Bread Wheat TaSPO11-1 exhibits evolutionary conserved function in meiotic recombination across distant plant species

Olivier da Ines, Robin Michard, Ian Fayos, Giacomo Bastianelli, Alain Nicolas, Emmanuel Guiderdoni, Charles I. White, Pierre Sourdille

Bread Wheat TaSPO11-1 exhibits evolutionary conserved function in meiotic recombination across distant plant species.

Da Ines Olivier[1*], Michard Robin[2,5*], Fayos Ian[3,4,5], Bastianelli Giacomo[5], Nicolas Alain[5,6], Guiderdoni Emmanuel[3,4], White Charles[1], Sourdille Pierre[2]

[1]Université Clermont Auvergne, CNRS, Inserm, GReD, F-63000 Clermont–Ferrand, France.

[2]Genetics, Diversity & Ecophysiology of Cereals, INRA, Université Clermont-Auvergne, 63000, Clermont-Ferrand, France

[3]CIRAD, UMR AGAP, 34398 Montpellier Cedex 5 France

[4]Université de Montpellier, CIRAD, INRA, Montpellier SupAgro, 34398 Montpellier, France

[5]Meiogenix, 27 rue du Chemin Vert, 75011 Paris, France

[6]Institut Curie, Centre de recherche, CNRS UMR 3244, PSL University, 26 rue d'Ulm, 75248 Paris Cedex 05, France

*: the two first authors contributed equally to this work.

Corresponding author: Pierre SOURDILLE

INRA, UMR 1095 Genetic, Diversity and Eco-physiology of Cereals,

5, Chemin de Beaulieu

63000 Clermont-Ferrand, France

Phone: 33 4 43 76 15 17

Fax: 33 4 43 76 15 10

E-mail: pierre.sourdille@inra.fr

Word count: 6016

With 1 Table & 11 Figures & 6 Supplementary Materials

Summary – Abstract (250 words)

The manipulation of meiotic recombination in crops is essential to rapidly develop new plant varieties, producing more in a sustainable manner. One option is to control the formation and repair of the meiosis-specific DNA double strand breaks (DSBs) that initiate recombination between the homologous chromosomes and ultimately lead to crossovers. These DSBs are introduced by the evolutionarily conserved topoisomerase-like protein SPO11 and associated proteins. Here, we characterized the homoeologous copies of the SPO11-1 protein in hexaploid bread wheat (*Triticum aestivum* L.). The genome contains three *SPO11-1* gene copies that exhibit 93-95% identity at the nucleotide level and clearly, the A and D copies originated from the diploid ancestors *Triticum urartu* and *Aegilops tauschii*, respectively. Further, phylogenetic analysis of 105 plant genomes revealed a clear partitioning between monocots and dicots, with the seven main motifs being almost fully conserved even between clades. The functional similarity of the proteins among monocots was confirmed through complementation analysis of the rice *spo11-1* mutant by the wheat *TaSPO11-1-5D* coding sequence. Also, remarkably, albeit the wheat and Arabidopsis SPO11-1 proteins share only 55% identity and the partner proteins also differ, the *TaSPO11-1-5D* cDNA significantly restored the fertility of the Arabidopsis *spo11-1* mutant indicating a robust functional conservation of the SPO11-1 protein activity, across distant plants. These successful heterologous complementation assays, using the Arabidopsis and rice hosts, are good surrogates to validate the functionality of candidate genes and cDNA, as well as variant constructs, when the transformation and mutant production in wheat is much longer and tedious.

Significance statement (69 words)

We analysed the three homoeologous copies of *SPO11-1* genes in the bread wheat genome. They are very similar but not identical, revealing their parental origin from diploid wheats. These SPO11-1 proteins contain the seven essential and evolutionary conserved motifs, now found in 105 plant species. The coding sequence of the wheat D-copy significantly complemented the rice and

Arabidopsis *spo11-1* mutants indicating a strong functional conservation between highly divergent species.

Keywords

SPO11-1; Meiotic recombination; Protein evolution; Wheat; Grasses.

Introduction

The improvement of genetic diversity in cultivated species is of utmost importance especially towards the development of a sustainable agriculture using for example lower amounts of water, fertilizers, pesticides or fungicides. One way to reach this challenge is explore the tremendous gene resources that already exist in nature. This can be achieved by crossing the elite lines with related and/or exotic varieties bearing advantageous alleles of agronomical interest. Such genetic admixture occurs naturally through meiotic recombination and gamete formation. The mechanisms of meiotic recombination leading to crossover products (COs) have long been studied in model organisms, including plants (reviewed by (Mercier *et al.* 2015)). Importantly, crossover recombination not only generates new combinations of parental alleles, but also ensures the faithful distribution of the homologous chromosomes during the reductional meiotic division (MI), key to the success of sexual reproduction and fertility (Hunter 2015, Mercier, *et al.* 2015).

Meiotic recombination is initiated by the formation of programmed DNA double-strand breaks (DSBs) catalysed by the topoisomerase-like protein SPO11 (Bergerat *et al.* 1997, Keeney *et al.* 1997); for reviews, see (de Massy 2013, Lam and Keeney 2014, Robert *et al.* 2016b)). Ressembling the archaeal topoisomerase VI complex, SPO11 (ortholog of the archaea A subunit) is likely working in a protein complex with one or several other proteins (Lam and Keeney 2014), in particular with the MTOPVIB protein (orthologous to the archaea B subunit) to form an active catalytic complex (Figure 1) (Robert *et al.* 2016a, Vrielynck *et al.* 2016).

SPO11 is evolutionarily conserved among sexually reproducing organisms and is encoded by a single gene in most organisms (Malik *et al.* 2007). However, there are several *SPO11* genes in plant genomes. The Arabidopsis genome carries three related genes named *SPO11-1*, *SPO11-2* and *SPO11-3* (Hartung and Puchta 2000, Grelon *et al.* 2001, Hartung and Puchta 2001, Sprink and Hartung 2014). Both *SPO11-1* and *SPO11-2* are essential for meiotic recombination (Grelon, *et al.* 2001, Hartung and Puchta 2001, Stacey *et al.* 2006, Hartung *et al.* 2007) while *SPO11-3* is involved in somatic endo-reduplication (Hartung *et al.* 2002, Sugimoto-Shirasu *et al.* 2002, Yin *et al.* 2002). The situation in rice (*Oryza sativa*) is even more complex with five *SPO11*-related genes (Jain *et al.* 2006, Yu *et al.* 2010,

An *et al.* 2011). Of these rice genes, to date only *OsSPO11-1* and *OsSPO11-2* have been clearly demonstrated to be required for meiotic recombination (Yu, *et al.* 2010, Fayos *et al.* 2019). The interactions between SPO11-1, SPO11-2 and MTOPVIB proteins have recently been described in Arabidopsis, rice and mice (Figure 1) (Fu *et al.* 2016, Robert, *et al.* 2016a, Vrielynck, *et al.* 2016, Xue *et al.* 2016).

Another layer of complexity in meiosis in plants is polyploidy, a common feature of plant kingdom, which concerns about 30% of existing flowering plants (Wood *et al.* 2009), including most of the world's important crops. Polyploid species can be divided into autopolyploid species derived from whole genome duplication, such as potato (*Solanum tuberosum*, 4x), strawberry (*Fragaria ananassa*, 8x), and allopolyploid species originating from crosses between closely related species such as oilseed rape (*Brassica napus*, 2n = 4x = 38, AACC), cotton (*Gossypium arboreum*, 2n = 4x = 52, AADD) or tobacco (*Nicotiana tabacum*, 2n = 4x = 48). Allopolyploid species thus contain different sets of related but not completely homologous chromosomes called homoeologues. It is now clearly established that all flowering plants have experienced at least one and usually several rounds of polyploidy (also called whole genome duplications; WGD) during the course of their evolution (Soltis *et al.* 2009, Van de Peer *et al.* 2009). Thus, other angiosperms, including Arabidopsis, rice or Brassica crops, are ancient polyploids (paleo-polyploids) that returned to a functionally diploid state by a massive elimination of some but not all duplicated genes post-WGD (a process called fractionation; (Doyle *et al.* 2008)). Given the prevalence of WGD in plants, it is critical to integrate and extend our knowledge of important biological processes such as meiotic recombination into the field of polyploidy. From this point of view, (Lloyd *et al.* 2014) showed that meiotic genes return to a single copy more rapidly than genome-wide average in Angiosperms. Analysis of the presence and expression of meiotic genes in two recent polyploid species (oilseed rape and bread wheat; ~10,000 years ago) suggested that their loss is passive and is a long-term process.

