OXFORD GENETICS

APOK3, a pollen killer antidote in *Arabidopsis thaliana*

Matthieu Simon,[1] Stéphanie Durand,[1,4] Anthony Ricou,[1] Nathalie Vrielynck,[1] Baptiste Mayjonade [ID] ,[2] Jérôme Gouzy [ID] ,[2] Roxane Boyer,[3,5] Fabrice Roux [ID] ,[2] Christine Camilleri [ID] ,[1] Françoise Budar [ID] [1,*]

[1]Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), 78000 Versailles, France,
[2]LIPME, Université de Toulouse, INRAE, CNRS, 31326 Castanet-Tolosan, France,
[3]INRAE, GeT-PlaGe, Genotoul, 31326 Castanet-Tolosan, France,
[4]Present address: Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany,
[5]Present address: Institute of Biology Leiden (IBL), 2333 BE Leiden, The Netherlands

*Corresponding author: IJPB, INRAE Centre IdF de Versailles-Grignon, Route de St-Cyr (RD10), 78026 Versailles Cedex, France. Email: francoise.budar@inrae.fr

Abstract

The principles of heredity state that the two alleles carried by a heterozygote are equally transmitted to the progeny. However, genomic regions that escape this rule have been reported in many organisms. It is notably the case of genetic loci referred to as gamete killers, where one allele enhances its transmission by causing the death of the gametes that do not carry it. Gamete killers are of great interest, particularly to understand mechanisms of evolution and speciation. Although being common in plants, only a few, all in rice, have so far been deciphered to the causal genes. Here, we studied a pollen killer found in hybrids between two accessions of *Arabidopsis thaliana*. Exploring natural variation, we observed this pollen killer in many crosses within the species. Genetic analyses revealed that three genetically linked elements are necessary for pollen killer activity. Using mutants, we showed that this pollen killer works according to a poison-antidote model, where the poison kills pollen grains not producing the antidote. We identified the gene encoding the antidote, a chimeric protein addressed to mitochondria. *De novo* genomic sequencing in 12 natural variants with different behaviors regarding the pollen killer revealed a hyper variable locus, with important structural variations particularly in killer genotypes, where the antidote gene recently underwent duplications. Our results strongly suggest that the gene has newly evolved within *A. thaliana*. Finally, we identified in the protein sequence polymorphisms related to its antidote activity.

Keywords: transmission ratio distortion; pollen killer; poison-antidote; structural variation; *Arabidopsis thaliana*

Introduction

First considered as genetic curiosities, loci that do not comply with Mendel's laws are now recognized to be common in fungi, plants, and animals, with a particularly high incidence in plants (Fishman and McIntosh 2019). These loci with transmission ratio distortion are of major interest to understand genomic evolution, adaptation, and speciation (Lindholm *et al.* 2016; Agren and Clark 2018; Fishman and Sweigart 2018; Presgraves 2010). In addition, they have significant impacts on plant breeding, disturbing QTL mapping, and hampering the use of genetic resources. Cases of transmission ratio distortion have been reported in multiple plant species, with most studies conducted in rice and *Arabidopsis thaliana*. In rice, such segregation biases are very common in interspecific and subspecific crosses, and often linked to hybrid sterility (Ouyang and Zhang 2018), limiting intersubspecific crosses suitable for breeding (Matsubara *et al.* 2011; Zhang *et al.* 2020). For instance, 18 genomic regions were recently found to be distorted in a cross between *Oryza sativa* ssp. *japonica* and *O. sativa* ssp. *indica* (Zhang *et al.* 2020). In the genus *Arabidopsis*, many examples of inter or intraspecific hybrid incompatibilities were also reported (reviewed in Vaid and Laitinen 2019). Within the species *A. thaliana*, transmission ratio distortion was observed in over half of a set of 17 F2 populations (Salomé *et al.* 2012). More recently, at

least one transmission ratio distortion was detected in ~25% of a set of 500 F2 populations, corresponding to more than 100 distorted genomic regions (Seymour *et al.* 2019).

Transmission ratio distortion can result from Bateson–Dobzhansky–Muller incompatibilities, where independently evolved alleles result in deleterious or suboptimal phenotypes when brought together, or from allele-specific gamete elimination, where an allele takes over the alternative allele in the gametes produced by a heterozygote (reviewed in Ouyang and Zhang 2013). Allele-specific gamete elimination can occur at the meiotic stage, thus designated as meiotic driver. The B chromosomes of cereals (Östergren 1945; Houben 2017) and the driving centromere in yellow monkeyflower (Finseth *et al.* 2015, 2021), for example, were identified as such meiotic drivers. Alternatively, gamete elimination occurs after meiosis when the favored (killer) allele induces a defect in the gametes that carry a sensitive allele, eventually causing its underrepresentation in the next generation. Such a situation was observed for the *wtf* genes in fission yeast (Nuckolls *et al.* 2017), the *Spok* genes in *Podospora anserina* (Grognet *et al.* 2014), the *SD* system in *Drosophila melanogaster* (Larracuente and Presgraves 2012), and the *Sa* (Long *et al.* 2008), *S1* (Xie *et al.* 2019), *S5* (Yang *et al.* 2012), and *qhMS7* (Yu *et al.* 2018) loci in rice. In plants, loci causing segregation distortion by allele-specific

gamete elimination are designated as gamete killers, or more specifically pollen killers when they affect the male gametes.

Meiotic drivers and gamete killers have retained particular attention for their potential in triggering genomic conflicts (Lindholm *et al.* 2016; Agren and Clark 2018). One intriguing feature of gamete killers is that killer alleles cause a defect only in the gametes that do not carry them. This can be explained by either of 2 main genetic models (Bravo Nunez *et al.* 2018), schematically represented in Fig. 1. In both models, all the genes responsible for gamete elimination are tightly linked and the killing factor is produced before (or during) meiosis, and still present in all the developing gametes. In the "killer-target" model, a partner (the target), necessary for the killing activity, is encoded by the sensitive allele and defects occur after meiosis, only in gametes that carry this allele. For instance, this model applies to the *SD* system of *D. melanogaster* (Larracuente and Presgraves 2012) and the *Sa* locus of rice (Long *et al.* 2008). In the "poison-antidote" model, inspired by poison-antidote bacterial systems (Burga *et al.* 2020), the killer allele produces an antidote that counteracts the poison. However, whereas the poison produced before meiosis subsists in the gametes, the antidote does not. Therefore, only the gametes that are able to produce the antidote are rescued. Poison-antidote type segregation distorters have been described in fungi and plants and include the fission yeast *wtf* genes (Nuckolls *et al.* 2017) and the rice pollen killer *qHMS7* (Yu *et al.* 2018).

In the unraveled cases of transmission ratio distortion in *A. thaliana*, the identification of causal genes showed that the distortions resulted from Bateson–Dobzhansky–Muller incompatibilities between parental alleles (or epialleles), most often located at physically unlinked loci (Bomblies *et al.* 2007; Bikard *et al.* 2009; Durand *et al.* 2012; Agorio *et al.* 2017; Jiao *et al.* 2021), or at 1 locus (Smith *et al.* 2011). However, in contrast to rice for which a few gamete killers have been deciphered to the causal genes (Long *et al.* 2008; Yang *et al.* 2012; Koide *et al.* 2018; Yu *et al.* 2018), pollen killers have been reported only once in *A. thaliana* (Simon *et al.* 2016). In this initial study, we showed that several pollen killers contributed to a male sterility observed in hybrids between 2 *A. thaliana* natural genotypes (also called accessions), Shahdara (Sha) and Mr-0.

Here, we deciphered one of these pollen killers, located at the bottom of chromosome 3 at a locus referred to as L3, which induces in the hybrid's progeny a deficit of plants homozygous Sha at this locus. We showed that this pollen killer belongs to the poison-antidote class and we identified the gene coding the antidote. This gene, of chimeric origin, encodes a mitochondrial protein. The antidote gene originated and has evolved in the species *A. thaliana* within a highly variable locus. We precisely characterized the locus structure, showing that it contains at least 2 elements, flanking the antidote gene, required for its killer activity. By exploring the diversity within *A. thaliana*, this pollen killer was found in a number of hybrids. We showed that the loci of the killer genotypes strongly differ from those of non killer ones by important structural variations, including duplications of the antidote gene. Finally, we identified in the antidote protein some polymorphisms associated with its protective effect.

Materials and methods
Plant material and growth conditions
Natural accessions of *A. thaliana* were provided by the Versailles Arabidopsis Stock Center (http://publiclines.versailles.inrae.fr/) and T-DNA lines were provided by the Nottingham Arabidopsis Stock Center (http://arabidopsis.info/). Crosses are written in the conventional order female parent × male parent. In the analysis of the Sha/Mr-0 segregation bias, all the genotypes used have the
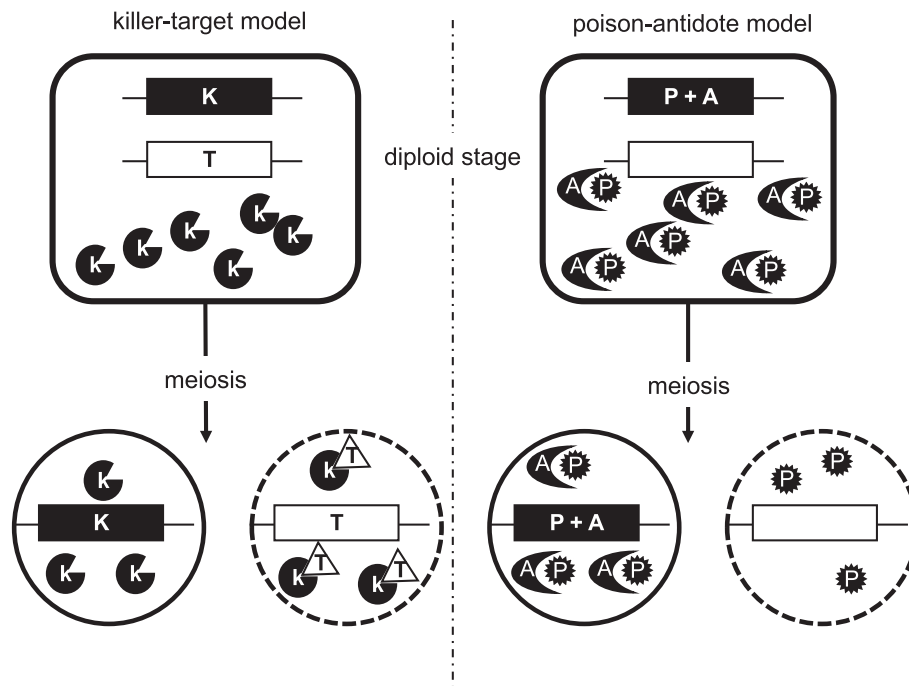
Fig. 1. Two genetic models of gamete killers. The locus is represented as a horizontal line with the killer and sensitive alleles as black and white boxes, respectively. For these general models, there is no hypothesis on the number of genes present in each box. In both models, the black allele is expressed before meiosis, and the killer (K) or poison (P) persists in all the gametes. In the killer-target model, only gametes with the white allele express the target (T), which interacts with K to trigger gamete death (represented as dashed outline). In the poison-antidote model, the black allele also produces a short-lived antidote (A) that counteracts P, so only the gametes that inherit the black allele are efficiently protected.

Mr-0 cytoplasm, unless specified. Because Mr-0 flowers more than 3 months after sowing, we constructed the Mr*fri* genotype, an early-flowering version of Mr-0, by inactivating the gene *FRIGIDA* (Supplementary File 1).

We use the following nomenclature: a genotype with the N nuclear background excepted at the locus L3 is designated NL3^M, NL3^S, or NL3^H where M, S, and H stand for homozygous Mr-0, homozygous Sha and heterozygous, respectively. The L3 locus at the bottom of chromosome 3, where a transmission ratio distortion due to a pollen killer was previously observed in ShaL3^H, (Simon *et al.* 2016) is delimited by the markers M1 and M15, respectively, at 23,007,336 and 23,289,927 bp on the Col-0 reference genomic sequence (Supplementary Table 1). To determine whether the segregation distortion depends on the nuclear background, we constructed the MrL3^H genotype by backcrossing the (Mr-0 × Sha) F1 by Mr-0 3 times and selecting a plant heterozygous at L3.

Plants were grown in the greenhouse under long-day conditions (16 h day, 8 h night) with additional artificial light (105 μE/m^2/s) when necessary.

Fine-mapping the pollen killer locus

The pollen killer was fine-mapped in the selfed progeny of a ShaL3^H plant. To reduce the interval responsible for the segregation distortion, we genotyped 4,717 plants and identified 42 recombinants within the L3 locus. Transmission ratio distortion was tested in 23 informative recombinants by genotyping their progenies. DNA extractions were conducted on leaves from seedlings as described by Loudet *et al.* (2002). Molecular markers are described in Supplementary Table 1. Cleaved amplified polymorphic sequences (CAPS) and single nucleotide polymorphism (SNP) markers were developed after Sanger-sequencing the locus in Sha and Mr-0 (see below). For CAPS markers M3 and M12, we used *Cac*8I (NEB) and *Bsp*1407I (ThermoFisher) restriction enzymes, respectively. SNP markers were genotyped by Sanger sequencing.

Cytological observations of pollen development

To determine the timing of pollen lethality, we observed several stages of male gamete development from male meiosis to mature pollen formation.

Analysis of male meiosis

DAPI (4',6-diamidino-2-phenylindole) staining of spread male meiotic chromosomes was performed according to Ross *et al.* (1996). All stages were observed in 3 independent ShaL3^H plants with both ShaL3^M and ShaL3^S as controls. Observations were made using a Zeiss Axio Imager2 microscope and photographs were taken using an AxioCam MRm (Zeiss) camera.

Pollen development

Propidium iodide (PI) staining of pollen from floral buds at different stages were performed as described by Durand *et al.* (2021). Images of tri-cellular pollen correspond to the accumulation of all slices from 1 stack with the 3D projection option (default parameters) of ImageJ 1.53k (http://imagej.nih.gov/ij). Alexander staining (Alexander 1969) was used to observe mature pollen as described by Durand *et al.* (2021). This technique stains pollen walls in blue–green and cytoplasm of viable pollen grains in red. Therefore, viable pollen grains appear filled in red while dead pollen grains appear blue–green. This technique was also used to check pollen viability in F1s to assess the activity of the pollen killer.

