were transcribed *in vitro* to generate the corresponding mRNAs. The mRNAs were then translated *in vitro* in BYL to yield the respective AGO4 protein variants and, by the addition of an siRNA to the translation reaction, the respective AGO4 variants were loaded ('programmed') *in statu nascendi* with this RNA. As AtAGO4 is known to associate preferentially with 24-nucleotide-long siRNAs containing a 5′-terminal adenosine nucleotide (Qi *et al.*, 2006; Mi *et al.*, 2008), we applied this type of siRNA. Most importantly, the guide strand of the chosen siRNA was complementary to the mRNA of the green fluorescent protein (GFP). Thus, programmed, functional AGO4/RISC was accordingly expected to show an endonucleolytic activity on this RNA. Indeed, on addition of a ^{32}P-labeled GFP target RNA to *in vitro* BYL reactions, which contained the respective *in vitro*-translated AGO4 variants, specific cleavage of this RNA was observed (Fig. 5). Importantly, all reactions, irrespective of whether they used the two rye AGO4B-like proteins or AtAGO4, showed the same pattern of cleavage products, that is, a larger 5′ product, which was clearly visible, and a 3′ product (*c.* 100 nucleotides), which was less easy to detect. Additional fragments with sizes of just above 100 nucleotides were suspected to result from further processing of the 5′ cleavage product. Specific cleavage products were not detected when the reactions were performed in the absence of the GFP-specific siRNA or when *AGO4* mRNA translation was omitted. Hence, we conclude that the AGO4B-like proteins, regardless of whether they are encoded by the As or Bs of rye, are able to form active RISC. Thus, each protein shows a slicer activity that is comparable and homologous to the activity of the AtAGO4-containing RISC.

Discussion

Bs of rye are transcriptionally active

The detection of active RNAPII in the proximity of B chromatin, the RNA-seq analysis-based identification of B transcripts in different tissues of rye and the characterization of B-located genes undergoing pseudogenization are in accordance with previous data on the transcription of B-located coding and non-coding sequences (Carchilan *et al.*, 2007, 2009; Banaei-Moghaddam *et al.*, 2013, 2015).

With increasing numbers of Bs, the amount of Sc*KIF4A*, Sc*SHOC1* and Sc*AGO4B* transcripts encoded by As and Bs changed. Sc*KIF4A* transcripts decreased with increasing numbers of Bs, with the exception that a significant increase was observed in the presence of 2Bs. For Sc*AGO4B,* the presence of 1B greatly increased the activity, which then decreased to 4B in a zig-zag-like pattern. This pattern could reflect the already described odd–even effect of B-containing organisms. Strong effects associated with odd numbers of Bs have been shown previously in different species (Jones & Rees, 1982). In the case of Sc*SHOC1*, the presence of Bs did not change the expression level significantly. A certain higher expression appeared in the presence of 2Bs, but the increase was less pronounced than that found for the other two genes analyzed.

In general, the quantitative analyses of the relative contribution of A- and B-originating transcripts showed that the amount of B-originating transcripts increased with increasing B number, whereas the amount of A-originating transcripts decreased. However, there were two exceptions. The B-derived Sc*SHOC1* transcripts decreased in 4Bs after increasing from 1B to 3Bs and, in comparison with 0B, the total expression of Sc*AGO4B* increased six-fold in the presence of one deficient B. As 54% of transcripts originate from A, this indirectly implies that there should also be an increase in A-originating expression. Similarly, expression changes caused by aneuploidy often have genome-wide impacts, and the spectrum of expression changes is often complex (Birchler, 2014). In maize, the compensation for the majority of genes on aneuploid chromosomes occurs at the RNA level (Guo *et al.*, 1996). In plants, *de novo*-formed genes show a high epigenetic variation among different individuals, which is reflected in their expression patterns. It seems that newly formed genes are especially prone to epigenetic variation (Silveira *et al.*, 2013). However, it is not yet known whether this is also true for B-located coding sequences. Another possibility is that the promoter regions of B-located sequences may have a higher affinity to transcription factors or RNAPII, causing the suppression of A-located genes in competition for regulatory factors.