Among allopolyploid species, hexaploid bread wheat (*Triticum aestivum* L.; 2n = 6x = 42) is derived from two successive interspecific crosses (Blake *et al.* 1999, Huang *et al.* 2002). These involved the three diploid species *T. monococcum* ssp *urartu* (AA), a yet unknown species related to the *Sitopsis* section (BB; the closest one being *Aegilops speltoides*; S genome) and *Aegilops tauschii* (DD). The

recent release of the anchored and annotated wheat whole-genome sequence (IWGSC 2018) thus revealed that more than 60% of the 39,474 genes are present in triplicates and more than 90% have at least two homoeologous copies.

The large size (17 Gb) and complexity of the hexaploid wheat genome makes it challenging for functional analyses of gene, in particular for gene involved in complex biological complex such as meiosis. Until now, only a dozen genes (*TaRAD51*, *TaRAD51C*, *TaRAD51D*, *TaDMC1*, *TaMRE11*, *TaRAD50*, *TaASY1*, *TaZYP1*, *TaPHS1*, *TaPh1*, *TaREC8*, *TaRECQ-7*), among more than 100 known as involved in meiotic recombination in plants, have been cloned and significantly analysed in wheat. These analyses have mainly been limited to comparisons of sequences of homoeologous copies, expression analyses and immunolocalization (Boden *et al.* 2007, de Bustos *et al.* 2007, Boden *et al.* 2009, Devisetty *et al.* 2010, Perez *et al.* 2011, Khoo *et al.* 2012a, Khoo *et al.* 2012b, Ma *et al.* 2018, Gardiner *et al.* 2019). *TaPh1*, which controls homoeologous recombination in wheat, remains the best-defined locus involved in meiosis in wheat (Griffiths *et al.* 2006, Moore 2014, Martin *et al.* 2017).

To date, there is no study analysing *SPO11-1* from a polyploid species. *SPO11-1* is a key gene at the onset of the recombination process. It is therefore essential to understand how it behaves in a polyploid context such as that of bread wheat. In this work, we wanted to know to what extent wheat *SPO11-1* homoeologous copies (TaSPO11-1-5A, TaSPO11-1-5B and TaSPO11-1-5D) are conserved between each other and with those of other plants. We thus sought to identify the hexaploid wheat *SPO11-1* gene and to assess its functionality during meiosis. We isolated the homoeologous (*A*, *B* and *D*) copies of *TaSPO11-1* and performed a phylogenetic analysis with *SPO11-1* from different species. The three copies were compared in detail with those of Arabidopsis and Rice to estimate their conservation. We extend these comparisons in demonstrating the functionality and evolutionary conservation of wheat TaSPO11-1 through heterologous complementation of the corresponding Arabidopsis and Rice mutant lines, i.e. across distant plant species.

Results

SPO11-1 is conserved between the three homoeologous genomes of the hexaploid wheat.

To identify the orthologous copies of SPO11-1 within the bread wheat genome, we used the newly released wheat genome sequence (IWGSC 2018) and the Arabidopsis SPO11-1 protein sequence as a query (At3g13170). Three copies were identified, one on each of the three homoeologous genomes from group 5: TraesCS5A02G391400, TraesCS5B02G396300 and TraesCS5D02G401100 further named *TaSPO11-1-5A*, *TaSPO11-1-5B* and *TaSPO11-1-5D*, respectively. The length of the loci was similar with approximately 3.9 kb for the genomic sequences and 1.2 kb for the coding sequences (CDS) (Table 1). The gene structures, previously described for land plants (Sprink and Hartung 2014) with 15 exons and 14 introns, are identical and exhibit 93-95% identity including the introns (98% for the coding sequences). The three coding sequences which share 99% identity between each other encode proteins of 387, 386 and 387 amino acids, respectively (Table 1; Figure S1). The single Amino acid Polymorphism (SAP) mapped at position 45 and 74 for the 5A copy, 189 and 291 for the 5B copy and 64 and 66 for the 5D copy. In addition, the 5B copy carries an in-frame deletion of 3 nucleotides, removing an Alanine amino acid at position 72 (Figure S1).

The *SPO11-1* genes are conserved between wheat and wheat ancestors.

We then sought to assess the sequences of the bread wheat *SPO11-1* copies with their ancestors. The ancestor of the B genome remains unknown but the wheat *SPO11-1*-5A and 5D sequences could be compared with their respective ancestors *Triticum urartu* and *Aegilops tauschii*, respectively. The *Triticum urartu* (TRIUR3_12346) and the *Aegilops tauschii* (LOC109743941) *SPO11-1* gene and the corresponding CDS sequences are reported in Table 1. The *Triticum urartu* and *Aegilops tauschii* *SPO11-1* genes share the conserved structure of 15 exons and 14 introns. Then, we re-annotated these sequences manually using the newly released wheat sequence as a basis for splicing sites and obtained two newly annotated CDS sequences of 1173 bp for *Triticum urartu* and 1164 bp for *Aegilops tauschii* encoding two proteins of 390 and 387 aa, respectively (Table 1).

Close comparison of the wheat TaSPO11-1-5A and -5D CDS and protein sequences with their ancestors showed that wheat TaSPO11-1-5D and *Aegilops tauschii*'s are 100% identical while TaSPO11-1-5A and *Triticum urartu*'s SPO11-1 share 99% identity (Figure S2). Furthermore, the SAP Cys/Ser at position 64 and Ser/Asp at position 66 between *Triticum urartu* and *Aegilops tauschii*'s SPO11-1 are conserved between the wheat A and D genomes (see above; Figures S1 and S2). These results suggest that the wheat SPO11-1 sequence has not changed since the polyploidization event. The extreme conservation of the homoeologous *SPO11-1* genes and proteins within the Triticeae tribe suggests that they are all functional and presumably under a strong selection pressure in regard to their essential role in meiosis. In accordance with this, expression data showed that all three homoeologous genes are equivalently expressed during meiosis (Martin *et al.* 2018).

The SPO11-1 proteins are highly conserved and shares key protein domains throughout plant kingdom.

To get deeper insight into the evolution of SPO11-1 within plants, we first compared the *Arabidopsis thaliana* and wheat SPO11-1 amino acid sequences (Figure S3). We found significant identity (54-55%) similar to the BLAST-P analysis (e-value = 4e-156) (Table 1; Figure S3). This relies on the presence of seven highly conserved motifs present on the archaebacterial subunit A of topoisomerase VI (Bergerat, *et al.* 1997, Keeney, *et al.* 1997, Diaz *et al.* 2002, Malik, *et al.* 2007), *Arabidopsis thaliana* (Hartung and Puchta 2000, Hartung, *et al.* 2007, Shingu *et al.* 2010), wheat (Figure 2 and Figure S3), and many other organisms (Malik, *et al.* 2007). We note that the first motif that contains the catalytic Tyrosine residue for DSB formation (Tyr103 in Arabidopsis) (Bergerat, *et al.* 1997) is located at position 129 in the SPO11-1 A and D homoeologous wheat copies and at position 128 in the B copy. The second motif contains the invariant Arginine 130 (R156 in the A and D copies and R155 in the B copy) that is essential for the function of SPO11-1 *in vivo* (Diaz, *et al.* 2002, Shingu, *et al.* 2010). The fourth motif contains conserved residues implicated in TopoVI DNA binding activity (Glycine 215 and Arginine 222, 223 and 226 in Arabidopsis) (Shingu, *et al.* 2010). Finally, the seventh motif (10 amino acids) is fully conserved between Arabidopsis and wheat (Figure 2 and

Figure S3). This Toprim domain contains two conserved residues invariant in all species examined in this study (Lysine 332 and Glutamic acid 334 in Arabidopsis) that affect DSB formation in yeast when mutated (Diaz, *et al.* 2002). Interestingly, a 3D predictive structural modelling of the TaSPO11-1-5D protein shows that all seven motifs are linked to each other (Figure 2). This also shows that the essential DNA binding (5Y-CAP) and cleavage (Toprim) domains are clearly physically distinguishable from one another (Figure 3). Altogether, these analyses indicate that all SPO11 key residues and domains are conserved in the wheat homoeologous SPO11-1 proteins.