Sanger sequencing and annotation of Sha and Mr-0 alleles at the pollen killer locus

A PCR-walking and Sanger-sequencing strategy was used to obtain the DNA sequences of Sha and Mr-0 alleles at the locus. Amplification products were sequenced by Beckman Coulter Genomics (http://www.beckmangenomics.com). Sequences were processed and aligned with Codon Code Aligner V5.0.2 (http://www.codoncode.com/aligner/). For Mr-0, because the locus was very different from that of Col-0, a fosmid library was constructed using the Copycontrol fosmid library production kit (Illumina technologies). The library contained 7,600 cfu/ml with an average insert size of 35 kb. PCR screening of the library using 3 pairs of primers distributed across the locus (Supplementary Table 2) identified several clones covering the entire region in Mr-0, which were afterward sequenced. Purification of fosmid DNA was performed using the FosmidMax DNA purification kit (Illumina technologies).

Transfer of structural and functional annotations from the Col-0 reference genome to Sha and Mr-0 was performed with the EGN-EP transfer pipeline (Sallet *et al.* 2019). These annotations were manually corrected, especially in Mr-0, to account for copy number variations. TE annotation was performed with the TEannot tool (available at http://urgi.versailles.inra.fr/; Quesneville *et al.* 2005) and occurrences larger than 1 kb were retained.

De novo sequencing, assembly, and annotation of the genomes of 10 accessions

Using both Nanopore and Illumina whole-genome shotgun sequencing, we *de novo* sequenced and assembled the entire genomes of 10 accessions with different behaviors regarding the pollen killer (4 killer, 3 sensitive, and 3 resistant accessions). High molecular weight DNA was extracted from plantlets using a protocol modified from Mayjonade *et al.* (2016), described on Protocols.io (Russo *et al.* 2021). The following library preparation and sequencing were performed at the GeT-PlaGe core facility, INRAE Toulouse.

Nanopore sequencing

ONT libraries were prepared using the EXP-NBD103 and SQK-LSK109 kits according to the manufacturer's instructions and using 4 μg of 40 kb sheared DNA (Megaruptor, Diagenode) as input. Pools of 6 samples were sequenced on one R9.4.1 flowcell. Between 0.014 and 0.020 pmol of library were loaded on each flowcell and sequenced on a PromethION instrument for 72 h.

Illumina sequencing

Illumina libraries were prepared using the Illumina TruSeq Nano DNA HT Library Prep Kit. Libraries were sequenced with 2 × 150 bp paired-end reads on an Hiseq3000 instrument (Illumina).

Assembly

Nanopore sequence datasets ranging from 4 to 14 Gb with a minimum N50 of 22 kb were assembled with CANU (Koren *et al.* 2017) software release 1.9 (genomeSize = 125M parameter). Nuclear contigs were scaffolded with AllMaps (Tang *et al.* 2015) using Col-0 genome as reference. Then, 2 rounds of consensus polishing using Illumina paired-end data were performed. For each round, the paired-end reads were first mapped with bwa (Li and Durbin 2009) (0.7.17-r1188 debian) with a minimal score of 50 (-T). Then, pilon software (Walker *et al.* 2014) (version 1.23) was used to

generate the polished consensus sequences (–fix snps, indels –flank 20 –min-depth 20). Unanchored contigs shorter than 40 kb were removed (considered as likely contaminants as very few Illumina reads were mapped on these short contigs). Sequence reads and assembly metrics are given in Supplementary Table 3. In accessions with several copies of the antidote, occasional sequence errors subsist in some copies in the automatic assemblies. To correct these errors, we Sanger-sequenced the antidote gene after amplification with the primers AT3G62530F1 and AT3G62530R1 (Supplementary Table 2).

Annotation

The genomes were annotated using the EuGène software version 4.2a via the integrative pipeline egnep (Sallet *et al.* 2019) version 1.6 (http://eugene.toulouse.inra.fr/Downloads/egnep-Linux-x86_64.1.6.tar.gz). In addition to standard tools for repeat masking and lncRNA prediction included in egnep, the most recent and comprehensive annotation of *A. thaliana* available was used both as training dataset and as source of evidences. The peptide database of Araport (Krishnakumar *et al.* 2015) version 11 (201606) was used for the similarity searches performed with NCBI-BLASTX software version 2.2.31+ (hits spanning more than 80% of the protein length were retained). The corresponding cDNA database (201606) was used as transcriptional evidences. The cDNAs were mapped with gmap (Wu and Watanabe 2005) version 2017-09-05 (hits spanning more than 30% of the transcript length at a minimum identity percentage of 94 were retained).

Sequence analyses

Structural comparison of the pollen killer locus among accessions

In order to reveal major structural variations at the locus, each of the *de novo* assembled sequences of 12 accessions was aligned to Col-0 with nucmer (MUMmer version 4.0.0 beta2, default settings, on http://bioinfo.genotoul.fr/). Alignments were visualized with mummerplot (options –nocolor –png). The coordinates of the sequences used for the alignments are indicated in Supplementary Table 4.

Structural annotation of AT3G62530

After its identification as the antidote gene, we searched for features in the AT3G62530 amino acid sequence using the Simple Modular Architecture Research Tool (SMART; http://smart.embl-heidelberg.de/; Letunic *et al.* 2021). The TargetP 2.0 program (http://www.cbs.dtu.dk/services/TargetP; Almagro Armenteros *et al.* 2019) was used for the prediction of the cellular localization of the protein. Col-0 genes encoding proteins similar to AT3G62530 were retrieved through BLASTP search (2.9.0+ default settings; Altschul *et al.* 1997) against Araport11 protein sequences on the *Arabidopsis* Information Resource (https://www.arabidopsis.org/, last accessed 15th April 2021).

Phylogenetic analysis

DNA sequences of *AT3G62530* genes from 27 accessions (File S2) were obtained after DNA amplification with the primers AT3G62530F1 and AT3G62530R1 (Supplementary Table 2). An unrooted phylogenetic tree was generated using Phylogeny with default parameters (http://www.phylogeny.fr; Dereeper *et al.* 2008).

Gene expression in microspores

Microspore purification

Microspores of plants being either Sha, Mr-0 or Col-0 at the pollen killer locus were purified by flow cytometry on the cytometry platform facility at Gif-sur-Yvette, CNRS France (https://www.i2bc.paris-saclay.fr). Flower buds were chopped with a razor blade in 0.1M mannitol. The crude suspension was filtered through a 50-μm nylon mesh (Sysmex-Partec) and collected in polypropylene tubes at 4°C. Microspores were sorted using a MoFloAstrios_EQ cytometer (Beckman Coulter, Roissy, France) in PuraFlow sheath fluid (Beckman Coulter) at 25 psi (pounds per square inch), with a 100-μg nozzle. We performed sorting in purity mode, with ~43 kHz drop drive frequency, plates voltage of 4,000–4,500 V and an amplitude of 30–50 V. The 488-SSC (Side Scatter) parameter was set as threshold. Microspore gate was first set on FITC-A (526/50 band-pass for autofluorescence) *vs* SSC plot. Then, only low Forward Scatter (FSC) events were kept. The singlet gate was established using autofluorescence parameters. Accuracy of gating was determined postsorting using microscopy with transmitted light and 40× dry. The flow cytometer-sorted microspores (300k) were collected in 1.5-ml tubes containing 300 μl TRI Reagent (Sigma) and conserved at −80°C until subsequent RNA extraction.

RNA extractions and RT-PCR

Five hundred microliters of purified microspore suspension ($\sim 2.10^5$ microspores/ml) were mixed with 300 μl of glass beads (Sigma, G8772-100G acid washed glass beads) and 2 sterilized metal beads (3 mm diameter) on ice. The tubes were shaken 2 × 2.5 min at a frequency of 1/30 in a Retsch MM400 shaker and the bead-free solution was collected in a new RNase-free tube by making a small hole in the bottom of the tube and centrifuging for 2 min at 3,900 rpm at 4°C. The beads were rinsed with 500 μl of TRI Reagent and centrifuged again. The eluate was centrifuged for 2 min at 11,000 rpm, 4°C, to remove cellular debris. RNA was then prepared using an RNA isolation kit (Zymo Research). Reverse transcription was performed with 200 ng of purified RNA using Maxima Reverse Transcriptase (ThermoFisher) primed with oligodT in 50 μl and incubated for 30 min at 50°C. After 5 min at 85°C for enzyme inactivation, 2 μl of cDNA was used for each PCR reaction. When possible, RT-PCR primers were designed flanking an intron so that the presence of remaining genomic DNA in RNA samples could be detected, if any.

Complementation of the antidote function in Sha

A 2.5-kb genomic fragment comprising the coding sequence of *AT3G62530*, plus 1 kb upstream and 0.5 kb downstream, was amplified from Col-0 using specific primers (530_compl_GB_F and 530_compl_GB_R; Supplementary Table 2). Using the GoldenBraid strategy (https://gbcloning.upv.es/; (Sarrion-Perdigones *et al.* 2013), this fragment was cloned into the pUPD2 vector and combined with a transcriptional unit conferring red fluorescence to seeds (pCMV: DSRed: tnos; Morineau *et al.* 2017) into a pDGB3-omega2 binary plasmid. The resulting construct was transformed into *Agrobacterium tumefaciens* strain C58C1 pMDC90 and integrated by floral dipping into ShaL3[H] plants. The transformants were selected on the red fluorescence of seeds, checked by PCR, and 3 independent ShaL3[H] transformants were selected to analyze the bias in their progenies. A ShaL3[H] untransformed sibling was used as a control with an active pollen killer.

Cellular localization of the AT3G62530 protein

Translational fusion of AT3G62530 with the red fluorescent protein (RFP)

Using the GoldenBraid strategy, we built a transcriptional unit comprising the UBIQ10 promoter, the *AT3G62530* coding region omitting the stop codon (between S2_530_S3_F and S2_530_S3_R primers, Supplementary Table 2), the *RFP* coding sequence and the *RbcS* terminator. This *AT3G62530-RFP* translational fusion gene was associated with a gene conferring resistance to glufosinate as a selection marker. The resulting construct was transferred into *A. tumefaciens* strain C58C1 pMDC90 and integrated by floral dipping into *A. thaliana* Col-0 mt-yk (CS16264) homozygous plants that express a mitochondria-targeted yellow fluorescent protein (YFP) under the 35S promoter (Nelson *et al.* 2007). Transformants were selected in the greenhouse on sand irrigated with a solution of 7.5 g/l glufosinate (Basta, Bayer), and transferred to soil for 1–3 weeks before observation under the confocal microscope.

Confocal microscopy

Cellular localization of fluorescent proteins was assessed on leaf epidermal cells using a Leica SP8 confocal microscope with a 40× magnification objective. The bidirectional sequential mode by lines was used to prevent leakage of fluorescence between channels (scan speed: 400 Hz). YFP was excited at 488 nm and measured at 535–557 nm, RFP was excited at 561 nm and measured at 590–630 nm. Images were acquired by averaging 4 scans per frame. Raw images were processed with ImageJ v2.3.0/1.53f (http://imagej.nih.gov/ij). Colocalization of fluorescence signals was quantified (Pearson's coefficient of correlation) using the JACoP Plugin (Bolte and Cordelieres 2006) in 3 independent transformants. For a better visualization of the images shown, signal intensity was increased using the image/adjust tool in ImageJ.

Results

Elements leading to a segregation bias at the L3 locus are common in natural variants of *A. thaliana*

We previously detected several pollen killers in the selfed progeny of Mr-0 × Sha hybrids (Simon *et al.* 2016). Here, our study focuses on the locus named L3 and located at the bottom of chromosome 3, where a segregation bias against the Sha allele was observed. This transmission ratio distortion was linked to the death of pollen grains carrying the Sha sensitive allele at this locus in the presence of the Mr-0 killer allele. We also observed a strong bias at L3 against the Rak-2 allele in the Mr-0 × Rak-2 F2 population (Simon *et al.* 2016), suggesting that other natural variants than Sha possess alleles conferring sensitivity to the Mr-0 killer effect at L3. Likewise, other accessions than Mr-0 may carry killer alleles. Therefore, we crossed 30 accessions belonging to diverse genetic groups (Simon *et al.* 2012) with Mr-0 and/or Sha, and analyzed the segregation at L3 in the progenies of the hybrids. Twenty-six accessions were crossed with Mr-0; similar to Sha, 14 of them showed a bias in the progeny of the hybrid, and were thus classified as sensitive, while the others were classified as resistant (Table 1). As expected in presence of a gametophytic defect in these hybrids, genotype proportions were consistent with a 1:1 distribution of homozygotes Mr-0 and heterozygotes in most of the biased segregations (Supplementary Table 5). In parallel, 18 accessions were crossed with Sha; 5 of the F2s presented a bias against the Sha allele (Table 1), with a distribution of genotypes that was consistent with a gametophytic defect (1:1

distribution of homozygotes not Sha and heterozygotes, Supplementary Table 6), and were thus classified as killers. Coherently, none of the 14 accessions tested in crosses with Mr-0 and Sha was found to be both sensitive and killer (Table 1), while 5 accessions, including Col-0, were resistant and not killer. Therefore, this segregation distortion at L3 appears to be widespread in crosses between natural accessions of *A. thaliana*. We focused on the Mr-0 × Sha cross to further characterize the pollen killer.

The pollen killer at the L3 locus induces an allele-specific defect of pollen development

A plant heterozygous Sha/Mr-0 at L3 while homozygous Sha in the rest of its nuclear genome (named ShaL3^H) shows at L3 a bias against Sha homozygous progenies, which was linked to a deficit in pollen grains carrying the Sha allele (Simon *et al.* 2016). We observed a strong distortion when ShaL3^H was used as male in a cross with Sha, but no distortion when it served as female parent (Table 2). These results confirmed that the segregation distortion was only due to a male defect. Along this study, we relied on the measure of a significant segregation distortion at the locus L3 in the selfed progeny of a heterozygote, associated with the observation of aborted pollen in its anthers, to assess the activity of the pollen killer. We tested the segregations in the selfed progenies of plants heterozygous at L3 either in Sha (ShaL3^H) or Mr-0 (MrL3^H) nuclear backgrounds. Both genotypes produced progenies with very similar deficits in homozygous Sha at L3 (Table 3), indicating that the segregation distortion was independent of the nuclear background Sha or Mr-0. Accordingly, anthers from both genotypes showed dead pollen (Supplementary Fig. 1). We further studied the ShaL3^H genotype that flowers earlier than MrL3^H.