Rye Bs undergo pseudogenization

Our comparison of A-located genes with their B-located counterparts showed that the mutation rate was different among the analyzed candidates. A high degree of polymorphisms may indicate the location of such genic fragments on Bs at the onset of B evolution. Alternatively, according to the gene balance hypothesis (Birchler, 2014), the activity of such genes may interfere with their A counterparts. As a result, rapid inactivation of the B genes after gene duplication and B insertion may have occurred. A subsequent integration of duplicated genes after proto-B formation might explain the presence of fragments with fewer polymorphisms.

The B-derived Sc*KIF4A* gene shows the best evidence of pseudogenization, as a premature stop codon was found for one B-encoded copy. However, the existence of functional active B-located Sc*KIF4A* variants cannot be excluded as different B-located Sc*KIF4A* variants were found. In addition to being derived from ancestral founder As, B-located genic sequences could also originate from hitchhiking of genomic fragments from As via transposable elements, as demonstrated for non-collinear Triticeae genes (Wicker *et al.*, 2011).

Despite degeneration, a pseudogene can still be functional by the regulation of its parental gene (Pennisi, 2012), for example, via encoded endo-siRNAs (Johnsson *et al.*, 2014). Probably, as

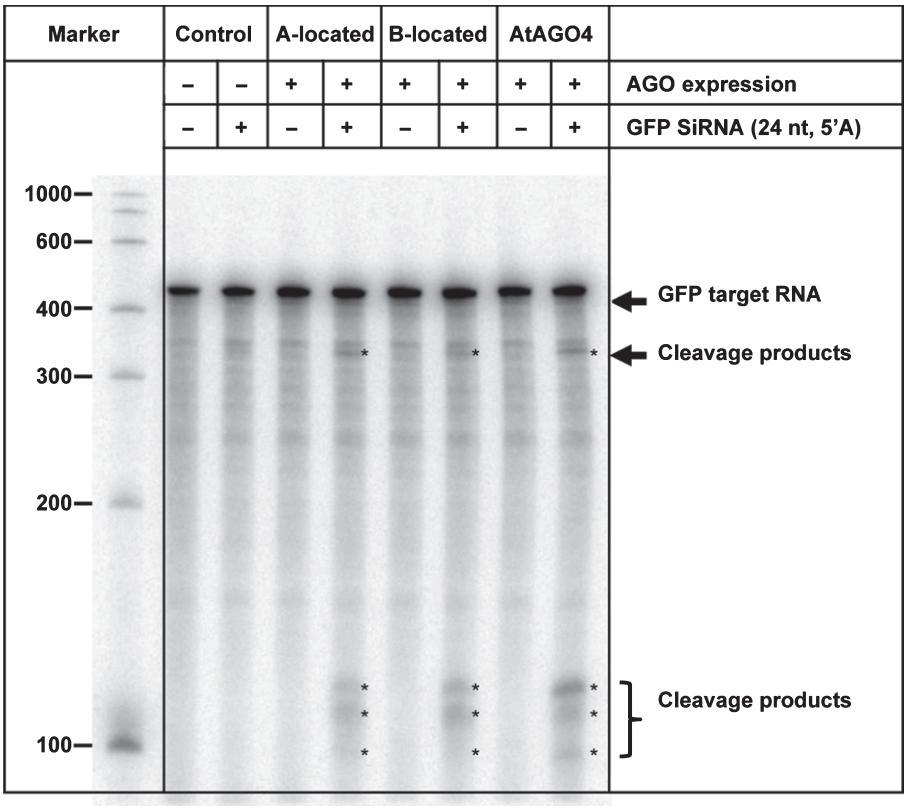

Fig. 5 Rye A- and B-derived AGO4B-like proteins show similar slicer activity. The mRNAs encoding the A- and B-derived ScAGO4B-like proteins were translated in *Nicotiana tabacum* BY-2 lysate (BYL) in the absence or presence of an exogenous, 24-nucleotide small interfering RNA (siRNA) targeting the mRNA of green fluorescent protein (GFP). Subsequently, a ^{32}P-labeled GFP mRNA fragment was added as an AGO4/RNA-induced silencing complex (RISC) target and cleavage products were analyzed by denaturing polyacrylamide gel electrophoresis (PAGE) and autoradiography. As negative and positive controls, the reactions were performed in the absence and presence of additionally expressed (*in vitro* translated) AGO and the mRNA encoding the *Arabidopsis thaliana* AGO4, respectively.