Then to extend our analysis to other plant species, we retrieved 155 SPO11-1 protein sequences from 153 plant species including monocots and dicots. A curation step based on the accuracy of the ATG and STOP codon position, the splicing and the integrity of the sequences allowed to retain and aligned 107 robust sequences from 105 species (Supplementary material S4). The amino acid sequence similarity ranged from 51% (*Cucumis melo*) to 100% (*Aegilops tauschii*). Compared to the wheat SPO11-1-5D copy, the mean similarity between these plant SPO11-1 proteins reach 65%. They all contain the seven most conserved motifs that landmark the SPO11 orthologs. We thus built a consensus sequence for each motif based on the alignment of the sequences (Figure 4) and calculated the average identities of each motif from the 107 sequences (Figure 4B). The motifs show very strong conservation with identities of 90.3, 89.1, 89.6, 93.3, 95.1, 99.1 and 96.5% for motifs 1, 2, 3, 4, 5, 6 and 7, respectively (Figure 4B). In particular, the active residues described above are fully conserved, except the last arginine in motif 4 (Arg226 in Arabidopsis) that is variant in 11/105 (10.5 %) of the sequences. Among these, the six variants with a Glycine instead of Arginine, specifically belong to the Rosids clade and specifically to the Rosales order. Finally, the alignment of the 107 complete sequences led us to generate a phylogenetic tree which revealed a perfect separation into two groups corresponding to monocots (blue, Figure 5) and dicots (red, Figure 5). Interestingly, *Amborella trichopoda*, sister of the angiosperms (flowering plants), is at the frontier of the groups, suggesting it shares properties of both groups. Altogether, our extensive computational analysis of the SPO11-1 protein sequences in plants highlighted extensive conservation of the 7 key protein sequence motifs as well as limited evolution since the separation of monocots and dicots.

Heterologous expression of the Wheat *TaSPO11-1-5D* coding region restores the fertility of the Rice *spo11-1* mutant.

The strong conservation of SPO11-1 protein sequences across plant species does not, in itself, prove that they are functionally interchangeable between species, especially given that they form parts of a multi-protein complex. To address this issue, we sought to determine whether or not the expression of the bread wheat SPO11-1 protein would rescue the meiotic phenotypes and sterility of the *spo11-1* mutant of another monocotyledonous plant, namely rice. To perform this heterologous complementation assay, the wheat *TaSPO11-1-5D* coding sequence placed under the maize Ubiquitin promoter was introduced in our *Oryza sativa* ssp *japonica* var. Kitaake *Osspo11-1-1* mutant line generated through CRISPR/Cas9 mutagenesis (see Methods; Fayos et al., unpublished). The *Osspo11-1-1* mutation is a frameshift resulting from a single nucleotide (A) insertion in the ATG sequence (ATAG). The homozygous *Osspo11-1-1* mutant lines fail to form chiasma during meiosis and are sterile (Yu, *et al.* 2010, Fayos, *et al.* 2019). Functional complementation can thus be easily visualized as an increase in seed production. Three *spo11-1-1* homozygous plants (out of 33 primary transformants) carrying a single copy of the *UBI::TaSPO11-1-5D* transgene were obtained by transformation and remarkably, were fertile (Figure 6A). In the following T1 generation, the restoration of the rice *spo11-1-1* fertility strictly co-segregated with the presence of the *UBI::TaSPO11-1-5D* transgene. Indeed, comparison of the number of filled spikelets in wild-type plants and in the progeny of two transformants grown in similar conditions shows that Rice *spo11-1-1* mutant plants expressing the *UBI::TaSPO11-1-5D* transgene exhibit a fertility comparable to that of wild-type plants (Figure 6B). Altogether, these results demonstrate that the *TaSPO11-1-5D* coding sequence is functional and the wheat protein can functionally replace the rice SPO11-1 in its essential meiotic function.

Heterologous expression of the Wheat *TaSPO11-1-5D* coding region restores the fertility of the Arabidopsis *spo11-1* mutant.

The success of the wheat-rice interspecies complementation prompted us to investigate whether or not

the wheat *TaSPO11-1-5D* complements the *spo11-1* meiotic defects in Arabidopsis, a more distant species belonging to a different clade. In Arabidopsis, the *Atspo11-1* mutant fails to form meiotic DSBs and exhibits a severe reduction in fertility (Grelon, *et al.* 2001). Thus, we placed the *TaSPO11-1-5D* coding sequence under the control of the Arabidopsis *RAD51* promoter and introduced this construct into the Arabidopsis *SPO11-1/spo11-1-2* heterozygous plants (Figure 7A). The *RAD51* promoter is well expressed in Arabidopsis meiocytes (Chen *et al.* 2010, Yang *et al.* 2011, Walker *et al.* 2018) and known to drive successful complementation of other Arabidopsis meiotic mutants (Da Ines *et al.* 2013). PCR genotyping of the *SPO11-1* locus of 41 *TaSPO11-1* primary transformants showed that 15 were homozygous for the *spo11-1-2* allele (*spo11-1-2/spo11-1-2)*, 15 heterozygous (*SPO11-1/spo11-1-2)* and 11 were wild-type (*SPO11-1/SPO11-1*). Remarkably, 14/15 *spo11-1-2* homozygous plants carrying the *TaSPO11-1-5D* transgene exhibited 15% to 70% fertility, instead of the 5% residual fertility observed in the absence of the transgene (Figure 7C and Table S1). Then, we monitored fertility in the progeny (T2 generation) of four randomly selected T1 plants. Consistently, the restoration of fertility (30 to 80%) strictly co-segregated with the presence of the transgene (Figure 7D and Table S2). Thus, the heterologous expression of the Wheat TaSPO11-1-5D protein is able to restore fertility in the Arabidopsis *spo11-1* mutant.

Restoration of the wild-type meiotic progression in the Arabidopsis *spo11-1* mutant expressing the *TaSPO11-1-5D* transgene.

In wild-type plants, the meiotic chromosomes condense, recombine and synapse during prophase I (Figure 8A-E). Full synapsis of the homologs is seen at late prophase I (Figure 8B). The chromosomes then further condense and five bivalents (homologous chromosome pairs attached together by chiasmata and sister chromatid cohesion) are observed at metaphase I (Figure 8C). Homologous chromosomes then segregate to opposite poles to give two sets of five chromosomes at metaphase II (Figure 8D). Meiosis II then proceeds and gives rise to four balanced haploid nuclei (Figure 8E). In contrast, the *spo11-1* mutants lack DSBs formation, hence recombination, pairing and synapsis of the homologs (Figure 8F-G), manifested by the presence of 10 univalents instead of bivalents at

metaphase I (Figure 8H). Chromosome mis-segregation eventually produces unbalanced metaphase II (Figure 8I) and polyads (Figure 8J). In sharp contrast, the cytogenetic analysis of pollen mother cells from the *spo11-1* plants expressing the *TaSPO11-1-5D* transgene revealed the presence of normal meiotic figures (Figure 8K-O). In particular, 5 bivalents were readily observed at metaphase I (Figure 8M). Subsequent proper homologous chromosome segregation at anaphase I (Figure 8N) followed by separation of sister chromatids during the second equational division resulted in 4 balanced meiotic products (Figure 8O). We note however that in accordance with the partial restoration of the fertility in the complemented plants, most metaphases I exhibited a mixture of bivalents and univalents (Figure 9A). Thus, to examine whether the expression of two copies of the transgene will quantitatively improve the faithful progression of meiosis, we characterized two independent transgenic lines homozygous for the *TaSPO11-1-5D* transgene. Their fertility (~ 40%) remained in the average range of the single copy transgene lines (line 3 and 36, see Figure 7). Cytologically, in the two complemented lines, we observed a mean of 1.7 (n=57) and 2.6 (n=74) bivalents per cell instead of 0.05 bivalents/cell (n = 19) in the *spo11-1* mutants and 5 (n=36) in the wild-type meiocytes (Figure 9B and C). We also noted that a majority of bivalents in the complemented lines exhibited a rod-shaped structure with a single chiasma, although ring-shaped bivalents with at least two chiasmata were also observed (10 to 30% of bivalents for line 3 and 36, respectively). Accordingly, in both complemented lines, the number of chiasmata per cell significantly increased reaching a mean of 2 (± 1.7, n = 57) and 3.6 (± 2.6, n = 74), respectively. This is a strong increase compared to the *spo11-1* mutant (0.05 ± 0.2, n = 19), yet lower than wild-type plants (9.3 ± 0.8, n = 36).