We used cytological approaches to determine whether the male dysfunction occurred during meiosis or during pollen development and to specify its timing. The male meiosis of ShaL3^H plants was identical to that of homozygous siblings (ShaL3^M and Sha) at all stages (Supplementary Fig. 2), excluding a meiotic defect as the source of the segregation bias in ShaL3^H plants. During male gametogenesis, abnormally developing pollen grains were observed in anthers of ShaL3^H plants from the bicellular stage, after the first pollen division (Fig. 2). The proportion of abnormal pollen increased in the tricellular stage, and about 35% of mature pollen grains were dead in ShaL3^H plants (Fig. 2; Supplementary Table 7), which reflects an incomplete penetrance since 50% dead pollen would be expected if the pollen killer effect was total.

We concluded that, in these plants, the Sha allele at L3 is poorly transmitted because most of the Sha pollen grains fail to develop properly from the binucleate stage and eventually die.

The pollen killer at the L3 locus contains three genetic elements

The L3 locus was previously mapped in a 280-kb interval at the bottom of chromosome 3 (Simon *et al.* 2016). We finely mapped the pollen killer in the genotype ShaL3^H using the presence of a segregation bias in the self-descent of recombinants to narrow down the L3 interval. Genetic markers homozygous for Sha or Mr-0 alleles in recombinants with a significant bias in their progenies were excluded from the candidate interval. This strategy allowed us to map all the genetic elements necessary for the pollen killer activity in an interval called PK3, corresponding to the region flanked by markers M5 and M13 (68 kb in Col-0) (Fig. 3a). Out of 4,717 genotyped plants, we detected 6 recombinants between M5 and M13, named Rec1 to Rec6. Only 3 of these plants, Rec4, Rec5, and Rec6, were recombined between markers M6 and

Table 1. Behaviors of natural accessions regarding the PK3.

Accession	Country	Tested in cross with Mr-0			Tested in cross with Sha		
		F2 plants genotyped	Frequency of Ho not Mr-0 in F2[a]	Behavior	F2 plants genotyped	Frequency of Ho Sha in F2[a]	Behavior
Shigu-2	Russia	183	0.30	Resistant	164	0.09***	Killer
Etna-2	Italy	169	0.27	Resistant	179	0.02***	Killer
Cant-1	Spain	152	0.25	Resistant	181	0.08***	Killer
Ct-1	Italy	184	0.24	Resistant	189	0.12***	Killer
Jea	France	172	0.26	Resistant	180	0.03***	Killer
Lov-5	Sweden	156	0.17	Resistant	422	0.23	Not killer
Blh-1	Czech Republic	174	0.25	Resistant	184	0.28	Not killer
Col-0	Poland	170	0.32	Resistant	182	0.30	Not killer
Bur-0	Ireland	184	0.22	Resistant	182	0.28	Not killer
Oy-0	Norway	181	0.20	Resistant	184	0.21	Not killer
N16	Russia	182	0.22	Resistant	–	nd	nd
Koch-1	Ukraine	178	0.24	Resistant	–	nd	nd
Sorbo	Tajikistan	98	0.04***	Sensitive	175	0.21	Not killer
Cvi-0	Cape Verde Islands	178	0.06***	Sensitive	191	0.27	Not killer
Ita-0	Morocco	184/184[b]	0.07/0.05***[b]	Sensitive	184	0.26	Not killer
Are-10	Portugal	188	0.01***	Sensitive	182	0.20	Not killer
Are-1	Portugal	162	0.00***	Sensitive	–	nd	nd
Kas-2	India	182	0.09***	Sensitive	–	nd	nd
Kz-1	Kazakhstan	182	0.12***	Sensitive	–	nd	nd
nov-01	Russia	173	0.12***	Sensitive	–	nd	nd
Kidr-1	Russia	184	0.03***	Sensitive	–	nd	nd
Kly-1	Russia	182	0.07***	Sensitive	–	nd	nd
N13	Russia	179	0.11***	Sensitive	–	nd	nd
Rak-2	Russia	182	0.02***	Sensitive	–	nd	nd
Stepn-1	Russia	174	0.07***	Sensitive	–	nd	nd
Zal-3	Kyrgyzstan	181	0.06***	Sensitive	–	nd	nd
Kz-9	Kazakhstan	–	nd	nd	167	0.25	Not killer
Kas-1	India	–	nd	nd	169	0.29	Not killer
Qar-8a	Lebanon	–	nd	nd	179	0.27	Not killer
Etn-0	Italy	–	nd	nd	145	0.28	Not killer

[a] ***P χ^2 for [1:2:1] distribution <0.001, see Supplementary Tables 5 and 6 for full data.
[b] Ita-0 was tested by crossing not with Mr-0 but with 2 other killer accessions, Ct-1 and Jea.
Ho, homozygotes (expected frequency in F2: 0.25); nd, not determined.

Table 2. Transmission through the male and female sides of the Sha allele in a heterozygous context.

Test cross mother × father	Number of plants[a]			P χ^2 (1:1)	f Sha[b]
	Sha	Hz	Total		
Sha × ShaL3[H]	40	132	172	2.3×10^{-12}***	0.23
ShaL3[H] × Sha	78	90	168	0.4[NS]	0.46

[a] Genotyped at marker M9 (Supplementary Table 1).
[b] Frequency of Sha homozygotes in the test-cross (expected frequency 0.5).
*** $P < 0.001$; NS, not significant.

Table 3. Segregation at L3 in selfed progenies in 2 different nuclear backgrounds.

Genotype	Number of plants[a]				P χ^2 (1:2:1)	f Sha[b]
	Sha	Hz	Mr-0	Total		
MrL3[H]	14	83	73	170	1.2×10^{-9} ***	0.08
ShaL3[H c]	18	87	73	178	4.0×10^{-8} ***	0.10

[a] Genotyped at marker M1 (Supplementary Table 1).
[b] Frequency of Sha homozygotes (expected frequency 0.25).
[c] From Simon et al. (2016).
*** $P < 0.001$.

M12, which are very close to M5 and M13, respectively (Supplementary Table 8). None of them presented a bias in its progeny, whereas Rec1, Rec2, and Rec3, whose heterozygous regions include M6 and M12, did. Pollen viability of the

recombinants, assessed by Alexander stainings, were consistent with the presence or absence of bias in their progenies (Fig. 3b). Further information on the genetic structure of the PK3 was obtained from the 3 plants recombined between M6 and M12. First, by crossing Rec4 and Rec5 with Sha, we converted their homozygous part of the interval into a heterozygous region (Fig. 3a; Supplementary Fig. 3 for the detail of crosses and segregating alleles). The offsprings of the hybrids Rec4 × Sha and Rec5 × Sha did not show any segregation bias, whereas their siblings heterozygous along the whole interval did (Supplementary Table 9). This indicated that the parts of the PK3 interval that were heterozygous in Rec4 and in Rec5 both contained elements necessary for the pollen killer activity. We therefore segmented the PK3 locus into 3 genetic intervals, named PK3A, PK3B, and PK3C (Fig. 3a), PK3A and PK3C both carrying elements necessary for a functional pollen killer. Then, in order to evaluate the PK3B interval, we crossed homozygous progenies of Rec4 and Rec6 that inherited the recombination events from their parents (see Supplementary Fig. 3 for the detail of crossed genotypes), and obtained a plant (Rec4 × Rec6) heterozygous at both PK3A and PK3C and homozygous Mr-0 at PK3B (Fig. 3a). The absence of segregation bias in the selfed progeny of this plant showed that PK3B also carried an element necessary for the pollen killer activity.

Therefore, each of the 3 parts of the PK3 interval contains at least 1 element required for the pollen killer activity. On one hand, Mr-0 alleles were required at PK3A and at PK3C, these 2 intervals thus contain killer elements. On the other hand, when the Sha allele was absent at PK3B while PK3A and PK3C
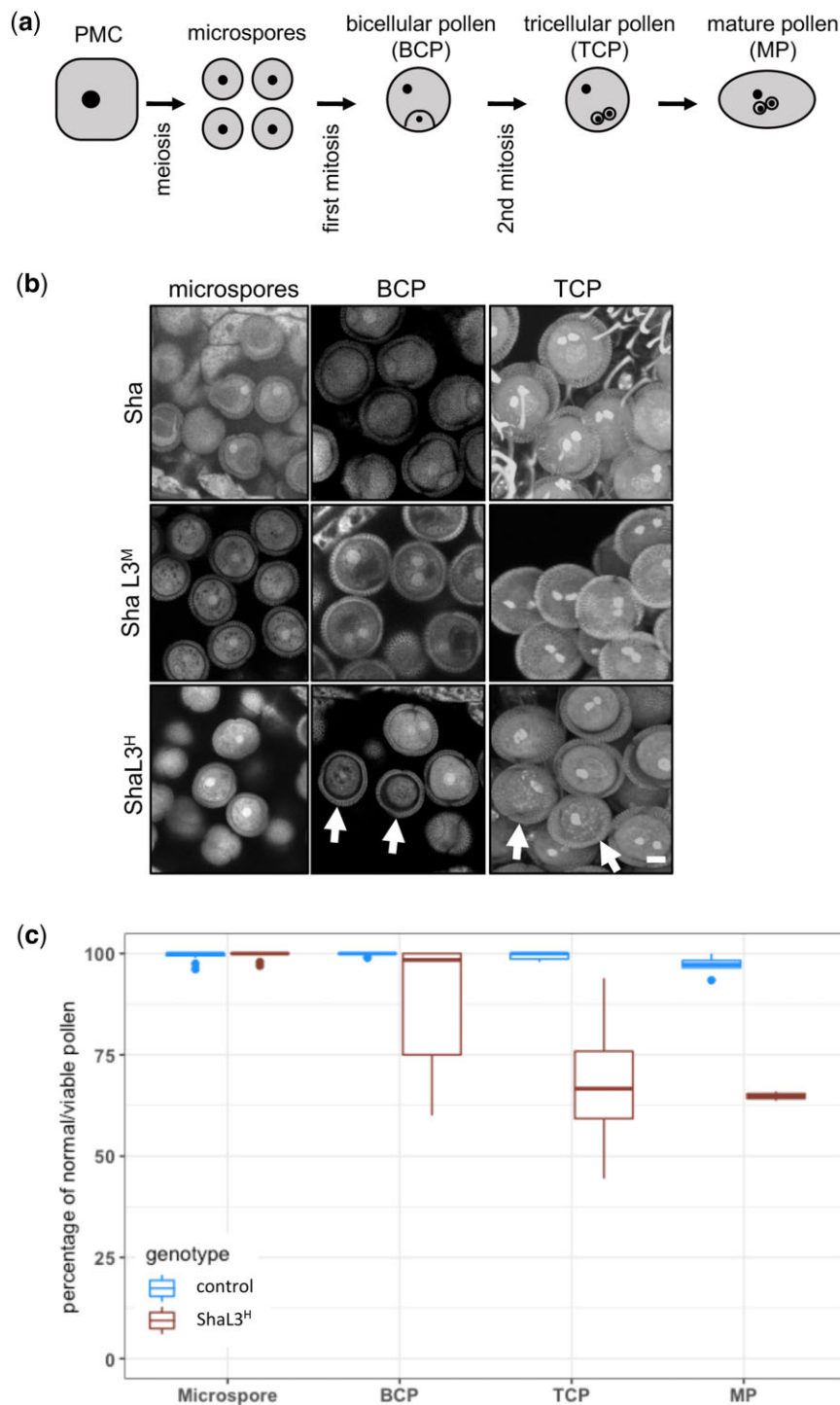
Fig. 2. Pollen from plants heterozygous at L3 abort progressively from the bicellular stage. a) The scheme outlines the main steps of pollen development in *A. thaliana*: the pollen mother cell (PMC) produces haploid microspores through meiosis; the first pollen mitosis produces bicellular pollen (BCP); the reproductive cell divides to give tricellular pollen (TCP), which evolves into mature pollen (MP). b) Typical confocal images (PI staining) of male gametogenesis in ShaL3^H plants (bottom row) and in Sha and ShaL3^M parental controls. White arrows indicate pollen grains with typical abnormal phenotypes: at the bicellular stage, they appear slightly smaller, with a less dense cytoplasm; at the tricellular stage, they have only one generative nucleus, and their cytoplasm appear less dense, with many punctuations. In order to clearly show the number of generative nuclei, the TCP images are accumulations of several confocal slices of the acquired stacks. Scale bar: 10μm. c) Percentage of normal developing (PI staining) or viable (Alexander staining) pollen (Supplementary Table 7) in anthers of parental (blue) and ShaL3^H (brown) plants at different stages of development (counts from confocal acquisitions of PI staining), and in mature pollen (counts from Alexander staining).

were heterozygous, the pollen killer was no longer active, indicating that either a target element from Sha was missing, or an antidote from Mr-0 was present in all the pollen grains produced.