B-specific transcripts would be aberrant as a result of less selective pressure, they may serve as a substrate for RNA-directed RNA polymerases producing double-stranded RNAs (dsRNAs). These dsRNAs may then be processed to small regulatory RNAs. Alternatively, pseudogene transcripts may function as indirect post-transcriptional regulators by acting as miRNA 'sponges'. As a result of the high similarity between parental and pseudogene transcripts, both could compete for miRNA, leading to the degradation of the parental gene transcripts (Muro *et al.*, 2011). Further, it has been shown that some pseudogene transcripts may be translated and then produce short functional peptides or truncated proteins (Johnsson *et al.*, 2014). Comparative profiling of small RNAs from individuals with and without Bs could address the question of whether B-derived regulatory siRNAs acting as regulators are indeed formed.

Could any duplicated gene be located on Bs? Dosage-sensitive genes are less likely to be part of an evolving B, as duplication of a B donor fragment containing such genes could result in a detrimental phenotype. By contrast, dosage-insensitive genes, such as structural genes, having no regulatory roles for transcription or translation (Bellott *et al.*, 2014), could be B-hosted, lost or undergo pseudogenization. Duplicated genes are often associated with detrimental effects and thus may be removed by natural selection. As most of the mutations are degenerative, it is more likely that a duplicated gene will undergo inactivation rather than acquire new functions. Nevertheless, there is evidence for the beneficial role of duplicated genes, especially under stress conditions (Tang & Amon, 2013). Under certain circumstances, the selective retention of duplicated genes may occur, that is, when their redundancy protects corresponding parental genes from immediate detrimental mutations, or when over-dominance between their products exists. In the second case, duplicated genes could convert to new genes by achieving beneficial mutations (Katju & Bergthorsson, 2013). Thus, B-located duplicated genes may accelerate the evolution of Bs. The potential function of the active gene on rye Bs discovered here and the mechanism of its regulation, especially in response to the presence of a different number of Bs, remain to be answered.

In line with previous findings on the B-specific amplification of coding sequences in Lake Victoria cichlid fishes (Yoshida *et al.*, 2011; Valente *et al.*, 2014) and different Canidae species (Graphodatsky *et al.*, 2005; Becker *et al.*, 2011; Trifonov *et al.*, 2013), the B-coding sequences of rye also displayed an increase in copy number. In analogy with homogeneously stained regions and double minute chromosomes, which exist in addition to the normal chromosome complement in vertebrates, and chromosome aberrations that are characterized by the amplification of sequences, in agreement with Makunin *et al.* (2014), we propose that the mechanisms involved in the sequence amplification on Bs may be common for these genomic elements.

The *in vitro* analysis of the A- and B-encoded ScAGO4B protein variants revealed that both operate a slicer activity, similar to that observed for *A. thaliana* AGO4. These data demonstrated unambiguously the presence of a functional Sc*AGO4B* gene on rye Bs and that these Bs carry both functional protein coding genes as well as pseudogene copies. What may be the functional consequences of an additional AGO4 gene being expressed via B

in rye? In concert with Pol IVA and Pol IVB, the RNA-dependent RNA polymerases RDR2 and RDR5, the Dicer-like protein DCL3 and DNA methyl transferases, AGO4 is indicated to function in RNA-directed DNA methylation (RdDM), for example, by controlling the maintenance of epigenetically silent states at repeated loci, transposons and heterochromatin (reviewed in Vaucheret, 2008; Zhang *et al.*, 2015). Transcripts of Pol IV are copied into dsRNAs by RDR2, and these dsRNAs are then processed by DCL3 into 24-nucleotide siRNAs, which are preferentially bound by AGO4. The AGO4-bound siRNAs may then guide the targeting of nascent scaffold transcripts from RNA polymerase V by sequence complementarity and recruit DNA methyltransferase activity to mediate *de novo* methylation (reviewed in Zhang *et al.*, 2015). As its functional presence was found to be important in the defense response of the plant cell during bacterial and viral infections, AGO4 has also been proposed to act as a linker of the transcriptional and post-transcriptional silencing pathways (Agorio & Vera, 2007; Ye *et al.*, 2009; Hamera *et al.*, 2012). Additional AGO4 pseudogenes on Bs may thus be important to ensure the silencing of certain sequence elements, such as transposons.

In short, this study suggests that B-encoded (pseudo)genes may provide an additional level of gene control and complexity in combination with their related A-located genes. Hence, physiological effects, associated with the presence of Bs, may be explained by the activity of B-located (pseudo)genes.