Altogether, these results demonstrate that the wheat SPO11-1 can functionally replace the absence of the Arabidopsis SPO11-1 ortholog, substantially restoring meiotic recombination and normal meiotic progression. Beyond the protein sequence homology, these results demonstrated its evolutionary conserved function.

Meiotic DSBs are formed in *Arabidopsis spo11-1* mutants expressing Wheat *TaSPO11-1-5D* transgene.

The presence of bivalents and chiasmata indicate that meiotic recombination occurs and hence that meiotic DSBs are formed in plants expressing *TaSPO11-1-5D* transgene. Given that fertility is not fully restored, this however suggest that less DSBs might be produced in the complemented lines or, alternatively, that DSBs are repaired without forming COs. We thus sought to analyse the ability of TaSPO11-1-5D to form DSBs in meiotic cells. We performed immunolocalization of the strand-exchange protein RAD51 as a marker for DSB formation in both wild-type and *spo11-1 + TaSPO11-1* complemented plants (line 36). As expected, numerous RAD51 foci were observed in early prophase I cells of wild-type plants (mean of 92 foci per cell, n = 39; Figure 10). RAD51 foci were also observed in pollen mother cell nuclei of *spo11-1* mutant plants expressing *TaSPO11-1-5D*. However, a strong two-fold reduction in RAD51 foci formation was detected in these plants (mean of 45 foci per cell, n = 47; Figure 10). This strongly suggests that DSB levels are reduced in the complemented lines and this may explain the limited complementation by TaSPO11-1.

Wheat SPO11-1-5D functionally interacts with Arabidopsis SPO11-2 and MTOPVIB to induce meiotic recombination.

Current knowledge suggests that SPO11-1 does not exhibit DNA cleavage activity alone but acts in a protein complex, physically interacting with the SPO11-2 and MTOPVIB proteins in Arabidopsis and functionally related orthologs in other organisms, in order to form an active topoisomerase VI-like complex that catalyses meiotic DSB formation (Figure 1, Robert *et al.*, 2016; Vrielynck *et al.*, 2016). So, to determine whether TaSPO11-5D also needs the presence of the Arabidopsis SPO11-2 and MTOPVIB proteins to induce meiotic recombination, we crossed our *spo11-1_TaSPO11-1-5D* transgenic plants with Arabidopsis *spo11-2* or *mtopVIB* mutant lines and analysed the fertility of the double mutants. Clearly, as shown in Figure 11, the presence of the *TaSPO11-1-5D* transgene did not rescue the sterility of the *spo11-1 spo11-2* and *spo11-1 mtopVIB* double mutant plants. This excludes

the possibility that TaSPO11-1 induce DSBs independently of SPO11-2 (non-plant organisms have a single SPO11 protein) and confirms the need for the MTOPVIB to form DSBs.

Discussion

TaSPO11-1 homoeologous copies are highly similar between each other and to those from angiosperms.

To identify *SPO11-1* genes from wheat (*T. aestivum*), we exploited the first assembled and annotated pseudomolecule sequence of the wheat genome (IWGSC 2018). Using *in silico* assignment of the Arabidopsis SPO11-1 protein (At3g13170), we readily identified the three homoeologous copies, mapping on chromosomes 5A, 5B and 5D and indicating good conservation of the protein sequence between the two species. The three homoeologous wheat copies are highly similar with ~95% identity at the nucleotide level for the genomic sequences. This is consistent with data from expression of wheat genes (Ramirez-Gonzalez *et al.* 2018) showing an homoeologous SNP diversity ranging from 95.0% to 97.2% within triads (genes present in only three homoeologous copies).

We also observed that the D copy (TaSPO11-1-5D) is identical to the copy from *Ae. tauschii*, the donor of the D genome, while the A copy (TaSPO11-1-5A) is only very slightly different from that of *T. monococcum ssp urartu*, the donor of the A genome. Divergence between the A and the B genome lineages occurred ~7 million years ago (MYA; (Marcussen *et al.* 2014)) while the D genome diverged from the A and B genomes, 1 to 2 million years after. The two successive polyploidization events giving rise to *T. aestivum* occurred at least 0.58 to 0.82 MYA for the first one and 0.23 to 0.43 MYA for the second one. In addition, it is suggested that only a few accessions of *Ae. tauschii* contributed to the D genome of bread wheat (Giles and Brown 2006).

SPO11-1 proteins are characterized by the presence of several conserved domains (Bergerat, *et al.* 1997, Keeney, *et al.* 1997). Accordingly, these seven domains are highly conserved in all the plant sequences that we examined. In particular, the essential residues for DNA cleavage (Tyr103 in Arabidopsis) or binding (Gly215, Arg222, Arg223 and Arg226 in Arabidopsis) are highly conserved although their position changed slightly according to the total size of the protein, which indeed varies

among species for yet unknown reasons. A previous study using 42 SPO11-1 sequences from land plants, but not wheat, indicated that SPO11-1 is highly conserved in plants (Sprink and Hartung 2014). Here, we analysed SPO11-1 sequences from more than 100 plants, including wheat, and show that SPO11-1 exhibits high sequence identities. In particular, more than 90% identity was observed in the most broadly evolutionary conserved functional domains. Overall, our extensive phylogenetic analyses based on sequence comparison of plant SPO11-1 indicates that this protein evolved slowly and exhibits an evolutionary pattern consistent with known relationships between plant species.

Heterologous complementation analyses reveal functional conservation of the SPO11 complex features.

Beyond the computational analyses of the *SPO11-1* genes, we asked to what extent the function of SPO11-1 is also evolutionary conserved throughout plants, testing the complementation of the rice and Arabidopsis *spo11-1* mutants with the *SPO11-1-5D* cDNA from bread wheat. Wheat and rice are monocots while *Arabidopsis thaliana* is a more distantly related dicot. Remarkably, our data show that expression of the wheat gene was able to complement both rice and Arabidopsis mutants, with full complementation in the former.

Furthermore, our analyses in Arabidopsis show that the complementation by TaSPO11-1 restores DSB formation (RAD51 foci) and recombination (chiasmata) and still requires the presence of the wild-type Arabidopsis *SPO11-2* and *MTOPVIB* genes. Wheat SPO11-1 and Arabidopsis SPO11-2 and MTOPVIB are able to interact and form a functional inter-species complex, albeit resulting in only a partial restoration (10 to 70% with most lines showing 20-40% restoration). Immunolocalization of the RAD51 recombinase indicates that less DSBs are formed in the complemented plants (50% of wild-type level in the tested line) and this likely explains the partial complementation.

Partial complementation in Arabidopsis with Arabidopsis clones has been frequently observed, as for Arabidopsis SPO11-1 (Xue *et al.* 2018). Many factors could influence the efficacy of complementation: the use of a recipient T-DNA mutant plant, T-DNA integration, the choice of the promoter or the use of CDS or genomic sequences. Although this cannot be excluded, we do not think that reduced DSBs formation in our transgenic plants result from lower expression of *TaSPO11-1*

gene. Using the same strategy as for TaSPO11-1, we could show that expression of *TaSPO11-2* cDNA driven by the RAD51 promoter is able to fully restore fertility of the Arabidopsis *spo11-2* mutant (Benyahya et al., unpublished). This indicates that RAD51 promoter allows sufficient transcription of *TaSPO11* genes. Conversely, translation may be affected. This is particularly true since codon usage bias is well known to be different in monocots and dicots (Plotkin and Kudla 2011, Camiolo *et al.* 2015). However, without specific TaSPO11-1 antibodies this hypothesis cannot be tested. In the present case of expression in an heterologous species, an additional key factor for the incomplete complementation and the reduced DSB formation is the amino acid sequence divergence of the transgenic and endogenous proteins that needs to interact in a multi-protein complex. Sufficient restoration of the mutant phenotypes has been obtained to conclude on the formation of functional interactions but the incomplete complementation uncovered subtle deficiencies of interest. For instance, TaSPO11-1 may be less prone to interact with other DSB-associated proteins and to form an active complex. In this context, it will be interesting to individually assay the other members of the SPO11 complex and attempt to co-express in Arabidopsis mutants the wheat TaSPO11-1, TaSPO11-2 and TaMTOPVIB once the likely homoeologous genes and coding regions have been well identified. It will also be interesting to more extensively analyse amino acid sequence divergence and its effect on interaction with other essential DSB-associated endogenous proteins.