The structure of the PK3 locus is highly variable

To highlight differences between Sha and Mr-0, we sequenced the entire locus in both accessions following a PCR-walking and Sanger-sequencing strategy. The overall structure of the PK3
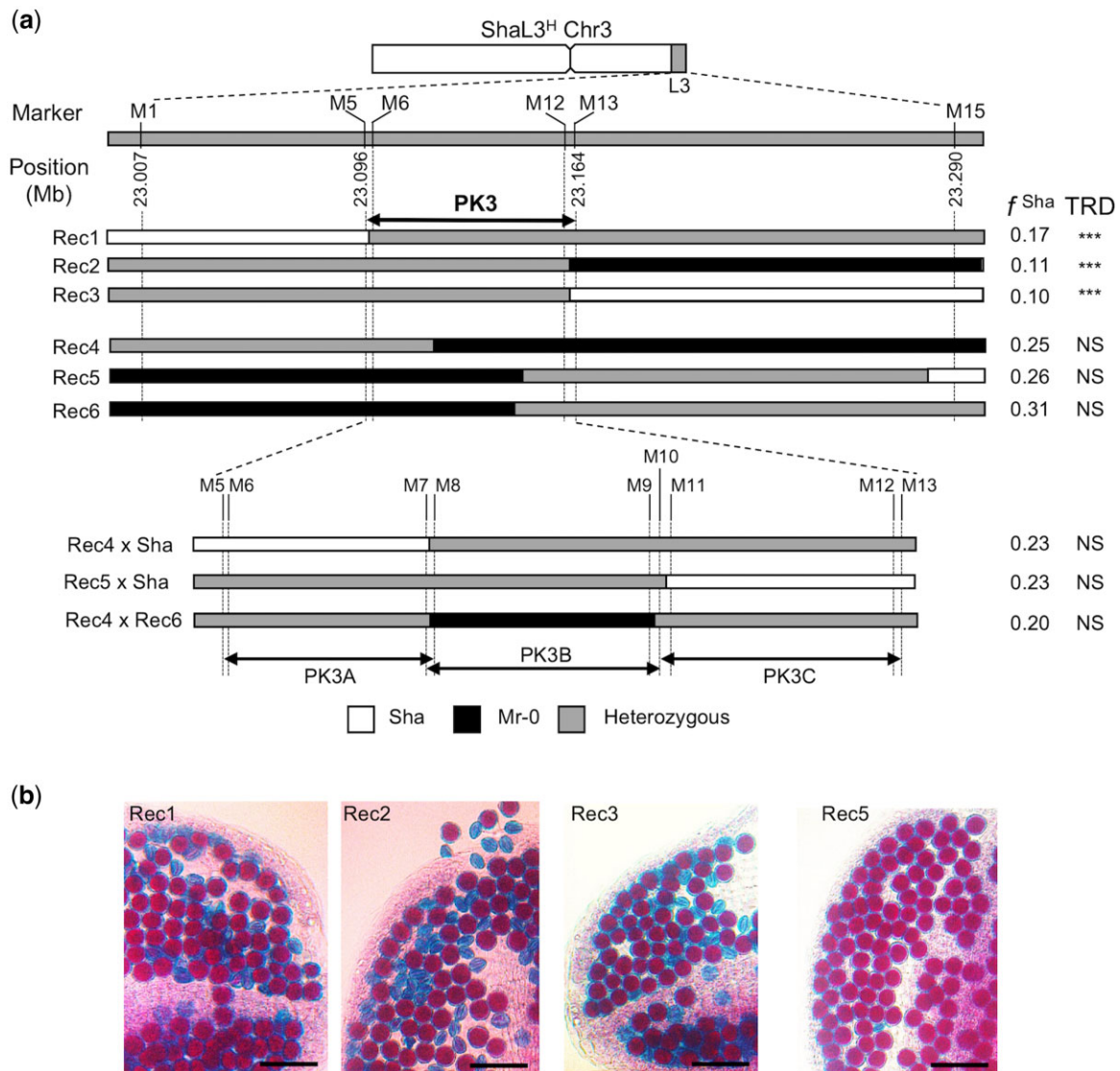
Fig. 3. Fine mapping of the PK3 locus. a) The upper panel shows the mapping of PK3 in the L3 locus delimited by the genetic markers M1 and M15 (Supplementary Tables 1 and 8). Positions of markers on the TAIR10 Col-0 genomic sequence are given. Markers M5 and M13 delimit the PK3 interval, where all the pollen killer elements required to cause the segregation bias are present. Markers M6 and M12 are very close to M5 and M13, respectively, they serve to map the recombination points of the recombinants Rec1, Rec2, and Rec3. The most relevant recombinants are shown, with the frequencies of Sha homozygotes (f^{Sha}) observed in their selfed progenies indicated on the right. TRD, transmission ratio distortion (χ^2 test for 1:2:1 segregation; sample size from 178 to 363, detailed results in Supplementary Table 8) ***$P < 0.001$, NS: not significant. The lower panel shows the further dissection of the PK3 locus into PK3A, PK3B, and PK3C. Because some markers are very close to each other, this diagram is not to scale. Genotypes used to dissect the PK3 derive from Rec4, Rec5, and Rec6 by different crosses (see Supplementary Fig. 3 for detail of crosses and Supplementary Table 8 for detailed segregation results). b) Pollen viability in recombinants used to delineate the PK3 interval. Representative images of Alexander staining of anthers from recombinants presented in (a). Aborted pollen grains are blue whereas viable pollen grains are red. Scale bars: 50 µm.

locus in Sha was very similar to that of Col-0 (Fig. 4a), the main differences being the absence of the transposable element (TE) *AT3G62455*, an insertion of 1,308 bp in the intron of *AT3G62460*, and an additional sequence of ~1 kb in the intergenic region between *AT3G62540* and *AT3G62550*. In contrast, the Mr-0 locus showed many structural variations such as large deletions, insertions, duplications, and inversions compared with Col-0 and Sha (Fig. 4a). Two TEs present in Col-0, *AT3G62455* and *AT3G62520*, were missing in Mr-0. The TEs *AT3G62475, AT3G62480*, and *AT3G62490* located in the PK3A region of Sha were absent from this region in Mr-0. However, homologs of *AT3G62475, AT3G62480*, and *AT3G62490* were present in an additional sequence of over 20 kb in the PK3B interval of Mr-0. Nonetheless, the same protein coding genes are present in the 3 genotypes,

even though Mr-0 has 2 copies of the gene *AT3G62510*, and 3 copies of the genes *AT3G62528, AT3G62530*, and *AT3G62540*, with 1 copy of *AT3G62530* and *AT3G62540* being inserted into the second intron of *AT3G62610* (Fig. 4a).

Because the PK3 locus is highly rearranged among these 3 genotypes, we looked at other accessions of known status for the pollen killer phenotype. The entire genomes of 4 killer, 3 sensitive and 3 resistant accessions, were *de novo* sequenced and assembled. PK3 sequence alignments revealed important structural variations relative to Col-0 in all the killer accessions (Fig. 4b). In addition, the killer alleles present structural differences (Supplementary Fig. 4). In contrast, both sensitive and resistant alleles, with the exception of the accession Bur-0, were mostly colinear with Col-0 (Fig. 4b). We compared the synteny of protein
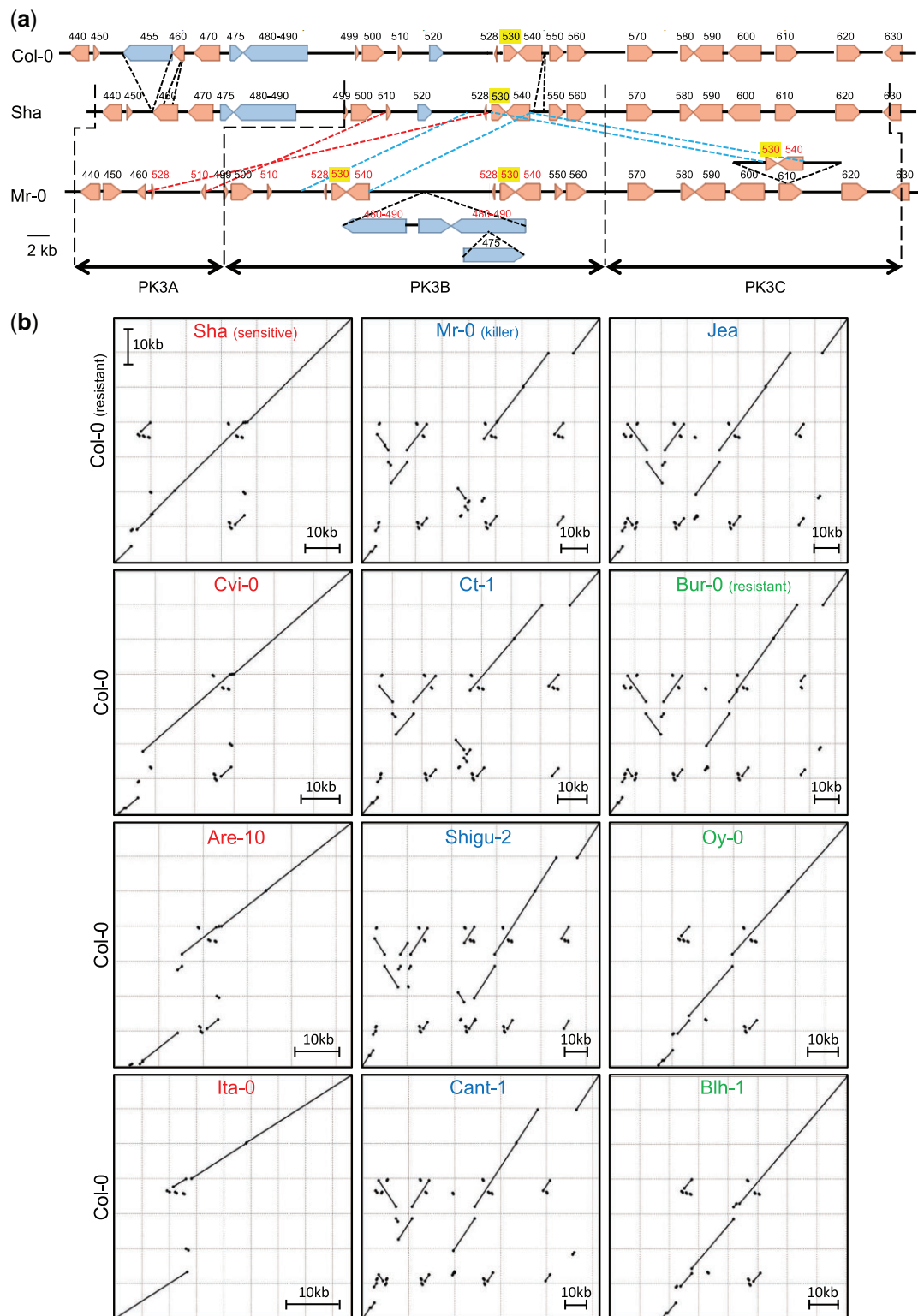
Fig. 4. Structural variation at the PK3 locus. a) PK3 locus in Col-0, Sha, and Mr-0. Beige and blue plain arrows represent protein coding genes and TEs, respectively, with their orientations. Gene labels on the Col-0 locus correspond to the last 3 digits of *AT3G62xxx* gene identifiers in the reference sequence (TAIR10), the same numbers on Sha and Mr-0 loci indicate genes homologous to the Col-0 genes; duplicated or triplicated genes in Mr-0 are labeled in red. Copies of the *AT3G62530* gene, studied further, are highlighted in yellow. Black, red, and blue dotted lines, respectively, represent insertions/deletions, inversions, and duplications. The genetically determined boundaries of the PK3A, PK3B, and PK3C intervals in Sha and Mr-0 are indicated (dashed broken lines). b) Dot plots of PK3 sequences of 12 accessions (*x* axes) against the reference Col-0 (*y* axes). Names of killer, resistant, and sensitive accessions are in blue, green, and red, respectively. Coordinates of the aligned sequences used are indicated in Supplementary Table 4. The vertical scale is for Col-0 sequence (identical in all plots) and a horizontal scale is given for each accession.
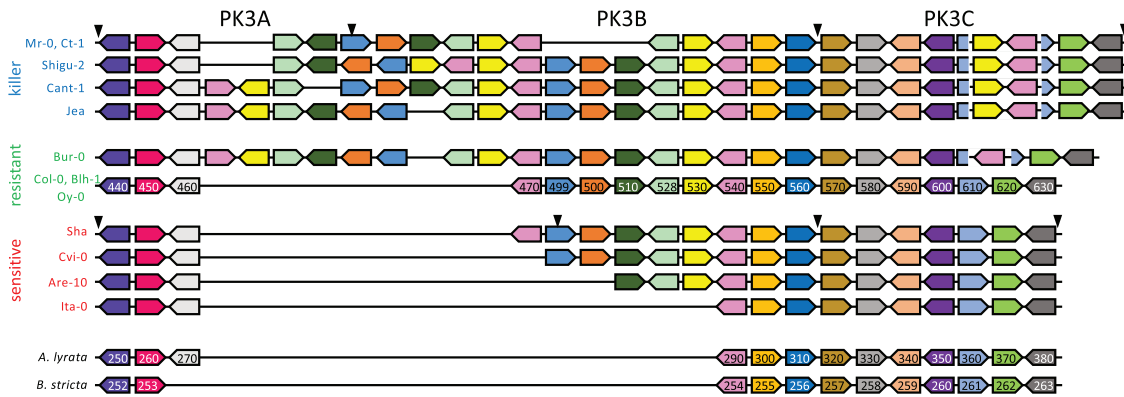
Fig. 5. Intra- and interspecies synteny of protein-coding genes at the PK3 locus. Representation of the PK3 alleles from 13 *A. thaliana* accessions and 2 related species, *Arabidopsis lyrata* and *Boechera stricta*, drawn to highlight synteny between protein coding genes. The diagram is not to scale and TEs are not represented. For *A. lyrata* and *B. stricta*, the structural annotation was obtained from Phytozome (https://phytozome.jgi.doe.gov/, last acessed 15th April 2021). Plain colored arrows represent coding genes with their orientations, each color refers to orthologs and paralogs of a same gene. Copies of the *AT3G62530* gene, studied further, are in yellow. Gene labels correspond to the last 3 digits of *AT3G62xxx*, *AL5G45xxx*, and *Bostr.13158s0xxx* gene identifiers in *A. thaliana* (Col-0 reference sequence TAIR10), *A. lyrata* (V2.1), and *B. stricta* (V1.2), respectively. Black arrowheads indicate the limits in Mr-0 and Sha of the genetically determined PK3A, PK3B, and PK3C intervals.

coding genes at the locus between *A. thaliana* accessions and its closest relative *A. lyrata* (Fig. 5). The locus in the sensitive accession Ita-0 is similar to that of *A. lyrata*, with orthologous genes in the same order and orientation. This structure is also found in other Brassicaceae species (Fig. 5), suggesting it represents the ancestral structure of the locus. On the other hand, Sha and the 3 resistant accessions Col-0, Blh-1, and Oy-0 have a duplication of the *A. lyrata* gene *AL5G45290*, that encodes a pentatricopeptide repeat protein (PPR), resulting in 2 nearly identical genes (*AT3G62470* and *AT3G62540*). Five additional protein-coding genes (*AT3G62499* to *AT3G62530*) compared with the *A. lyrata* and Ita-0 sequences were found between the 2 PPR paralogs in these 4 accessions. We also observed a great variability in the number of copies of these 5 genes among accessions. Compared with Col-0, Blh-1, Oy-0, and Sha, where they are present once, some of them are absent in the other sensitive accessions Cvi-0 and Are-10. In contrast, in the killer genotypes, they have undergone a variable number of duplications, with 1–4 copies depending on the identity of both the gene and the accession (Fig. 5).