Acknowledgements

We are grateful to all members of the Chromosome Structure & Function Group (IPK, Gatersleben, Germany) for fruitful discussions: Susann Hesse and Lala Aliyeva-Schnorr for supporting the FISH experiments; Marie Kubaláková, Zdeňka Dubská and Romana Šperková for sorting B and DNA amplification; Karla Meier, Katrin Kumke and Oda Weiß for excellent technical assistance; Anne Fiebig for the submission of sequence data to EMBL/ENA; Eva Bauer (Technical University of Munich, Munich, Germany) for early access to the rye WGS assembly. This work was supported by the China CSC scholarship, DFG Germany (grant nos. HO 1779/10-1/14-1, CRC 648 C1 and A7), National Program of Sustainability (grant no. LO 2014 to J.V. and J.D.) and the Czech Science Foundation (grant number P501-12-G090).

Author contributions

W.M., A.M.B-M. and A.Ho. designed the research; W.M., T.S.G., T.G., J.V., J.D., H.G., L.A. and A.M.B-M. performed the experiments. W.M., M.M.M., V.S., S-E.B., A.Hi, U.S. and A.M.B-M. analyzed the data; W.M., M.M.M., T.G., V.S., A.M.B-M., S-E.B. and A.Ho wrote the article.

References

Agorio A, Vera P. 2007. ARGONAUTE4 is required for resistance to *Pseudomonas syringae* in *Arabidopsis*. *Plant Cell* **19**: 3778–3790.

Banaei-Moghaddam AM, Martis MM, Macas J, Gundlach H, Himmelbach A, Altschmied L, Mayer KF, Houben A. 2015. Genes on B chromosomes: old questions revisited with new tools. *Biochimica et Biophysica Acta* **1849**: 64–70.

Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R, Houben A. 2013. Formation and expression of pseudogenes on the B chromosome of rye. *Plant Cell* **25**: 2536–2544.

Becker SE, Thomas R, Trifonov VA, Wayne RK, Graphodatsky AS, Breen M. 2011. Anchoring the dog to its relatives reveals new evolutionary breakpoints across 11 species of the Canidae and provides new clues for the role of B chromosomes. *Chromosome Research* **19**: 685–708.

Beier S, Himmelbach A, Schmutzer T, Felder M, Taudien S, Mayer KF, Platzer M, Stein N, Scholz U, Mascher M. 2015. Multiplex sequencing of bacterial artificial chromosomes for assembling complex plant genomes. *Plant Biotechnology Journal* **14**: 1511–1522.

Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho TJ, Koutseva N, Zaghlul S, Graves T, Rock S *et al.* 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**: 494–499.

Birchler JA. 2014. Facts and artifacts in studies of gene expression in aneuploids and sex chromosomes. *Chromosoma* **123**: 459–469.

Bougourd SM, Jones RN. 1997. B chromosomes: a physiological enigma. *New Phytologist* **137**: 43–54.

Carbonell A, Carrington JC. 2015. Antiviral roles of plant ARGONAUTES. *Current Opinion in Plant Biology* **27**: 111–117.

Carchilan M, Delgado M, Ribeiro T, Costa-Nunes P, Caperta A, Morais-Cecílio L, Jones RN, Viegas W, Houben A. 2007. Transcriptionally active heterochromatin in rye B chromosomes. *Plant Cell* **19**: 1738–1749.

Carchilan M, Kumke K, Mikolajewski S, Houben A. 2009. Rye B chromosomes are weakly transcribed and might alter the transcriptional activity of A chromosome sequences. *Chromosoma* **118**: 607–616.

Carlson W. 2009. The B chromosome of maize. In: Bennetzen JJ, Hake S, eds. *Handbook of maize*. New York, USA: Springer, 459–480.

Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, Schmutz J, Taga M, White GJ, Zhou S *et al.* 2009. The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genetics* **5**: e1000618.

Donald TM, Houben A, Leach CR, Timmis JN. 1997. Ribosomal RNA genes specific to the B chromosomes in *Brachycome dichromosomatica* are not transcribed in leaf tissue. *Genome* **40**: 674–681.

Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**: 257–258.