Eventually, working with hypomorphic mutants of the meiotic SPO11/MTOPVIB complex with reduced DSBs could be very valuable to better understand the relationship between DSB formation and CO regulation (CO assurance, homeostasis and interference).

Conclusion

In this study, we isolated the three wheat homoeologous copies for SPO11-1: TaSPO11-1-5A, TaSPO11-1-5B, TaSPO11-15D. The three copies are highly similar between each other and with those from diploid ancestors, *T. urartu* and *Ae. tauschii*, suggesting that all three are functional. SPO11-1 protein is very well conserved across angiosperms with conserved domains. Remarkably, due to the high level of similarity, TaSPO11-1-5D protein was able to restore the fertility of rice and Arabidopsis

406 *spo11-1* mutants. This also showed that the wheat proteins could be used (and hence further studied)
407 in other more tractable model plants. This is of particular interest in polyploid species in which the
408 redundancy of function brought by the homoeologous genes and variant alleles adds an additional
409 level of genetic complexity in the wild type context, and where the construction of appropriate mutants
410 remains technically difficult and time-consuming due to the polyploidy of the genome (Ramirez-
411 Gonzalez, *et al.* 2018).

412

Experimental Procedures

Plant material and growth conditions

The rice mutant was obtained through CRISPR/Cas9 genome editing as described in Fayos et al., (unpublished). Rice plants were cultivated in controlled conditions with a temperature of 28°C during the day and 24°C during the night, with 60% hygrometry. The natural light is completed by artificial sodium light (700µmol/m²/s). The *Arabidopsis thaliana spo11-1-2*, *spo11-2-3* and *mtopVIB-2* mutants used in this work have been described previously (Grelon, *et al.* 2001, Hartung, *et al.* 2007, Vrielynck, *et al.* 2016). Arabidopsis plants were grown under the following standard conditions: seeds were stratified in water at 4°C for 2 days and grown on soil or in vitro on 0.8% agar plates, 1% sucrose and half-strength Murashige and Skoog salts (M0255; Duchefa Biochemie). Plants were cultivated in a greenhouse or a growth chamber with a 16/8 hour light/dark cycle, at 23°C and 60% relative humidity.

Recovery and synthesis of *TaSPO11-1*

Wheat genome D *SPO11-1* DNA sequence (TraesCS5A02G391400) was first retrieved through a BLAST analysis research on the newly annotated wheat genome sequence (IWGSC 2018) using the *Arabidopsis thaliana SPO11-1* protein sequence (At3g13170) as an input with basic BLAST parameters. The annotated CDS sequence was determined using Triannot pipeline (Leroy *et al.* 2012). CDS sequence was synthesized with a short additional sequence at the 5' end of the gene (coding for the peptide PEFMAMEAPGIR) and flanked with GATEWAY attB sites. Synthesized product was inserted into pDONR/Zeo and further verified by sequencing.

3D modelling and rendering

3D structural model of TaSPO11-1-5D protein was generated by homology modelling on PHYRE2 online pipeline ((Kelley *et al.* 2015), http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) with intensive modelling mode and TaSPO11-1-5D full sequence as amino acid sequence. Rendering was made with PyMOL 2.3.3 software.

Recovery of SPO11-1 protein sequences from multiple plant species, alignment, logo and phylogeny

We retrieved all plant SPO11-1 sequences using Wheat and Arabidopsis SPO11-1 sequences as query on NCBI BLAST website (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with nr database. The Barley SPO11-1 sequence was retrieved using IPK Barley BLAST Server website (https://webblast.ipk-gatersleben.de/barley_ibsc/viroblast.php) and the Barley AA (HC and LC) Morex v2.0 database. Multiple Sequence Alignment was done using ClustalW with basic parameters. Motifs logo were built using WebLogo website (https://weblogo.berkeley.edu/) (Crooks *et al.* 2004). Phylogeny was done with the following settings: ClustalW alignment and PHYLIP Neighbor Joining for the construction of the tree.

Cloning of *TaSPO11-1* and plant transformation

For rice complementation, a LR GATEWAY recombination cassette was inserted in a pZmUBI-tNos vector at MCS location using BamH1 restriction sites to form a pZmUBI-LR-tNos vector. The complete *TaSPO11-1* CDS fragment was inserted in this vector under the control of the UBI promoter using GATEWAY cloning sites. Rice seed embryo-derived callus from line segregating the *Osspo11-1-1* mutation were then transformed accordingly to the method described in (Sallaud *et al.* 2003). Primary transformants expressing the *TaSPO11-1* transgene were selected on hygromycin selection medium. Mutation in the ATG of *OsSPO11-1* (Loc_Os03g54091) was ascertain by PCR (primers: SPO11-R1 ccaaaattcttgtgggtgct and SPO11-F2 cggaggagcagtagttctgg) and sequencing. Presence and integrity of the transgene was also verified by PCR (primers: pUBI-F cttgatatacttggatgatggc and tNOS-R cgcaagaccggcaacaggattc) and sequencing.

For Arabidopsis complementation, the complete *SPO11-1* CDS fragment was cloned into the GATEWAY destination vector pMDC32 (Curtis and Grossniklaus 2003) in which the *35S* promoter was replaced with the Arabidopsis *RAD51* promoter (1031bp upstream of the *RAD51* ATG; (Da Ines, *et al.* 2013) with a HindIII/AscI digest . The plasmid was then inserted in an *Agrobacterium tumefaciens* C58C1 strain which was subsequently used to transform *Atspo11-1-2* heterozygous

mutant plants by the floral dip method (Clough and Bent 1998). T1 seeds from the Agrobacterium-transformed plants were sown on soil and T1 transformants were selected for Hygromycin resistance on 0.5X MS/ 1% sucrose/ 0.8% agar plates containing 15µg/ml Hygromycin B Gold (InvivoGen). Presence of the transgene and genotypes of transformants were verified by PCR.

Arabidopsis male meiotic chromosome spreads

Chromosome spreads were prepared according to (Ross *et al.* 1996). Whole inflorescences were fixed in ice-cold ethanol/glacial acetic acid (3:1) for 3 x 30 min and stored at -20°C until further use. Immature flower buds were rinsed twice at room temperature in distilled water for 5 min. This was followed by two washes in citrate buffer for 5 min. Buds of appropriate size were selected under a binocular microscope and incubated for 75 to 90 minutes on a slide in 100µL of enzyme mixture (0.3% w/v cellulase (Sigma), 0.3% w/v pectolyase (Sigma) and 0.3% cytohelicase (Sigma)) in a moist chamber at 37°C. Each bud was then softened for 1 minute in 15µL of acetic acid (60%) on a microscope slide at 45°C, fixed with ice-cold ethanol/glacial acetic acid (3:1) and air-dried. Eventually, slides were mounted in Vectashield mounting medium with DAPI (1.5 $\mu g.mL^{-1}$; Vector Laboratories Inc.).

For chiasma counting, number of chiasmata at metaphase I was estimated based on bivalent configuration: rod-shaped bivalents were considered to contain a single chiasma and ring-shaped bivalents, two (one on each arm) (Sanchez Moran *et al.* 2001).

Immunolocalization of proteins in pollen mother cells (PMCs)

Spreads of PMCs for immunolocalization of RAD51 were performed as described previously (Armstrong *et al.* 2002). Primary antibodies used for immunostaining were: anti-ASY1 raised in guinea Pig (1:500) (Higgins *et al.* 2004) and anti-RAD51 raised in rat (1:500) (Kurzbauer *et al.* 2012).

Microscopy

All observations were made with a motorized Zeiss AxioImager.Z1 epifluorescence microscope (Zeiss) using a PL Apochromat 100X/1.40 oil objective. Photographs were taken with an AxioCam MRm camera (Zeiss) driven by ZEN Pro software (Zeiss). Captured images were further processed and adjusted for brightness and contrast on ZEN Pro and ImageJ/FIJI software.

Statistical analysis

All graphs and statistical analyses were performed using software GraphPad PRISM 6. To determine whether differences between two groups were statistically significant, groups were compared using ordinary one-way ANOVA and Holm-Sidak test to account for multiple comparisons. A *P*-value of 0.05 or less was considered to be statistically significant.