Given the complexity and diversity of the locus, no obvious candidate gene appeared for killer or target/antidote elements. For the killer elements, an additional difficulty results from the fact that 2 elements are necessary for the killing activity, one located in the PK3A interval and the other in the PK3C interval, thereby preventing to draw conclusions from the comparison of killer and non killer alleles in a single interval. We thus focused on the PK3B interval, which contains either a target element in sensitive alleles or an antidote in killer and resistant alleles.

AT3G62530, expressed in young developing pollen, encodes the antidote

Col-0 has a resistant behavior regarding the pollen killer (Table 1) meaning that, in a poison-antidote model, Col-0 would carry an antidote element whose inactivation should let Col-0 pollen unprotected against the Mr-0 killer. A hybrid between Col-0 with an inactivated version of the antidote and Mr-0 would therefore present dead pollen, and produce an F2 with a segregation bias against the Col-0 allele at L3. We thus tested the poison-antidote model by using T-DNA insertion mutants available in this genetic background for the eight PK3B genes (Supplementary Table 10). None of the homozygous mutants (Col^{mut}) was distinguishable at the phenotypic level from Col-0 in our greenhouse conditions.

We crossed each Col^{mut} with Mr*fri*, an early-flowering Mr-0 genotype (File S1). All the Mr*fri* × Col^{mut} F1 plants exhibited only viable pollen, with the exception of the hybrid with a T-DNA insertion in *AT3G62530* (Col^{mut530}), which presented aborted pollen (Fig. 6a). Genotyping F2 populations at L3 revealed that they all followed the expected Mendelian proportions, with the exception of the Mr*fri* × Col^{mut530} and Mr*fri* × Col^{mut499} F2 populations (Fig. 6b; Supplementary Table 11). The latter presented a slightly biased segregation against the Col^{mut499} allele, but when the Mr*fri* × Col^{mut499} F1 was crossed as female or male parent using Mr*fri* as a tester, no mutant allele transmission bias was detected (Supplementary Table 12). In contrast, the bias against Col^{mut530} homozygotes observed in the Mr*fri* × Col^{mut530} F2 population was comparable to those observed in the progenies of hybrids with an active pollen killer (Table 4). In addition, this bias was observed when the Mr*fri* × Col^{mut530} F1 was used as male, but not as female, in a cross with Mr*fri* (Supplementary Table 12). Both sequencing the T-DNA insertion in *AT3G62530* and estimating expression level of *AT3G62530* revealed Col^{mut530} to be a knockout mutant (Supplementary Fig. 5). No dead pollen was observed in the anthers of homozygous and heterozygous Col^{mut530} (Supplementary Fig. 5). In addition, no segregation bias against the Col^{mut530} allele was detected in the selfed progeny of the heterozygous Col^{mut530} and in the progeny of the Sha × Col^{mut530} hybrid (Table 4). The presence of a killer allele was therefore necessary to trigger the death of Col^{mut530} pollen. This was further confirmed by crossing Col^{mut530} with another killer accession, Ct-1: while no bias was found in the progeny of the Ct-1 × Col-0 hybrid (Supplementary Table 13), a bias against the Col^{mut530} allele was observed in the progeny of Ct-1 × Col^{mut530} (Table 4), which was very similar to the bias against the Sha allele observed in the progeny of the Ct-1 × Sha cross (Supplementary Table 6). We concluded that *AT3G62530* was necessary to confer to Col-0 pollen its resistance against the PK3 killer activity.

In order to confirm this protective activity, we introduced the Col-0 genomic sequence of *AT3G62530* into a ShaL3^H background. In the progenies of 3 independent transformants, the segregation distortion against the Sha allele at L3 was corrected, whereas it was still observed in the progeny of an untransformed sibling (Table 5). The Col-0 genomic sequence of *AT3G62530* is therefore sufficient to rescue sensitive pollen from killer activity.
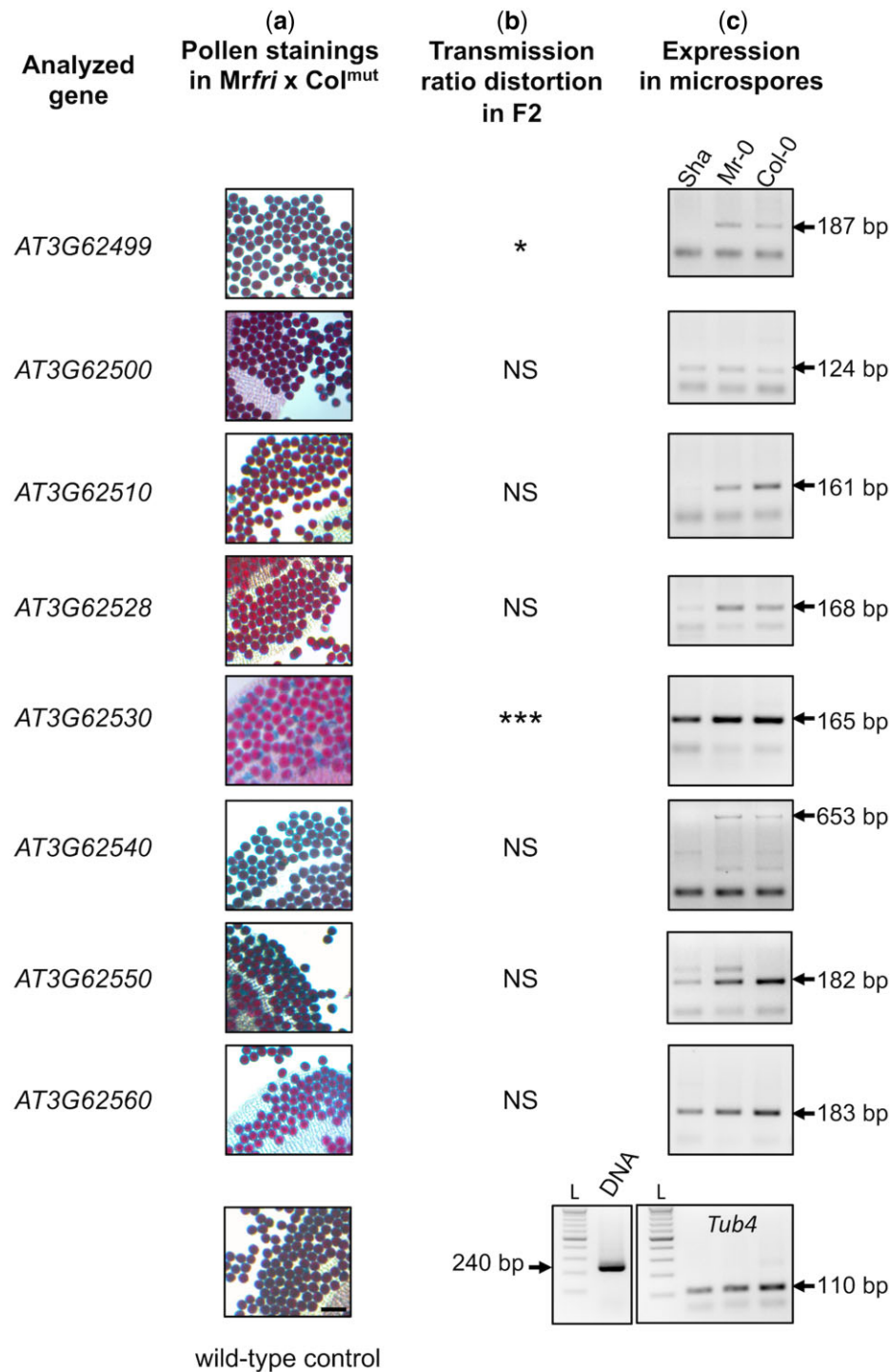
Fig. 6. Evaluation of the PK3B protein coding genes for a possible antidote function. Each homozygous Col-0 mutant was crossed with Mr*fri*, an early-flowering version of Mr-0. a) Typical Alexander stainings of anthers for each Mr*fri* × Col^mut F1 (viable pollen grains are stained in red, aborted pollen grains appear in blue. Scale bar: 50μm). b) Transmission ratio distortion at PK3 in F2 progeny (~180 individuals) of each F1 (NS, not significant; *P-value <0.05; ***P-value < 0.001; data in Supplementary Table 11). c) RT-PCR results for expression of each PK3B gene in microspores from plants either Sha, Mr-0, or Col-0 at the locus. The Mr-0 allele was analyzed in ShaL3^M, the 2 other alleles in the natural accessions. The *Tub4* gene was used as a positive control for RT-PCR; amplification of genomic DNA on the left shows its absence in RT-PCR products. PCR primers used (Supplementary Table 2) do not distinguish the 2 or 3 copies of *AT3G62510*, *AT3G62528*, *AT3G62530*, and *AT3G62540* that exist in the Mr-0 allele. The expected sizes of the amplification products are given. L, gene ruler ladder mix (Thermo Scientific).

In the poison-antidote model, the antidote must be expressed in cells that need to be protected from the poison, in particular in developing pollen. RT-PCR assays showed that *AT3G62530*, in addition to being expressed in leaves (Supplementary Fig. 5), is expressed in microspores (young developing pollen, before the first pollen mitosis; Fig. 2) from plants being either Col-0, Sha or Mr-0 at the locus (Fig. 6c).

Altogether, the above results are perfectly in line with a poison-antidote system where *AT3G62530* codes the antidote, with Col-0 and Mr-0 carrying functional forms of the antidote

Table 4. Segregation at L3 in F2s where the Col^{mut530} allele is confronted to a killer allele (Mrfri or Ct-1) or to a non killer allele (Col-0 or Sha)

Genotype	Number of plants				P χ^2 (1:2:1)	f Col$^{mut530\ a}$
	Not Col^{mut530}	Hz	Col^{mut530}	Total		
Mrfri × Col$^{mut530\ b}$	66	99	18	183	2×10^{-6} ***	0.10
Col-0 × Col$^{mut530\ c}$	49	85	46	180	0.5^{NS}	0.26
Sha × Col$^{mut530\ b}$	46	101	37	184	0.3^{NS}	0.20
Ct-1 × Col$^{mut530\ b}$	74	89	21	184	2×10^{-7} ***	0.11

[a] Frequency of Col^{mut530} homozygotes (expected frequency 0.25).
[b] Genotyped at marker M2, M1, and M17, respectively (Supplementary Table 1).
[c] Genotyped with PCR primers used to characterize the mutant (Supplementary Fig. 5 and Table 2). The F2s Mrfri × Col-0 and Ct-1 × Col-0 present no segregation distortion (Supplementary File 1 and Table 13).
*** $P < 0.001$; NS, not significant.

Table 5. Segregation at L3 in selfed progenies of ShaL3^{H} plants transformed with the Col-0 genomic sequence of AT3G62530.

Genotype	Number of plants[a]				P χ^2 (1:2:1)	f Sha[b]
	Sha	Hz	Mr-0	Total		
Control ShaL3^{H}	20	89	75	184	$7\ 10^{-8}$ ***	0.11
Transformant #1	42	90	52	184	0.56^{NS}	0.23
Transformant #2	37	103	44	184	0.21^{NS}	0.20
Transformant #3	46	101	37	184	0.27^{NS}	0.25

[a] Genotyped at marker M1 (Supplementary Table 1).
[b] Frequency of Sha homozygotes (expected frequency 0.25).
*** $P < 0.001$; NS, not significant.

while Sha has a nonfunctional antidote allele. We therefore named the gene AT3G62530 as APOK3, for **A**NTIDOTE OF **PO**LLEN **K**ILLER ON CHROMOSOME **3**.

APOK3 encodes a chimeric protein addressed to mitochondria

Because APOK3 is expressed at similar levels in microspores carrying the sensitive Sha allele and microspores carrying resistant Col-0 or Mr-0 alleles (Fig. 6c), we hypothesized that the allelic differences for antidote activity are due to differences at the protein level.