Graphodatsky AS, Kukekova AV, Yudkin DV, Trifonov VA, Vorobieva NV, Beklemisheva VR, Perelman PL, Graphodatskaya DA, Trut LN, Yang F. 2005. The proto-oncogene C-KIT maps to canid B-chromosomes. *Chromosome Research* **13**: 113–122.

Green DM. 1990. Muller's Ratchet and the evolution of supernumerary chromosomes. *Genome* **33**: 818–824.

Guo M, Davis D, Birchler JA. 1996. Dosage effects on gene expression in a maize ploidy series. *Genetics* **142**: 1349–1355.

Hamera S, Song X, Su L, Chen X, Fang R. 2012. Cucumber mosaic virus suppressor 2b binds to AGO4-related small RNAs and impairs AGO4 activities. *Plant Journal* **69**: 104–115.

Houben A, Schroeder-Reiter E, Nagaki K, Nasuda S, Wanner G, Murata M, Endo TR. 2007. CENH3 interacts with the centromeric retrotransposon cereba and GC-rich satellites and locates to centromeric substructures in barley. *Chromosoma* **116**: 275–283.

Huang W, Du Y, Zhao X, Jin W. 2016. B chromosome contains active genes and impacts the transcription of A chromosomes in maize (*Zea mays* L.). *BMC Plant Biology* **16**: 1.

Iki T, Yoshikawa M, Nishikiori M, Jaudal MC, Matsumoto-Yokoyama E, Mitsuhara I, Meshi T, Ishikawa M. 2010. *In vitro* assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Molecular Cell* **39**: 282–291.

Irvine DV, Zaratiegui M, Tolia NH, Goto DB, Chitwood DH, Vaughn MW, Joshua-Tor L, Martienssen RA. 2006. Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313**: 1134–1137.

Jimenez M, Romera E, Puertas M, Jones R. 1994. B-chromosomes in inbred lines of rye (*Secale cereale* L.). *Genetica* **92**: 149–154.

Johnsson P, Morris KV, Grander D. 2014. Pseudogenes: a novel source of trans-acting antisense RNAs. *Methods in Molecular Biology* **1167**: 213–226.

Jones RN. 1995. Tansley review no 85. B chromosomes in plants. *New Phytologist* **131**: 411–434.

Jones RN, Houben A. 2003. B chromosomes in plants: escapees from the A chromosome genome? *Trends in Plant Science* **8**: 417–423.

Jones RN, Puertas MJ. 1993. The B-chromosomes of rye (*Secale cereale* L.). In: Dhir KK, Sareen TS, eds. *Frontiers in plant science research*. Dehli, India: Bhagwati Enterprises Dehli (India), 81–112.

Jones RN, Rees H. 1982. *B chromosomes, 1st edn*. London, UK, New York, NY, USA: Academic Press.

Katju V, Bergthorsson U. 2013. Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Frontiers in Genetics* **4**: 273.

Kishikawa H. 1965. Cytogenetic studies of B chromosomes in rye, *Secale cereale* L. Japan. *Agricultural Bulletin of Saga University* **21**: 1–81.

Komarnitsky P, Cho EJ, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & Development* **14**: 2452–2460.

Leach CR, Houben A, Field B, Pistrick K, Demidov D, Timmis JN. 2005. Molecular evidence for transcription of genes on a B chromosome in *Crepis capillaris*. *Genetics* **171**: 269–278.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Lindström J. 1965. Transfer to wheat of accessory chromosomes from rye. *Hereditas* **54**: 149–155.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y *et al.* 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* **1**: 18.

Ma L, Vu GT, Schubert V, Watanabe K, Stein N, Houben A, Schubert I. 2010. Synteny between *Brachypodium distachyon* and *Hordeum vulgare* as revealed by FISH. *Chromosome Research* **18**: 841–850.

Macaisne N, Novatchkova M, Peirera L, Vezon D, Jolivet S, Froger N, Chelysheva L, Grelon M, Mercier R. 2008. SHOC1, an XPF endonuclease-related protein, is essential for the formation of class I meiotic crossovers. *Current Biology* **18**: 1432–1437.

Makunin AI, Dementyeva PV, Graphodatsky AS, Volobouev VT, Kukekova AV, Trifonov VA. 2014. Genes on B chromosomes of vertebrates. *Molecular Cytogenetics* **7**: 1–10.