Acknowledgements

We thank Mathilde Grelon for providing *spo11-1*, *spo11-2 and mtopVIB* mutants, and Chris Franklin and Peter Schlögelhofer for ASY1 and RAD51 antibodies, respectively. The authors thank members of the GeCO group and members of the recombination group for helpful discussions. This work was supported by CNRS, INRA, INSERM, Université Clermont Auvergne, and the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP20-25). RM was funded by a PhD grant from the Association Nationale Recherche Technologie (ANRT CIFRE Grant No. 2014/1020).

Supplementary Materials Legends

Figure S1. Sequence alignment of Bread wheat SPO11-1 proteins. Alignment was generated using ClustalW. Numbers indicate amino acid positions. Red letters indicate single amino acid polymorphism and blue letters denote amino acid deleted in the B genome. Identical residues are highlighted in black and asterisks, colons and dots under the sequences, indicate identical, conserved and semi-conserved residues, respectively.

Figure S2. Sequence alignment of SPO11-1 proteins from bread wheat and ancestors. Alignment of SPO11-1 from genome A, D, and their ancestors *Triticum urartu* and *Aegilops tauschii* was generated using ClustalW. Numbers indicate amino acid positions. Amino acid highlighted in cyan and yellow designate single amino acid polymorphism.

Figure S3. Sequence alignment of bread wheat and Arabidopsis proteins. Alignment was generated using ClustalW. Numbers indicate amino acid positions. Red squares with roman numerals indicate the conserved motifs. Identical residues are highlighted in black and asterisks, colons and dots under the sequences, indicate identical, conserved and semi-conserved residues, respectively.

Figure S4. Sequence alignment of 107 plant SPO11-1 proteins. Alignment was generated using ClustalW 2.0. Numbers indicate amino acid positions. Asterisks, colons and dots under the sequences, indicate identical, conserved and semi-conserved residues, respectively.

Table S1. Seed number per silique in wild-type, *Atspo11-1* and *Atspo11-1_TaSPO11-1* primary transformants. 10 to 12 fruits were counted per plant. (n.d. : not determined). P-value were calculated using ordinary one-way ANOVA and Holm-Sidak test to account for multiple comparisons.

Table S2. Seed number per silique in wild-type and in the progeny of *Atspo11-1-2_TaSPO11-1* transformants. Seeds were counted in 4 plants per genotype and 8 fruits per plant. (n.d. : not determined). P-value were calculated using ordinary one-way ANOVA and Holm-Sidak test to account for multiple comparisons.

References

An, X.J., Deng, Z.Y. and Wang, T. (2011) OsSpo11-4, a rice homologue of the archaeal TopVIA protein, mediates double-strand DNA cleavage and interacts with OsTopVIB. *PLoS One*, **6**, e20327.

Armstrong, S.J., Caryl, A.P., Jones, G.H. and Franklin, F.C. (2002) Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in Arabidopsis and Brassica. *J Cell Sci*, **115**, 3645-3655.

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P.C., Nicolas, A. and Forterre, P. (1997) An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature*, **386**, 414-417.

Blake, N.K., Lehfeldt, B.R., Lavin, M. and Talbert, L.E. (1999) Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: the B genome of wheat. *Genome*, **42**, 351-360.

Boden, S.A., Langridge, P., Spangenberg, G. and Able, J.A. (2009) TaASY1 promotes homologous chromosome interactions and is affected by deletion of *Ph1*. *Plant J*, **57**, 487-497.

Boden, S.A., Shadiac, N., Tucker, E.J., Langridge, P. and Able, J.A. (2007) Expression and functional analysis of *TaASY1* during meiosis of bread wheat (*Triticum aestivum*). *BMC Mol Biol*, **8**, 65.

Camiolo, S., Melito, S. and Porceddu, A. (2015) New insights into the interplay between codon bias determinants in plants. *DNA Res*, **22**, 461-470.

Chen, C., Farmer, A.D., Langley, R.J., Mudge, J., Crow, J.A., May, G.D., Huntley, J., Smith, A.G. and Retzel, E.F. (2010) Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated Arabidopsis male meiocytes. *BMC Plant Biol*, **10**, 280.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J*, **16**, 735-743.

Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res*, **14**, 1188-1190.

Curtis, M.D. and Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol*, **133**, 462-469.

Da Ines, O., Degroote, F., Goubely, C., Amiard, S., Gallego, M.E. and White, C.I. (2013) Meiotic recombination in Arabidopsis is catalysed by DMC1, with RAD51 playing a supporting role. *PLoS Genet*, **9**, e1003787.

de Bustos, A., Perez, R. and Jouve, N. (2007) Characterization of the gene *Mre11* and evidence of silencing after polyploidization in *Triticum*. *Theor Appl Genet*, **114**, 985-999.

de Massy, B. (2013) Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes. *Annu Rev Genet*, **47**, 563-599.

Devisetty, U.K., Mayes, K. and Mayes, S. (2010) The *RAD51* and *DMC1* homoeologous genes of bread wheat: cloning, molecular characterization and expression analysis. *BMC Res Notes*, **3**, 245.

Diaz, R.L., Alcid, A.D., Berger, J.M. and Keeney, S. (2002) Identification of residues in yeast Spo11p critical for meiotic DNA double-strand break formation. *Mol Cell Biol*, **22**, 1106-1115.

Doyle, J.J., Flagel, L.E., Paterson, A.H., Rapp, R.A., Soltis, D.E., Soltis, P.S. and Wendel, J.F. (2008) Evolutionary genetics of genome merger and doubling in plants. *Annu Rev Genet*, **42**, 443-461.

Fayos, I., Mieulet, D., Petit, J., Meunier, A.C., Perin, C., Nicolas, A. and Guiderdoni, E. (2019) Engineering meiotic recombination pathways in rice. *Plant Biotechnol J*.

Fu, M., Wang, C., Xue, F., Higgins, J., Chen, M., Zhang, D. and Liang, W. (2016) The DNA Topoisomerase VI-B Subunit OsMTOPVIB Is Essential for Meiotic Recombination Initiation in Rice. *Mol Plant*, **9**, 1539-1541.

Gardiner, L.J., Wingen, L.U., Bailey, P., Joynson, R., Brabbs, T., Wright, J., Higgins, J.D., Hall, N., Griffiths, S., Clavijo, B.J. and Hall, A. (2019) Analysis of the recombination landscape

of hexaploid bread wheat reveals genes controlling recombination and gene conversion frequency. *Genome Biol*, **20**, 69.

Giles, R.J. and Brown, T.A. (2006) GluDy allele variations in *Aegilops tauschii* and *Triticum aestivum*: implications for the origins of hexaploid wheats. *Theor Appl Genet*, **112**, 1563-1572.

Grelon, M., Vezon, D., Gendrot, G. and Pelletier, G. (2001) *AtSPO11-1* is necessary for efficient meiotic recombination in plants. *EMBO J*, **20**, 589-600.

Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I. and Moore, G. (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. *Nature*, **439**, 749-752.

Hartung, F., Angelis, K.J., Meister, A., Schubert, I., Melzer, M. and Puchta, H. (2002) An archaebacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation of plants. *Curr Biol*, **12**, 1787-1791.

Hartung, F. and Puchta, H. (2000) Molecular characterisation of two paralogous SPO11 homologues in *Arabidopsis thaliana*. *Nucleic Acids Res*, **28**, 1548-1554.

Hartung, F. and Puchta, H. (2001) Molecular characterization of homologues of both subunits A (SPO11) and B of the archaebacterial topoisomerase 6 in plants. *Gene*, **271**, 81-86.

Hartung, F., Wurz-Wildersinn, R., Fuchs, J., Schubert, I., Suer, S. and Puchta, H. (2007) The catalytically active tyrosine residues of both SPO11-1 and SPO11-2 are required for meiotic double-strand break induction in Arabidopsis. *Plant Cell*, **19**, 3090-3099.

Higgins, J.D., Armstrong, S.J., Franklin, F.C. and Jones, G.H. (2004) The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. *Genes Dev*, **18**, 2557-2570.

Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R. and Gornicki, P. (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc Natl Acad Sci U S A*, **99**, 8133-8138.

Hunter, N. (2015) Meiotic Recombination: The Essence of Heredity. *Cold Spring Harb Perspect Biol*, **7**.

IWGSC (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, **361**.

Jain, M., Tyagi, A.K. and Khurana, J.P. (2006) Overexpression of putative topoisomerase 6 genes from rice confers stress tolerance in transgenic Arabidopsis plants. *FEBS J*, **273**, 5245-5260.