In Col-0, APOK3 encodes a protein of 221 amino acids, belonging to the ARM-repeat superfamily (https://www.arabidopsis.org/), defined by the presence of tandem repeats generally forming alpha-helixes (https://supfam.org/). Further analysis of structural domains identified 3 HEAT repeat domains (Fig. 7a). Different parts of APOK3 are very similar to parts of proteins encoded by 3 genes located on chromosome 3, i.e. AT3G62460 (95% identity and 98% similarity on residues 1–44), AT3G43260 (66% identity and 73% similarity on residues 44–142) and AT3G58180 (57% identity and 78% similarity on residues 81–221) (Fig. 7b). Interestingly, AT3G62460 is located in the PK3A interval. AT3G43260 and AT3G58180 are both annotated as related to deoxyhypusine hydroxylases of other organisms. However, protein sequence alignments indicate that the canonical Col-0 deoxyhypusine hydroxylase is encoded by AT3G58180 (Supplementary Fig. 6). The N-terminal part of APOK3, homologous to AT3G62460, includes a mitochondria-targeting peptide (Fig. 7a), suggesting that APOK3 is addressed to mitochondria. We confirmed the mitochondrial location of APOK3 by using a translational fusion with the fluorescent RFP protein (Fig. 7c). We therefore hypothesized that, if the antidote acts in the mitochondria, the strength of the bias due to the pollen killer could be influenced by the mother plant cytoplasmic background. As a first insight in this direction, we

compared the segregation distortion in the progenies of plants heterozygous Sha/Mr-0 at PK3 differing only by their cytoplasmic backgrounds. We observed that the bias was stronger in the Sha than in the Mr-0 cytoplasmic background (Fig. 7d; Supplementary Table 14). We concluded that the antidote function of APOK3 is likely to be sensitive to variation in the mitochondrial genome.
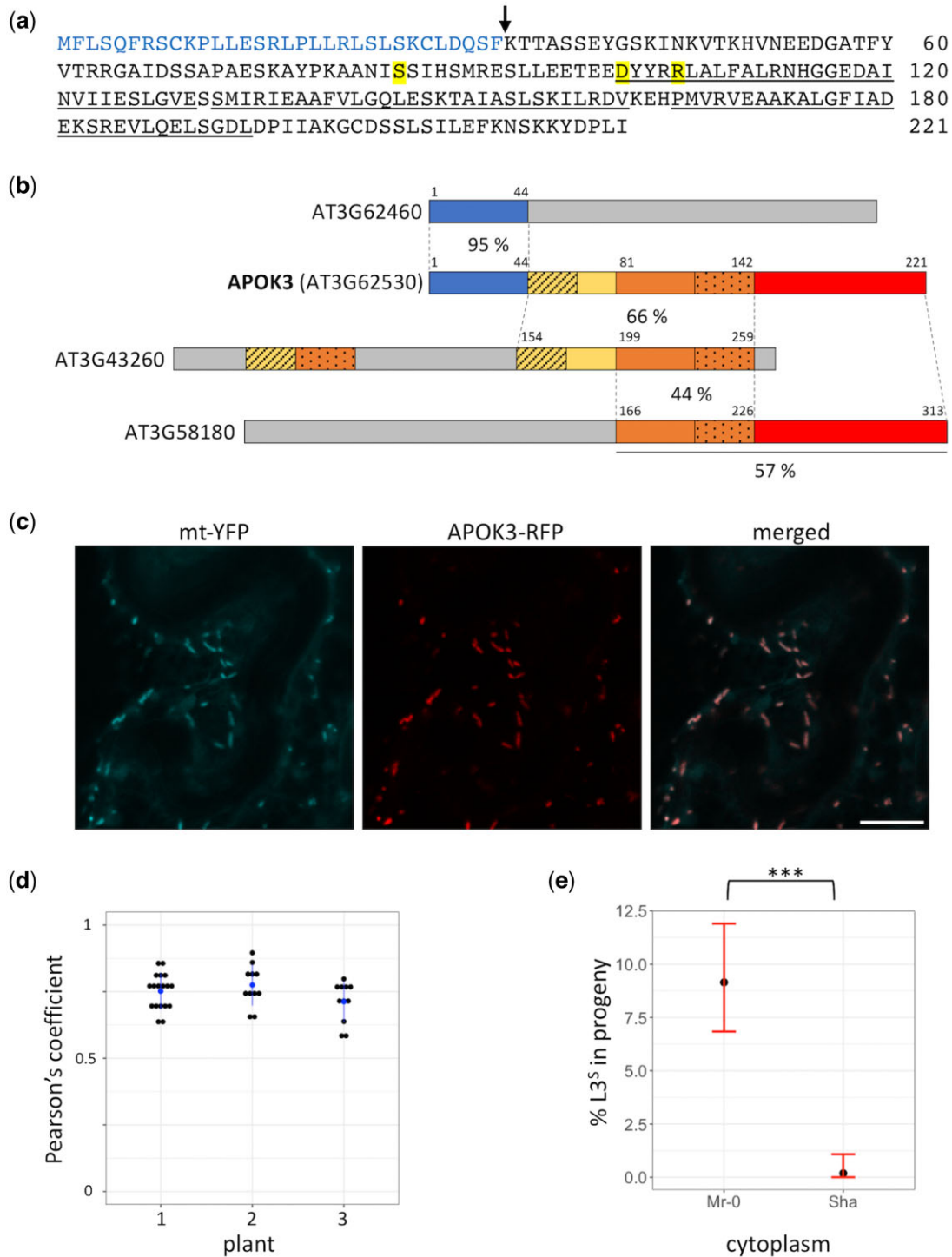
APOK3 has undergone several duplication events within PK3 killer alleles

Mr-0 has 2 strictly identical copies of APOK3 in the PK3B interval, which are nearly identical to Col-0 and Sha genes. Mr-0 also has a third copy inserted with other sequences in the intron of AT3G62610 in the PK3C interval (Fig. 8a). The N-terminal part of this copy, however, differs from that of the other two and is not homologous to AT3G62460, which suggests that this copy is functionally different from the others. This third copy was therefore named APOK3-*like*.

All the killer accessions analyzed in this study have an APOK3-*like* copy inserted in AT3G62610, and multiple copies of APOK3 (Fig. 5). Similar to Mr-0, Ct-1 has 2 APOK3 copies, when Cant-1, Jea, and Shigu-2 have 3 copies. Bur-0, which has a resistant but not killer behavior, also possesses 3 copies of APOK3, but no APOK3-*like*. The other resistant accessions and the sensitive ones have only one copy of APOK3, excepted Ita-0, in which no copy exists at the locus (Fig. 5). No other copy of APOK3 was found elsewhere in the Ita-0 genome, neither by searching in the genomic sequence nor by PCR amplification. Similarly, we did not find any gene encoding a protein more similar to APOK3 than to AT3G58180 neither in A. *lyrata* nor in other Brassicaceae sequences available in public databases. Altogether, these results suggest that APOK3 is specific to A. *thaliana* and has evolved within this species.

Sequence polymorphisms differentiate antidote and non antidote forms of APOK3

To further explore the sequence variation of APOK3 in relation with its antidote activity, we amplified and Sanger-sequenced APOK3 in all the accessions whose behavior for the PK3 had been determined (Table 1). The APOK3 copies found in each killer accession whose genome was *de novo* sequenced were identical. In Etna-0, the only killer accession for which no genomic sequence was available, we obtained a unique Sanger sequence for APOK3, indicating that, whatever the number of copies this accession has, they are identical. Phylogenetic analysis clustered the sequences into 3 clades, one grouping all copies found in accessions possessing the antidote activity. A second cluster groups most copies from the sensitive accessions, with the exception of

(a)

MFLSQFRSCKPLLESRLPLLRLSLSKCLDQSFKTTASSEYGSKINKVTKHVNEEDGATFY   60
VTRRGAIDSSAPAESKAYPKAANISSIHSMRESLLEETEEDYYRRLALFALRNHGGEDAI 120
NVIIESLGVESSMIRIEAAFVLGQLESKTAIASLSKILRDVKEHPMVRVEAAKALGFIAD 180
EKSREVLQELSGDLDPIIAKGCDSSLSILEFKNSKKYDPLI   221

(b)

(c)

(d)

(e)

Fig. 7. Structure and cellular localization of APOK3. a) Protein sequence of APOK3 in Col-0. The N-terminal mitochondria-targeting peptide (likelihood = 0.9) is written in blue, with an arrow indicating the potential cleavage site between amino acids 32 and 33. The 3 HEAT-repeat domains are underlined. The 3 amino acids highlighted in yellow differ between antidote and non antidote forms of the protein (see below). b) Representation of the composite APOK3 structure. Colors indicate the homologous regions between APOK3 and AT3G62460 (blue), AT3G43260 (yellow), or AT3G58180 (red). The orange box corresponds to a region homologous between APOK3 and both AT3G43260 and AT3G58180. Percentages of identity are indicated between proteins, that for APOK3 and AT5G18580 is indicated under the latter. The hatched and stippled boxes represent duplicated regions of AT3G43260 homologous to APOK3 or AT3G58180. Gray boxes indicate regions of the proteins not aligning with APOK3. c) Mitochondrial localization of the APOK3-RFP fusion protein in the Col-0 mt-yk marker line. A typical acquisition of epidermal cell is shown. Mitochondria appear as small organelles resembling bacteria: cyan, mitochondria-targeted YFP (mt-YFP); red, APOK3-RFP. Scale bar = 10 μm. d) Quantification of fluorescence colocalization. Black dots represent signal overlaps measured from acquisitions for 3 independent transformants; blue dots, and vertical lines indicate mean and SD. e) Effect of cytoplasmic background on the strength of the PK3-induced distortion. Three independent pairs of reciprocal F1s in either Mr-0 or Sha cytoplasmic background were obtained by crossing both ways with Sha 3 mapping recombinants and selecting F1s heterozygous on the whole PK3 interval. Percentages of Sha homozygotes at PK3 (L3^S) were determined in the progenies of these F1s (Supplementary Table 14). Dots indicate the average of Sha homozygote percentage for each cytoplasmic background. Red vertical bars indicate 0.95 confidence intervals. ***Fisher's test $P < 0.001$.
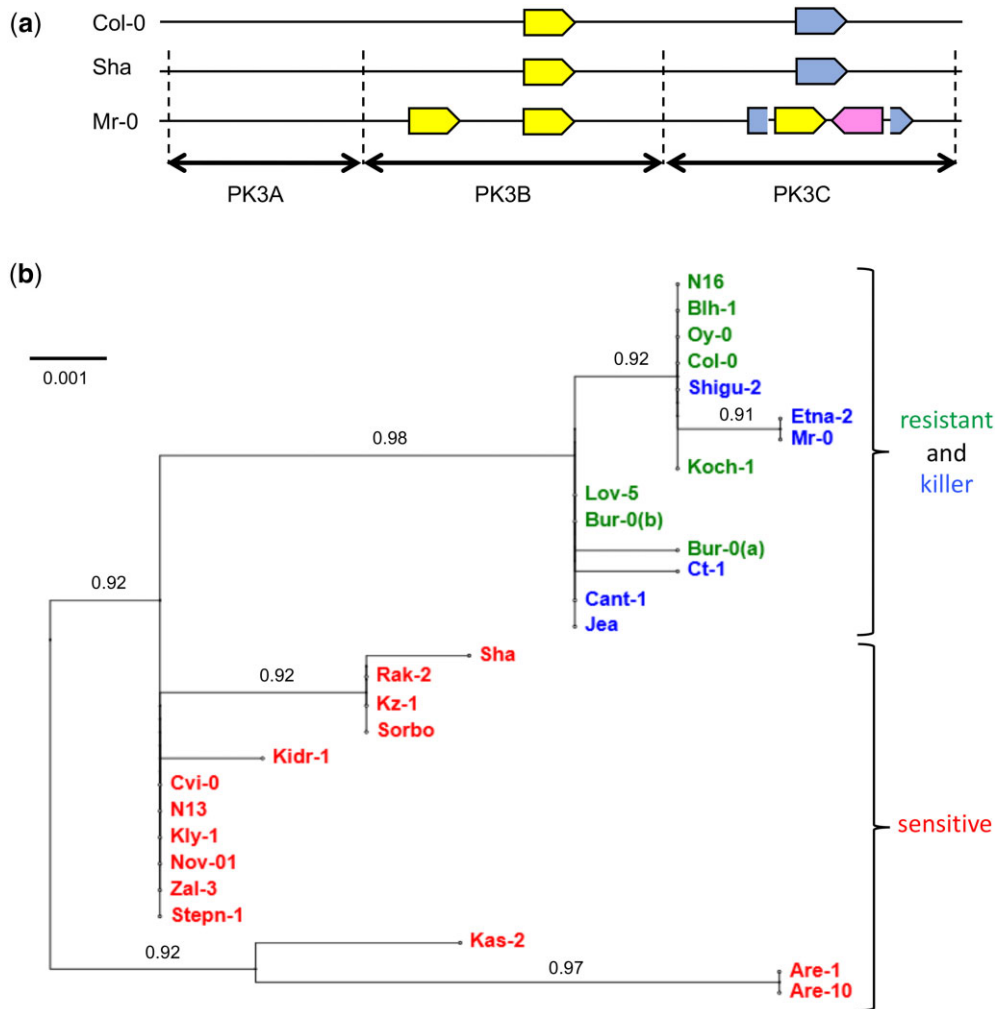
Fig. 8. Copy number variation and phylogeny of *APOK3*. a) Schematic representation of *APOK3* homologous sequences (yellow arrows) in the PK3 locus in Col-0, Sha and Mr-0. The scheme is not to scale. Gene colors are as in Fig. 5. The blue arrows represent *AT3G62610*, in which *APOK3-like* is inserted in Mr-0 together with *AT3G62540* (pink arrow). b) Variation in *APOK3* sequences: unrooted phylogenetic tree generated from *APOK3* copies of 27 accessions of known status for the pollen killer. Accessions with several copies of *APOK3* are displayed only once because their copies are identical, except Bur-0 where 2 copies (a) differ from the third (b) by one SNP. Branch support values are displayed over the branches.

Are-1, Are-10, and Kas-2 which branch together as a third clade (Fig. 8b).

We identified 11 different *APOK3* haplotypes, 5 from resistant and killer accessions and 6 from sensitive accessions (Table 6). One polymorphism located 36 pb upstream of the ATG start codon and 3 nonsynonymous SNPs in the coding sequence distinguished resistant and killer from sensitive alleles. Consequently, APOK3 proteins with antidote activity differ from those without antidote activity by the 3 amino acid changes C85S, V101D, and C105R (Table 6). Interestingly, the last 2 of these amino acids are located in the first HEAT-repeat domain of the protein (Fig. 7a), suggesting these changes could modify the protein interactions and could make it functional or non functional as an antidote.

Considering the 3 non synonymous SNPs correlated with APOK3 antidote activity in our set of 28 accessions tested, we defined 2 main haplotypes, called AAC, found in the resistant and killer accessions, and TTT, found in the 14 sensitive accessions possessing *APOK3* (Table 6). To obtain a wider picture of *APOK3* variation within the species, we looked at the SNPs detected in the Arabidopsis 1001 Genomes Project (https://1001genomes.org/) (Fig. 9; Supplementary Table 15). Among 1,042 accessions for which data were available, *APOK3* was missing in 9% of the accessions, and the AAC and TTT haplotypes were present in 37% and 10% of the accessions, respectively. Another haplotype, TAC, absent in our sample, was present in 44% of the accessions. The proportions of the different haplotypes were roughly even in several genetic groups defined within the 1001 Genomes accessions (The 1001 Genomes Consortium 2016)—referred to as "admixed," "South Sweden," "Spain," and "Italy-Balkan-Caucasus." In contrast, the AAC haplotype was predominant in the "Central Europe" group and present in all the "North Sweden" accessions, when it is uncommon in the "Western Europe," "Germany," and "Asia" groups. The haplotype TTT is present in 88% of the accessions from the "Asia" group, whereas the haplotype TAC is highly common in the "Western Europe" and "Germany" groups.

Discussion

As pointed out by Burga *et al.* (2020), the discovery of poison-antidote systems in eukaryotes is most often fortuitous, especially because nearly all these systems are species specific. In *A. thaliana*, we found pollen killers adventitiously when looking at a hybrid male sterility, of which they turned out to be unexpected components, in a cross between Sha and Mr-0 natural variants

Table 6. *APOK3* haplotype diversity among accessions of known status for antidote activity.