Martinez G, Slotkin RK. 2012. Developmental relaxation of transposable element silencing in plants: functional or byproduct? *Current Opinion in Plant Biology* **15**: 496–502.

Martis MM, Klemme S, Banaei-Moghaddam AM, Blattner FR, Macas J, Schmutzer T, Scholz U, Gundlach H, Wicker T, Simkova H *et al.* 2012. Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. *Proceedings of the National Academy of Sciences, USA* **109**: 13343–13346.

Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C. 2008. Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5′ terminal nucleotide. *Cell* **133**: 116–127.

Moss JP. 1966. The adaptive significance of B-chromosomes in rye. *Chromosomes Today* **1**: 15–23.

Müntzing A. 1963. Effects of accessory chromosomes in diploid and tetraploid rye. *Hereditas* **49**: 371–426.

Muro EM, Mah N, Andrade-Navarro MA. 2011. Functional evidence of post-transcriptional regulation by pseudogenes. *Biochimie* **93**: 1916–1921.

Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a *de novo* assembly approach to fill the gap within paired reads. *BMC Bioinformatics* **13**: S8.

Pennisi E. 2012. ENCODE project writes eulogy for junk DNA. *Science* **337**: 1159–1161.

Poletto AB, Ferreira IA, Martins C. 2010. The B chromosomes of the African cichlid fish *Haplochromis obliquidens* harbour 18S rRNA gene copies. *BMC Genetics* **11**: 1.

Pop M, Kosack DS, Salzberg SL. 2004. Hierarchical scaffolding with bambus. *Genome Research* **14**: 149–159.

Qi Y, He X, Wang XJ, Kohany O, Jurka J, Hannon GJ. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**: 1008–1012.

R Core Team. 2014. *R version 3.0.2: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2012.

Ribeiro T, Pires B, Delgado M, Viegas W, Jones N, Morais-Cecílio L. 2004. Evidence for 'cross–talk' between A and B chromosomes of rye. *Proceedings of the Royal Society of London B: Biological Sciences* **271**: S482–S484.

Roberts A, Pimentel H, Trapnell C, Pachter L. 2011. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**: 2325–2329.

Ruban A, Fuchs J, Marques A, Schubert V, Soloviev A, Raskina O, Badaeva E, Houben A. 2014. B chromosomes of *Aegilops speltoides* are enriched in organelle genome-derived sequences. *PLoS One* **9**: e90214.

Ruiz-Estévez M, Badisco L, Broeck JV, Perfectti F, Lopez-Leon MD, Cabrero J, Camacho JP. 2014. B chromosomes showing active ribosomal RNA genes contribute insignificant amounts of rRNA in the grasshopper *Eyprepocnemis plorans*. *Molecular Genetics and Genomics* **289**: 1209–1216.

Ruiz-Estévez M, Lopez-Leon MD, Cabrero J, Camacho JPM. 2012. B-chromosome ribosomal DNA is functional in the grasshopper *Eyprepocnemis plorans*. *PLoS One* **7**: e36600.

Sandery MJ, Forster JW, Blunden R, Jones RN. 1990. Identification of a family of repeated sequences on the rye B-chromosome. *Genome* **33**: 908–913.

Schuck J, Gursinsky T, Pantaleo V, Burgyán J, Behrens SE. 2013. AGO/RISC-mediated antiviral RNA silencing in a plant *in vitro* system. *Nucleic Acids Research* **41**: 5090–5103.

Sekine Y, Okada Y, Noda Y, Kondo S, Aizawa H, Takemura R, Hirokawa N. 1994. A novel microtubule-based motor protein (KIF4) for organelle transports, whose expression is regulated developmentally. *Journal of Cell Biology* **127**: 187–201.

Siegel JJ, Amon A. 2012. New insights into the troubles of aneuploidy. *Annual Review of Cell and Developmental Biology* **28**: 189–214.

Silva DM, Pansonato-Alves JC, Utsunomia R, Araya-Jaime C, Ruiz-Ruano FJ, Daniel SN, Hashimoto DT, Oliveira C, Camacho JP, Porto-Foresti F *et al.* 2014. Delimiting the origin of a B chromosome by FISH mapping, chromosome painting and DNA sequence analysis in *Astyanax paranae* (Teleostei, Characiformes). *PLoS ONE* **9**: e94896.