Keeney, S., Giroux, C.N. and Kleckner, N. (1997) Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, **88**, 375-384.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, **10**, 845-858.

Khoo, K.H., Able, A.J. and Able, J.A. (2012a) The isolation and characterisation of the wheat molecular ZIPper I homologue, *Ta*ZYP1. *BMC Res Notes*, **5**, 106.

Khoo, K.H., Able, A.J. and Able, J.A. (2012b) Poor Homologous Synapsis 1 Interacts with Chromatin but Does Not Colocalise with ASYnapsis 1 during Early Meiosis in Bread Wheat. *Int J Plant Genomics*, **2012**, 514398.

Kurzbauer, M.T., Uanschou, C., Chen, D. and Schlogelhofer, P. (2012) The recombinases DMC1 and RAD51 are functionally and spatially separated during meiosis in Arabidopsis. *Plant Cell*, **24**, 2058-2070.

Lam, I. and Keeney, S. (2014) Mechanism and regulation of meiotic recombination initiation. *Cold Spring Harb Perspect Biol*, **7**, a016634.

Leroy, P., Guilhot, N., Sakai, H., Bernard, A., Choulet, F., Theil, S., Reboux, S., Amano, N., Flutre, T., Pelegrin, C., Ohyanagi, H., Seidel, M., Giacomoni, F., Reichstadt, M., Alaux, M., Gicquello, E., Legeai, F., Cerutti, L., Numa, H., Tanaka, T., Mayer, K., Itoh, T., Quesneville, H. and Feuillet, C. (2012) TriAnnot: A Versatile and High Performance Pipeline for the Automated Annotation of Plant Genomes. *Front Plant Sci*, **3**, 5.

Lloyd, A.H., Ranoux, M., Vautrin, S., Glover, N., Fourment, J., Charif, D., Choulet, F., Lassalle, G., Marande, W., Tran, J., Granier, F., Pingault, L., Remay, A., Marquis, C., Belcram,

H., Chalhoub, B., Feuillet, C., Berges, H., Sourdille, P. and Jenczewski, E. (2014) Meiotic gene evolution: can you teach a new dog new tricks? *Mol Biol Evol*, **31**, 1724-1727.

Ma, G., Zhang, W., Liu, L., Chao, W.S., Gu, Y.Q., Qi, L., Xu, S.S. and Cai, X. (2018) Cloning and characterization of the homoeologous genes for the Rec8-like meiotic cohesin in polyploid wheat. *BMC Plant Biol*, **18**, 224.

Malik, S.B., Ramesh, M.A., Hulstrand, A.M. and Logsdon, J.M., Jr. (2007) Protist homologs of the meiotic *Spo11* gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. *Mol Biol Evol*, **24**, 2827-2841.

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., International Wheat Genome Sequencing, C., Jakobsen, K.S., Wulff, B.B., Steuernagel, B., Mayer, K.F. and Olsen, O.A. (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, **345**, 1250092.

Martin, A.C., Borrill, P., Higgins, J., Alabdullah, A., Ramirez-Gonzalez, R.H., Swarbreck, D., Uauy, C., Shaw, P. and Moore, G. (2018) Genome-Wide Transcription During Early Wheat Meiosis Is Independent of Synapsis, Ploidy Level, and the *Ph1* Locus. *Front Plant Sci*, **9**, 1791.

Martin, A.C., Rey, M.D., Shaw, P. and Moore, G. (2017) Dual effect of the wheat *Ph1* locus on chromosome synapsis and crossover. *Chromosoma*, **126**, 669-680.

Mercier, R., Mezard, C., Jenczewski, E., Macaisne, N. and Grelon, M. (2015) The molecular biology of meiosis in plants. *Annu Rev Plant Biol*, **66**, 297-327.

Moore, G. (2014) The control of recombination in wheat by *Ph1* and its use in breeding. *Methods Mol Biol*, **1145**, 143-153.

Perez, R., Cuadrado, A., Chen, I.P., Puchta, H., Jouve, N. and De Bustos, A. (2011) The *Rad50* genes of diploid and polyploid wheat species. Analysis of homologue and homoeologue expression and interactions with *Mre11*. *Theor Appl Genet*, **122**, 251-262.

Plotkin, J.B. and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*, **12**, 32-42.

Ramirez-Gonzalez, R.H., et al. (2018) The transcriptional landscape of polyploid wheat. *Science*, **361**.

Robert, T., Nore, A., Brun, C., Maffre, C., Crimi, B., Bourbon, H.M. and de Massy, B. (2016a) The TopoVIB-Like protein family is required for meiotic DNA double-strand break formation. *Science*, **351**, 943-949.

Robert, T., Vrielynck, N., Mezard, C., de Massy, B. and Grelon, M. (2016b) A new light on the meiotic DSB catalytic complex. *Semin Cell Dev Biol*, **54**, 165-176.

Ross, K.J., Fransz, P. and Jones, G.H. (1996) A light microscopic atlas of meiosis in *Arabidopsis thaliana*. *Chromosome Res*, **4**, 507-516.

Sallaud, C., Meynard, D., van Boxtel, J., Gay, C., Bes, M., Brizard, J.P., Larmande, P., Ortega, D., Raynal, M., Portefaix, M., Ouwerkerk, P.B., Rueb, S., Delseny, M. and Guiderdoni, E. (2003) Highly efficient production and characterization of T-DNA plants for rice (Oryza sativa L.) functional genomics. *Theor Appl Genet*, **106**, 1396-1408.

Sanchez Moran, E., Armstrong, S.J., Santos, J.L., Franklin, F.C. and Jones, G.H. (2001) Chiasma formation in Arabidopsis thaliana accession Wassileskija and in two meiotic mutants. *Chromosome Res*, **9**, 121-128.

Shingu, Y., Mikawa, T., Onuma, M., Hirayama, T. and Shibata, T. (2010) A DNA-binding surface of SPO11-1, an Arabidopsis SPO11 orthologue required for normal meiosis. *FEBS J*, **277**, 2360-2374.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Depamphilis, C.W., Wall, P.K. and Soltis, P.S. (2009) Polyploidy and angiosperm diversification. *Am J Bot*, **96**, 336-348.

Sprink, T. and Hartung, F. (2014) The splicing fate of plant *SPO11* genes. *Front Plant Sci*, **5**, 214.

Stacey, N.J., Kuromori, T., Azumi, Y., Roberts, G., Breuer, C., Wada, T., Maxwell, A., Roberts, K. and Sugimoto-Shirasu, K. (2006) Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. *Plant J*, **48**, 206-216.

Sugimoto-Shirasu, K., Stacey, N.J., Corsar, J., Roberts, K. and McCann, M.C. (2002) DNA topoisomerase VI is essential for endoreduplication in Arabidopsis. *Curr Biol*, **12**, 1782-1786.

Van de Peer, Y., Fawcett, J.A., Proost, S., Sterck, L. and Vandepoele, K. (2009) The flowering world: a tale of duplications. *Trends Plant Sci*, **14**, 680-688.

Vrielynck, N., Chambon, A., Vezon, D., Pereira, L., Chelysheva, L., De Muyt, A., Mezard, C., Mayer, C. and Grelon, M. (2016) A DNA topoisomerase VI-like complex initiates meiotic recombination. *Science*, **351**, 939-943.

Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J.D. and Feng, X. (2018) Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. *Nat Genet*, **50**, 130-137.

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B. and Rieseberg, L.H. (2009) The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U S A*, **106**, 13875-13879.

Xue, M., Wang, J., Jiang, L., Wang, M., Wolfe, S., Pawlowski, W.P., Wang, Y. and He, Y. (2018) The Number of Meiotic Double-Strand Breaks Influences Crossover Distribution in Arabidopsis. *Plant Cell*, **30**, 2628-2638.

Xue, Z., Li, Y., Zhang, L., Shi, W., Zhang, C., Feng, M., Zhang, F., Tang, D., Yu, H., Gu, M. and Cheng, Z. (2016) OsMTOPVIB Promotes Meiotic DNA Double-Strand Break Formation in Rice. *Mol Plant*, **9**, 1535-1538.

Yang, H., Lu, P., Wang, Y. and Ma, H. (2011) The transcriptome landscape of Arabidopsis male meiocytes from high-throughput sequencing: the complexity and evolution of the meiotic process. *Plant J*, **65**, 503-516.