	SNP position	−36	6	17	27	56	68	152	241	248	253	332	341	363	373	378	384	415	583	765	840
Antidote	**Col-0, Shigu-2, Oy-0, Blh-1,** N16, Koch-1	c	T	T	T	T	C	T	G	A	**A**	a	t	A	**A**	T	**C**	G	G	t	T
	Mr-0, Etna-2	.	.	.	.	.	.	.	T	.	**.**	.	.	.	**.**	.	**.**	.	.	.	.
	Ct-1	.	.	.	.	.	.	.	.	.	**.**	.	.	T	**.**	.	**.**	.	.	g	.
	Bur-0 APOK3-1 & 2	.	.	.	.	.	.	.	.	.	**.**	.	.	.	**.**	.	**.**	.	C	g	.
	Jea, Cant-1, Bur-0 APOK3-3, Lov-5	.	.	.	.	.	.	.	.	.	**.**	.	.	.	**.**	.	**.**	.	.	g	.
Non antidote	**Cvi-0**, N13, Kly-1, Nov-01, Zal-3, Stepn-1	t	.	.	.	.	.	.	.	.	**T**	.	.	.	**T**	.	**T**	.	.	g	.
	Kidr-1	t	.	.	.	.	.	.	.	.	**T**	.	c	.	**T**	.	**T**	.	.	g	.
	Sorbo, Kz-1, Rak-2	t	.	.	.	.	.	A	.	.	**T**	.	.	.	**T**	.	**T**	T	.	g	.
	Sha	t	.	.	.	.	A	A	.	.	**T**	.	.	.	**T**	.	**T**	T	.	g	.
	Kas-2	t	.	.	.	C	.	.	.	C	**T**	.	.	.	**T**	G	**T**	.	.	g	G
	Are-10, Are-1	g	C	G	C	.	.	.	.	C	**T**	g	.	.	**T**	G	**T**	.	.	g	.
	Amino acid change position		6	19	23	51	81	83			**85**			98	**101**	103	**105**	115	141		
	reference (Col-0)		F	L	S	V	A	N			**S**			T	**D**	Y	**R**	G	V		
	other		C	S	Y	D	S	T			**C**			S	**V**	D	**C**	V	L		

The names of the accessions whose entire PK3 sequences are known are in bold. Base 1 of the nucleotide sequence is the A of the start codon. The nucleotides in the exons are in upper case and those in the untranslated regions in lower case. Nucleotides identical to the Col-0 reference sequence are represented as dots. SNPs and amino acids specific of antidote vs non antidote forms are in bold. Multiple sequence alignment was done with ClustalW (https://www.genome.jp/tools-bin/clustalw).
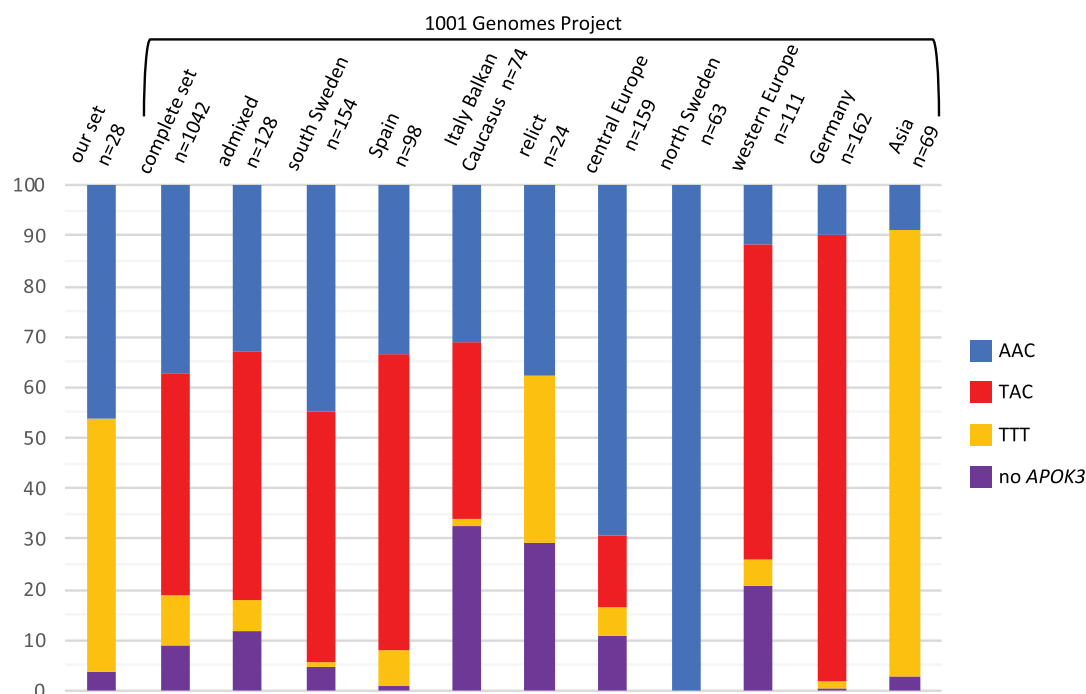
Fig. 9. Percentages of main *APOK3* haplotypes in different sets of accessions. Haplotypes were defined on the basis of the 3 nonsynonymous SNPs correlated with APOK3 antidote activity in our set of 28 accessions (A/T, A/T, C/T, Table 6). Polymorphism data for 1,042 accessions and diversity groups are from the 1001 Genomes Project (https://1001genomes.org/, last accessed 20th December 2021).

(Gobron *et al.* 2013; Simon *et al.* 2016). In this study, we focused on the pollen killer responsible for a deficit in Sha homozygotes at the bottom of chromosome 3 in the progeny of Mr-0/Sha hybrids. A transmission ratio distortion at the same locus was found in the progenies of 19 out of 44 hybrids involving either Mr-0 or Sha as one of the parents, indicating that sensitive or killer alleles are not rare in *A. thaliana*. This is in agreement with the report by Seymour *et al.* (2019) of frequent transmission ratio distortion in F2 populations from 80 *A. thaliana* founders, although these authors captured all types of segregation distorters, using a genome wide detection approach. Indeed, F2 distortion at one locus can result from allelic interactions either at the same locus or between different loci (Fishman and McIntosh 2019). We cannot formally exclude that some of the biases observed in our F2 families are partially caused by other allelic interactions than the pollen killer studied here. However, the presence of a distortion was subsequently correlated with sequence variations in the antidote genes of the parental accessions, supporting our conclusion that the bias observed in these crosses is due to the PK3. In addition, in biased F2 populations, most killer:heterozygote ratios are close to the 1:1 ratio that is expected for a bias due to a gametophytic defect controlled by a single locus. Exceptions were the F2s of Mr-

0 with Kas-2, Kly-1, N13, and Cvi-0, which presented an excess in heterozygotes (>55%). These cases might be explained by the presence, at loci partially linked to the PK3, of other segregation distorters that affect the transmission of the Mr-0 allele.

In ShaL3^H plants, mature anthers contain only ~35% of dead pollen, instead of 50% expected if the killer effect was total. If all these dead pollen grains carry the Sha allele, we theoretically expect 11.5% of Sha homozygotes in the progeny, which is very close to the 10% observed. Incomplete penetrance was also observed in other reported gamete killers and meiotic drivers, for example, in tomato (Rick 1966) or rice (Matsubara et al. 2011). In some cases, as in distorters from Drosophila (Larracuente and Presgraves 2012), rice (Koide et al. 2012), and yellow monkeyflower (Finseth et al. 2021), the incomplete penetrance of the phenotype can be due to the presence of unlinked modifier loci that modulate the strength of the distorter. It is conceivable that the variation in the strength of the biases observed among our F2s results from the presence in some hybrids of such unlinked epistatic loci. Besides this, changing the Mr-0 cytoplasmic background to the Sha one had a strong effect on the elimination of homozygous Sha progenies, making it almost absolute (only one Sha among 512 progenies). Considering that APOK3 is a mitochondrial protein, it is tempting to hypothesize that variations in the mitochondrial genome can modulate pollen sensitivity to the killer. In this case, the mitochondrial genome might be considered as a "modifier locus" of the pollen killer. Interestingly, the antidote of the rice qHMS7 locus is also addressed to mitochondria (Yu et al. 2018), but the effect of the cytoplasmic background was not explored.

Resistant alleles have been reported not only for gamete killers in rice (Wang et al. 2005; Liu et al. 2011; Yu et al. 2016; Koide et al. 2018) and tomato (Rick 1971), but also for spore killers in fungi such as Neurospora (Turner 2001) and for the meiotic driver wtf in fission yeast (Nuckolls et al. 2017). Resistant genotypes are valuable resources to distinguish between killer-target and poison-antidote models of gamete killers. While resistant genotypes miss the target in the first case, they have an antidote like the killer genotypes in the second case. Here, we took advantage of available mutants in the Col-0 resistant accession to establish that PK3 functions as a poison-antidote system and to identify the gene encoding the antidote, APOK3, as AT3G62530.

The molecular function of APOK3 remains to be elucidated. It is annotated belonging to the ARM-repeat superfamily, which mediates numerous cellular processes including signal transduction, cytoskeletal regulation, nuclear import, transcriptional regulation, and ubiquitination (Samuel et al. 2006). The remarkable features of this protein are its mitochondrial location, its interacting domains and its chimeric structure. Consistently with our results, APOK3 has been repeatedly found in A. thaliana mitochondrial proteomes (Heazlewood et al. 2004; Klodmann et al. 2011; Taylor et al. 2011; Konig et al. 2014; Senkler et al. 2017). It has 3 HEAT repeat domains, predicted to form superhelixes (Andrade and Bork 1995) and to mediate protein–protein interactions (Andrade et al. 2001). Specific investigations will be needed to determine whether these domains are involved in interactions with other mitochondrial proteins. Nevertheless, it is noteworthy that APOK3 was reported to associate with ISOCITRATE DEHYDROGENASE 1, a regulatory subunit of NAD-dependent enzyme of the tricarboxylic acid cycle (TCA), and SUCCINATE DEHYDROGENASE 4, a subunit of the mitochondrial respiratory complex II, in a study exploring the protein–protein interaction network of the plant TCA cycle (Zhang et al. 2018). Although the functional relevance of these protein interactions as well as the ability of APOK3 to bind zinc bivalent ions (Tan et al. 2010) remain to be assessed, they represent promising entry points

to functional studies aiming at discovering how APOK3 fulfills its antidote activity. This question has been elucidated in very few systems, among which the wtf spore killer of fission yeast in which the antidote was shown to coassemble with the corresponding poison and address the toxic aggregates to sequestering vacuoles (Nuckolls et al. 2020). Interestingly, these authors also found that genes involved in mitochondrial functioning counteract the poison activity, suggesting a role, yet to be uncovered, for mitochondria in this antidote mechanism.

APOK3 is not essential to A. thaliana and seems to have no other biological role than being the PK3 antidote. In Col-0, the apok3 KO mutation did not cause any visible phenotype alteration in our greenhouse conditions. Moreover, APOK3 is missing in the Moroccan accession Ita-0, as well as in accessions of almost all genetic groups identified in the 1001 Genomes Project (The 1001 Genomes Consortium 2016). In addition, APOK3 has no ortholog in the A. thaliana closest relative A. lyrata and was not found in the publically available Brassicaceae sequences. Together, these observations argue for the emergence of APOK3 after A. thaliana divergence from its common ancestor with A. lyrata. The APOK3 structure suggests that the gene originated from a duplication of a member of the AT3G43260 and AT3G58180 gene family, followed by the recruitment of the AT3G62460 mitochondrial targeting sequence. This way of gene emergence seems not rare. For instance, in O. sativa ssp. japonica, a comparative genome analysis detected 28 new O. sativa specific genes on chromosome 3, 14 of them being chimerical (Zhang et al. 2013). However, to our knowledge, their biological functions remain unknown.

Because of the absence of APOK3 in any related species, we could not conclusively infer whether non antidote forms preceded antidote ones or rather derived from them. However, because A. thaliana Madeiran strains such as Are-1 and Are-10 have been described as archaic (Fulgione et al. 2018), our phylogenetic analysis of APOK3 suggests that antidote forms evolved more recently. This will be an interesting question to address in order to understand the evolutionary trajectory of the gene in relation to its antidote function. Indeed, poison-antidote systems are often considered as selfish genetic elements, but their selfish nature is not always easy to properly establish, requiring thorough population and evolutionary genetic studies not amenable in all cases (Sweigart et al. 2019). In the present case, the question is still open and would involve the exploration of APOK3 sequence diversity in a wider sampling of natural variants, in particular in local polymorphic natural populations of A. thaliana. In addition, identifying the behavior of accessions possessing the TAC haplotype by crossing them with Mr-0 will help specifying the role of the different amino acid changes in the antidote activity of the protein.

Examination of the organization and gene content of the PK3 locus in 10 genotypes in addition to Col-0, Sha, and Mr-0 revealed a highly variable structure. In eukaryotes, poison-antidote elements are generally found within regions of high divergence and structural variation associated with low recombination (Burga et al. 2020). In our case, the structural differences between Sha and Mr-0 alleles easily explain the scarcity of recombinants found inside the PK3 interval during fine mapping of the locus. Using these rare recombinants and derived crosses allowed us to establish that the antidote gene is flanked by 2 intervals, each carrying elements necessary for the killer activity. Poison-antidote systems whose genetic factors have been identified so far in eukaryotes other than fungi most commonly involve 2 components (Burga et al. 2020). However, in plants, 3 components poison-antidote systems have been previously described in rice distorter loci. In rice S1, 3 linked genes are

necessary for toxicity (Xie *et al.* 2019). In rice S5, in addition to the protector, 2 sporophytic factors, not carried by the same allele, are necessary for the killer activity (Yang *et al.* 2012). It is noticeable that the structure of the PK3 locus favors the poison-antidote system by preventing its split apart at meiotic recombination ("stick together or die," McLaughlin and Malik 2017).