Silveira AB, Trontin C, Cortijo S, Barau J, Del-Bem LEV, Loudet O, Colot V, Vincentz M. 2013. Extensive natural epigenetic variation at a *de novo* originated gene. *PloS Genetics* **9**: e1003437.

Sims RJ, Mandal SS, Reinberg D. 2004. Recent highlights of RNA-polymerase-II-mediated transcription. *Current Opinion in Cell Biology* **16**: 263–271.

Szczerbal I, Switonski M. 2003. B chromosomes of the Chinese raccoon dog (*Nyctereutes procyonoides procyonoides* Gray) contain inactive NOR-like sequences. *Caryologia* **56**: 213–216.

Tang YC, Amon A. 2013. Gene copy-number alterations: a cost–benefit analysis. *Cell* **152**: 394–405.

Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**: 635–641.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.

Trifonov VA, Dementyeva PV, Larkin DM, O'Brien PCM, Perelman PL, Yang F, Ferguson-Smith MA, Graphodatsky AS. 2013. Transcription of a protein-coding gene on B chromosomes of the Siberian roe deer (*Capreolus pygargus*). *BMC Biology* **11**: 90.

Valente GT, Conte MA, Fantinatti BE, Cabral-de-Mello DC, Carvalho RF, Vicari MR, Kocher TD, Martins C. 2014. Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Molecular Biology and Evolution* **31**: 2061–2072.

Vaucheret H. 2008. Plant argonautes. *Trends in Plant Science* **13**: 350–358.

Wicker T, Mayer KF, Gundlach H, Martis M, Steuernagel B, Scholz U, Simkova H, Kubalakova M, Choulet F, Taudien S *et al.* 2011. Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *Plant Cell* **23**: 1706–1718.

Ye J, Qu J, Zhang JF, Geng YF, Fang RX. 2009. A critical domain of the cucumber mosaic virus 2b protein for RNA silencing suppressor activity. *FEBS Letters* **583**: 101–106.

Yoshida K, Terai Y, Mizoiri S, Aibara M, Nishihara H, Watanabe M, Kuroiwa A, Hirai H, Hirai Y, Matsuda Y *et al.* 2011. B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. *PLoS Genetics* 7: e1002203.

Zhang H, Xia R, Meyers BC, Walbot V. 2015. Evolution, functions, and mysteries of plant ARGONAUTE proteins. *Current Opinion in Plant Biology* **27**: 84–90.

Zhu C, Jiang W. 2005. Cell cycle-dependent translocation of PRC1 on the spindle by Kif4 is essential for midzone formation and cytokinesis. *Proceedings of the National Academy of Sciences, USA* **102**: 343–348.

Zilberman D, Cao X, Jacobsen SE. 2003. *ARGONAUTE4* control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**: 716–719.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Phylogenetic analysis of kinesin (KIF), shortage in chiasmata 1 (SHOC1) and Argonaute (AGO).

Fig. S2 Gene structural 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 Sc*KIF4*A-like fragments were subjected to structural modifications.

Fig. S5 Chromosomal location of Sc*KIF4A* is outside of the non-disjunction controlling region.

Fig. S6 Cleaved amplified polymorphic sequences (CAPS) analysis for A- and B-derived Sc*KIF4A* transcripts.

Fig. S7 Cleaved amplified polymorphic sequences (CAPS) analysis for A- and B-derived Sc*SHOC1* transcripts.

Fig. S8 Gene structural model of A- and B-derived Sc*AGO4B* transcripts.

Fig. S9 Domain structure of ScAGO4B protein.

Table S1 List of primer sequences for PCR, reverse transcription-polymerase chain reaction (RT-PCR), fluorescence *in situ* hybridization (FISH) probe preparations, *in vitro* transcription and quantitative RT-PCR

Table S2 List of sequence identifiers and description of kinesin (KIF), shortage in chiasmata 1 (SHOC1) and Argonaute (AGO) sequences used for phylogenetic tree construction

Table S3 Accession numbers of Sc*KIF4A*, Sc*SHOC1* and Sc*AGO4B* genes

Table S4 Categories of B-located transcripts

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

Table S6 The functional annotation and classification of B-chromosome-located gene candidates in generative (anther) tissue

Methods S1 Details on plant cultivation, *in silico* analysis of B transcripts, genomic DNA and RNA extraction, PCR and reverse transcription-polymerase chain reaction (RT-PCR), as well as the target cleavage *in vitro* assay.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.