Yin, Y., Cheong, H., Friedrichsen, D., Zhao, Y., Hu, J., Mora-Garcia, S. and Chory, J. (2002) A crucial role for the putative Arabidopsis topoisomerase VI in plant growth and development. *Proc Natl Acad Sci U S A*, **99**, 10191-10196.

Yu, H., Wang, M., Tang, D., Wang, K., Chen, F., Gong, Z., Gu, M. and Cheng, Z. (2010) OsSPO11-1 is essential for both homologous chromosome pairing and crossover formation in rice. *Chromosoma*, **119**, 625-636.

Table

Table 1: Wheat, its ancestors and *Arabidopsis thaliana SPO11-1* gene characteristics. gDNA: genomic DNA; CDS: coding sequence; bp: base pair; aa: amino acid.

Gene	Species	Gene ID	gDNA (bp)	CDS (bp)	Nb of exons	Protein ID			protein (aa)	Annotation status
						NCBI	Genbank	Uniprot		
SPO11-1-5A	*T. aestivum*	TraesCS5A02G391400	3937	1164	15	-	-	A0A3B6KM14	387	newly annotated
SPO11-1-5B	*T. aestivum*	TraesCS5B02G396300	3928	1161	15	-	-	A0A3B6LTD1	386	newly annotated
SPO11-1-5D	*T. aestivum*	TraesCS5D02G401100	3970	1164	15	-	-	A0A3B6MY04	387	newly annotated
SPO11-1	*T. urartu*	TRIUR3_12346	3946	1161	15	-	EMS54133.1	M7YTX6	390	reannotated
SPO11-1	*Aegilops tauschii*	LOC109743941	3970	1164	15	XP_020158624.1	-	-	387	reannotated
AtSPO11-1	*A. thaliana*	At3G13170	2812	1089	15	NP_187923.1	-	Q9M4A2	362	unmodified

Figures Legends

Figure 1: Schematic representation of the putative SPO11 meiotic DSB catalytic complex.

This protein complex is suggested to be an heterotetramer composed of one heterodimer (SPO11-1 and SPO11-2; green and blue, respectively) and one homodimer (MTOPVIB; orange) (Robert, *et al.* 2016b).

Figure 2: **Protein sequence alignment of *Triticum aestivum* and *Arabidopsis thaliana* SPO11-1.**

Only the seven conserved motifs are shown. Conserved amino acids are highlighted in green and similar amino acids are highlighted in yellow. Catalytically active tyrosine and conserved glycine and arginine involved in DNA-binding activity are in red.

Figure 3: **3D modelling of TaSPO11-1-5D protein.**

A 3D structural model of SPO11-1 obtained by sequence homology modelling with PHYRE2 online pipeline. This SPO11-1 model consists of amino acid residues 1 to 387. 336 residues over 387 (87%) were modelled with >90% accuracy. A and B are mirror views of the structural model rendered with PyMol software. The seven conserved motifs are shown in green, the Toprim domain in light pink and the 5Y-CAP domain in light yellow. Catalytic tyrosine is depicted in magenta and the glycine and arginine essential for DNA-binding activity are depicted in red and blue, respectively.

Figure 4: **Sequence and conservation level of the seven conserved motifs in SPO11-1 proteins from 105 species.** (**A)** Consensus sequence of the seven conserved motifs extracted from 107 SPO11-1 sequences. **(B)** Conservation level of the seven motifs within 107 SPO11-1 sequences. White boxes: lower and upper quartile of conservation rate for each motif. Black lines: mean conservation rate for each motif. Dashed lines: maximum and minimum conservation rate values. Circles: outliers.

Figure 5: **Phylogenetic tree of plant SPO11-1 proteins from 105 species.**

Dicotyledons are shown in red, monocotyledons in blue and *Amborella trichopoda*, sister of the angiosperms, is shown in purple.

Figure 6: **Wheat *TaSPO11-1-5D* restores fertility of the Rice *spo11-1* mutant.**

(A) Pictures of rice *spo11-1* mutant plants and panicles expressing or not the wheat *TaSPO11-1-5D* transgene. Rice *spo11-1* plants are sterile and develop panicles with empty spikelets (right panel). In contrast, *spo11-1* mutant plants expressing wheat *TaSPO11-1-5D* are fertile and develop panicles with filled spikelets (left panel). **(B)** Percentage of filled spikelet per panicle in wild-type plants and progeny of two *spo11-1* transformants expressing wheat *TaSPO11-1-5D* (T2 and T64). In the box and whiskers, each dot represents the percentage of filled spikelet per panicle in one plant (n = 6 for wild-type plants, n = 5 for *spo11-1*, and n = 5 for *spo11-1 TaSPO11-1-5D* plants). Means are represented by a + and horizontal bars denote medians.

Figure 7: **Wheat *TaSPO11-1-5D* restores fertility of the Arabidopsis *spo11-1* mutant**

(A) Schematic representation of the *pRAD51:TaSPO11-1* construct. **(B)** Wild-type plants have long siliques full of seeds, while *Atspo11-1* mutants are sterile and exhibit very short siliques. Expression of the *TaSPO11-1* in *Atspo11-1* mutants restores fertility. **(C)** number of seeds per silique in Wild-type, *Atspo11-1,* and 15 *Atspo11-1* + *pRAD51:TaSPO11-1* independent primary transformants. Each dot represents the number of seeds in one silique. **(D)** number of seeds per silique in Wild-type, and T2 progeny of *4 Atspo11-1* + *pRAD51:TaSPO11-1* independent primary transformants (T3, T12, T13 and T36 as indicated under graph). Each dot represents the number of seeds in one silique. Blue dots show number of seeds per silique in wild-type, red dots *Atspo11-1* mutants and black dots represent *Atspo11-1* mutants expressing *TaSPO11-1*.

Figure 8: **Meiotic progression in wild-type, *Atspo11-1* mutants and *Atspo11-1* mutants complemented with wheat *TaSPO11-1*.**

DAPI staining of chromosomes during meiosis in Arabidopsis **(A-E)** wild-type, **(F-J)** *spo11-1* and **(K-O)** *spo11-1* + TaSPO11-1 plants. **(A, F, K)** Early prophase I, **(B, G, L)** Late prophase I, **(C, H, M)**

Metaphase I, **(D, I, N)** Metaphase II, and **(E, J, O)** Tetrad. In wild-type, cells show pairing and synapsis of homologous chromosomes at late prophase I **(B)**, five bivalents at metaphase I **(C)**, two groups of five chromosomes at Metaphase II **(D)** and balanced tetrads **(E)**. *spo11-1* mutants exhibit defective synapsis **(G)**, univalent in Metaphase I **(H)** and unbalance Metaphase II **(I)** and polyads **(J)**. In *spo11-1* expressing wheat TaSPO11-1 **(K-O)**, wild-type meiotic figures can be observed. (Scale Bar: 10 µm).

Figure 9: **Expression of wheat *TaSPO11-1* in *Atspo11-1* mutant promotes bivalent formation**

(A) Representative images of Metaphase I are shown (Scale Bar: 10 µm).

(B) Mean number of bivalents (dark grey) and pairs of univalent (grey) per meiosis

(C) Bivalents per cell (in percentage). *Atspo11-1* plants expressing *TaSPO11-1* show a significant increase of bivalent formation when compared to *Atspo11-1* mutants. Number of cells analyzed is indicated in parentheses.

Figure 10: Reduced numbers of RAD51 foci in *Atspo11-1* mutants complemented with wheat *TaSPO11-1*.

(A) Immunolocalization of RAD51 (green) and the chromosome axis protein ASY1 (red) on leptotene/zygotene meiotic chromosome spreads. (Scale Bars: 5 µm). **(B)** Quantification of RAD51 foci per positive cell throughout prophase I in both wild-type and *Atspo11-1* mutants expressing *TaSPO11-1* (T36). (p-value < 0.0001, Mann-Whitney test).

Figure 11: **Complementation of *Atspo11-1* by *TaSPO11-1* requires presence of *AtSPO11-2* and *AtMTOPVIB*.**

(A) Pictures of siliques from wild-type, and mutant plants expressing or not *TaSPO11-1*. Genotype of the mutants are indicated above pictures. Fertile plants have long siliques while sterile plants have short siliques. Fertility of *Atspo11-1 + TaSPO11-1* is lost by deletion of either *AtSPO11-2* or *AtMTOPVIB*. **(B)**: Mean number of seeds per silique. Each dot represents the number of seeds in one silique.

819