The PK3 locus is particularly dynamic and prone to structural variations. Beside the sensitive Ita-0 allele that is structurally similar to that of *A. lyrata*, the other sensitive and the resistant alleles (excepted Bur-0) present the same organization, even if some protein coding genes are missing in Are-10 and Cvi-0. Among the resistant alleles, Bur-0 is an interesting exception since it resembles a killer both in structure and gene content. We therefore hypothesize that it evolved from a killer allele that lost its killer activity. Notably, the Bur-0 locus is nearly identical to that of the killer Jea: only the *APOK3-like* copy inserted into *AT3G62610* in the PK3C interval of all the killers is missing in Bur-0. This makes this *APOK3-like* copy, which is absent in all the non killer alleles, a gene to be tested for killer activity, although it cannot be excluded that Bur-0 lost its killer activity due to another polymorphism in PK3C or in its unknown PK3A killer element. A more thorough comparison between the Bur-0 and Jea sequences at PK3 will help identifying killer elements. The most complex alleles are found in killer genotypes, which are highly variable, with different groups of genes duplicated in one or the other orientation and in different relative positions. Having more than one copy of APOK3 is one remarkable common feature of killer alleles. It is likely that having multiple copies is important for killer alleles not to be the victims of their own activity. Gene dosage has been reported to be critical in a poison-antidote system in *C. elegans* (Seidel *et al.* 2008, 2011). In addition, the different copies of APOK3 found in each killer are identical, but may be slightly different between accessions, indicating that recent duplications occurred independently in the respective lineages of the different accessions we examined. These results suggest that killer alleles have experienced a more intense structural evolution than non killer ones, and raise the questions of the evolutionary mechanisms and forces leading to these structures. The presence of several TEs at the locus, and their variation in occurrence between Sha, Mr-0, and Col-0 could be relevant to this question, since TE mobility is a main source of genetic variation in *A. thaliana* (Baduel *et al.* 2021). It has also been suggested that the proximity of transposons has facilitated the duplication of the fission yeast *wtf* genes (Eickbush *et al.* 2019).

Even if gamete killers likely exist in many plant species, none had been investigated in *A. thaliana* until now, to our knowledge. The pollen killer we dissected here has some general features of eukaryotic poison-antidote systems, including its species-specific nature and its presence within a hyper variable locus. The layout of the killer alleles is particular, with diverse duplications of sequence blocks that contain antidote genes trapped between killer elements. Continuing to exploit the natural variation of the species should help identifying the killer elements, and provide clues toward the underlying mechanisms responsible for the pollen killer activity, the role of the mitochondria, and eventually the forces driving the evolution of the locus.

Data availability

The 10 newly assembled genome sequences and the sequences of Sha and Mr-0 PK3 loci are available at the doi listed in Supplementary Table 4. Strains and plasmids are available upon request. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

Supplemental material is available at *GENETICS* online.

Conflicts of interest

None declared.

Literature cited

Agorio A, Durand S, Fiume E, Brousse C, Gy I, Simon M, Anava S, Rechavi O, Loudet O, Camilleri C, *et al.* An *Arabidopsis* natural epiallele maintained by a feed-forward silencing loop between histone and DNA. PLoS Genet. 2017;13(1):e1006551.

Agren JA, Clark AG. Selfish genetic elements. PLoS Genet. 2018; 14(11):e1007700.

Alexander MP. Differential staining of aborted and nonaborted pollen. Stain Technol. 1969;44(3):117–122.

Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, von Nielsen H. Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance. 2019;2:e201900429.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25(17):3389–3402.

Andrade MA, Bork P. HEAT repeats in the Huntington's disease protein. Nat Genet. 1995;11(2):115–116.

Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P. Comparison of ARM and HEAT protein repeats. J Mol Biol. 2001; 309(1):1–18.

Baduel P, Leduque B, Ignace A, Gy I, Gil J, Loudet O, Colot V, Quadrana L, Jr. Genetic and environmental modulation of transposition shapes the evolutionary potential of *Arabidopsis thaliana*. Genome Biol. 2021;22(1):138.

Bikard D, Patel D, Le Metté C, Giorgi V, Camilleri C, Bennett MJ, Loudet O. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. Science. 2009;323(5914):623–626.

Bolte S, Cordelieres FP. A guided tour into subcellular colocalization analysis in light microscopy. J Microsc. 2006;224(Pt 3):213–232.

Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. PLoS Biol. 2007;5(9):e236.

Bravo Nunez MA, Nuckolls NL, Zanders SE. Genetic villains: killer meiotic drivers. Trends Genet. 2018;34(6):424–433.

Burga A, Ben-David E, Kruglyak L. Toxin-antidote elements across the tree of life. Annu Rev Genet. 2020;54:387–415.

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008;36(Web Server issue):W465–W469.

Durand S, Bouche N, Perez Strand E, Loudet O, Camilleri C. Rapid establishment of genetic incompatibility through natural epigenetic variation. Curr Biol. 2012;22(4):326–331.

Durand S, Ricou A, Simon M, Dehaene N, Budar F, Camilleri C. A restorer-of-fertility-like pentatricopeptide repeat protein promotes cytoplasmic male sterility in *Arabidopsis thaliana*. Plant J. 2021;105(1):124–135.

Eickbush MT, Young JM, Zanders SE. Killer meiotic drive and dynamic evolution of the *wtf* gene family. Mol Biol Evol. 2019;36(6):1201–1214.

Finseth FR, Dong Y, Saunders A, Fishman L. Duplication and adaptive evolution of a key centromeric protein in mimulus, a genus with female meiotic drive. Mol Biol Evol. 2015;32(10):2694–2706.

Finseth FR, Nelson TC, Fishman L. Selfish chromosomal drive shapes recent centromeric histone evolution in monkeyflowers. PLoS Genet. 2021;17(4):e1009418.

Fishman L, McIntosh M. Standard deviations: the biological bases of transmission ratio distortion. Annu Rev Genet. 2019;53:347–372.

Fishman L, Sweigart AL. When two rights make a wrong: the evolutionary genetics of plant hybrid incompatibilities. Annu Rev Plant Biol. 2018;69:707–731.

Fulgione A, Koornneef M, Roux F, Hermisson J, Hancock AM. Madeiran *Arabidopsis thaliana* reveals ancient long-range colonization and clarifies demography in Eurasia. Mol Biol Evol. 2018;35(3):564–574.

Gobron N, Waszczak C, Simon M, Hiard S, Boivin S, Charif D, Ducamp A, Wenes E, Budar F. A cryptic cytoplasmic male sterility unveils a possible gynodioecious past for *Arabidopsis thaliana*. PLoS ONE. 2013;8(4):e62450.

Grognet P, Lalucque H, Malagnac F, Silar P. Genes that bias Mendelian segregation. PLoS Genet. 2014;10(5):e1004387.

Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH. Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. Plant Cell. 2004;16(1):241–256.

Houben A. B chromosomes—a matter of chromosome drive. Front Plant Sci. 2017;8:210.

Jiao W-B, Patel V, Klasen J, Liu F, Pecinkova P, Ferrand M, Gy I, Camilleri C, Effgen S, Koornneef M, *et al.* The evolutionary dynamics of genetic incompatibilities introduced by duplicated genes in *Arabidopsis thaliana*. Mol Biol Evol. 2021;38(4):1225–1240.

Klodmann J, Senkler M, Rode C, Braun HP. Defining the protein complex proteome of plant mitochondria. Plant Physiol. 2011;157(2):587–598.

Koide Y, Ogino A, Yoshikawa T, Kitashima Y, Saito N, Kanaoka Y, Onishi K, Yoshitake Y, Tsukiyama T, Saito H, *et al.* Lineage-specific gene acquisition or loss is involved in interspecific hybrid sterility in rice. Proc Natl Acad Sci U S A. 2018;115(9):E1955–E1962.

Koide Y, Shinya Y, Ikenaga M, Sawamura N, Matsubara K, Onishi K, Kanazawa A, Sano Y. Complex genetic nature of sex-independent transmission ratio distortion in Asian rice species: the involvement of unlinked modifiers and sex-specific mechanisms. Heredity (Edinb). 2012;108(3):242–247.)

Konig AC, Hartl M, Boersema PJ, Mann M, Finkemeier I. The mitochondrial lysine acetylome of *Arabidopsis*. Mitochondrion. 2014;19(Pt B):252–260.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–736.

Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, Kim M, Rosen BD, Cheng C-Y, Moreira W, Mock SA, *et al.* Araport: the *Arabidopsis* information portal. Nucleic Acids Res. 2015;43(Database issue):D1003–1009.

Larracuente AM, Presgraves DC. The selfish segregation distorter gene complex of *Drosophila melanogaster*. Genetics. 2012;192(1):33–53.

Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 2021;49(D1):D458–D460.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–1760.

Lindholm AK, Dyer KA, Firman RC, Fishman L, Forstmeier W, Holman L, Johannesson H, Knief U, Kokko H, Larracuente AM, *et al.* The ecology and evolutionary dynamics of meiotic drive. Trends Ecol Evol. 2016;31(4):315–326.

Liu B, Li JQ, Liu XD, Shahid MQ, Shi LG, Lu YG. Identification of neutral genes at pollen sterility loci *Sd* and *Se* of cultivated rice (Oryza sativa) with wild rice (O. rufipogon) origin. Genet Mol Res. 2011;10(4):3435–3445. 10.4238/2011.October.31.10 22057998

Long Y, Zhao L, Niu B, Su J, Wu H, Chen Y, Zhang Q, Guo J, Zhuang C, Mei M, *et al.* Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. Proc Natl Acad Sci U S A. 2008;105(48):18871–18876.

Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F. Bay-0 x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. Theor Appl Genet. 2002;104(6–7):1173–1184.

Matsubara K, Ebana K, Mizubayashi T, Itoh S, Ando T, Nonoue Y, Ono N, Shibaya T, Ogiso E, Hori K, *et al.* Relationship between transmission ratio distortion and genetic divergence in intraspecific rice crosses. Mol Genet Genomics. 2011;286(5–6):307–319.

Mayjonade B, Gouzy J, Donnadieu C, Pouilly N, Marande W, Callot C, Langlade N, Muños S. Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. Biotechniques. 2016;61(4):203–205.

McLaughlin RN, Jr, Malik HS. Genetic conflicts: the usual suspects and beyond. J Exp Biol. 2017;220(Pt 1):6–17.

Morineau C, Bellec Y, Tellier F, Gissot L, Kelemen Z, Nogué F, Faure J-D. Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid *Camelina sativa*. Plant Biotechnol J. 2017;15(6):729–739.

Nelson BK, Cai X, Nebenfuhr A. A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants. Plant J. 2007;51(6):1126–1136.

Nuckolls NL, Bravo Núñez MA, Eickbush MT, Young JM, Lange JJ, Yu JS, Smith GR, Jaspersen SL, Malik HS, Zanders SE, *et al. wtf* genes are prolific dual poison-antidote meiotic drivers. eLife. 2017;6:e26033.

Nuckolls NL, Mok AC, Lange JJ, Yi K, Kandola TS, Hunn AM, McCroskey S, Snyder JL, Bravo Núñez MA, McClain M, *et al.* The *wtf4* meiotic driver utilizes controlled protein aggregation to generate selective cell death. eLife. 2020;9:e55694.

Östergren G. Parasitic nature of extra fragment chromosomes. Bot Not. 1945;2:157–163.

Ouyang Y, Zhang Q. Understanding reproductive isolation based on the rice model. Annu Rev Plant Biol. 2013;64:111–135.

Ouyang Y, Zhang Q. The molecular and evolutionary basis of reproductive isolation in plants. J Genet Genomics. 2018;45(11): 613–620.

Presgraves DC. The molecular evolutionary basis of species formation. Nat Rev Genet. 2010;11(3):175–180.

Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. Combined Evidence Annotation of Transposable Elements in Genome Sequences. PLoS Comp Biol. 2005;1(2):e22 10.1371/journal.pcbi.0010022

Rick CM. Abortion of male and female gametes in the tomato determined by allelic interaction. Genetics. 1966;53(1):85–96.

Rick CM. The tomato ge locus: linkage relations and geographic distribution of alleles. Genetics. 1971;67(1):75–85.

Ross KJ, Fransz P, Jones GH. A light microscopic atlas of meiosis in *Arabidopsis thaliana*. Chromosome Res. 1996;4(7):507–516.

Russo A, Potente G, Mayjonade B. HMW DNA extraction from diverse plants species for PacBio and Nanopore sequencing. *protocols.io*; 2021.

Sallet E, Gouzy J, Schiex T. EuGene: an automated integrative gene finder for eukaryotes and prokaryotes. Methods Mol Biol. 2019; 1962:97–120.

Salomé PA, Bomblies K, Fitz J, Laitinen RAE, Warthmann N, Yant L, Weigel D. The recombination landscape in *Arabidopsis* thaliana F2 populations. Heredity (Edinb). 2012;108(4):447–455.

Samuel MA, Salt JN, Shiu SH, Goring DR. Multifunctional arm repeat domains in plants. Int Rev Cytol. 2006;253:1–26.

Sarrion-Perdigones A, Vazquez-Vilar M, Palaci J, Castelijns B, Forment J, Ziarsolo P, Blanca J, Granell A, Orzaez D. GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. PLANT PHYSIOLOGY. 2013;162(3):1618–1631. 10.1104/pp.113.217661

Seidel HS, Ailion M, Li J, van Oudenaarden A, Rockman MV, Kruglyak L. A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. PLoS Biol. 2011; 9(7):e1001115.

Seidel HS, Rockman MV, Kruglyak L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. Science. 2008;319(5863):589–594.

Senkler J, Senkler M, Eubel H, Hildebrandt T, Lengwenus C, Schertl P, Schwarzländer M, Wagner S, Wittig I, Braun H-P, *et al.* The mitochondrial complexome of *Arabidopsis thaliana*. Plant J. 2017;89(6): 1079–1092.

Seymour DK, Chae E, Arioz BI, Koenig D, Weigel D. Transmission ratio distortion is frequent in *Arabidopsis thaliana* controlled crosses. Heredity (Edinb). 2019;122(3):294–304.

Simon M, Durand S, Pluta N, Gobron N, Botran L, Ricou A, Camilleri C, Budar F. Genomic conflicts that cause pollen mortality and raise reproductive barriers in *Arabidopsis thaliana*. Genetics. 2016; 203(3):1353–1367.

Simon M, Simon A, Martins F, Botran L, Tisné S, Granier F, Loudet O, Camilleri C. DNA fingerprinting and new tools for fine-scale discrimination of *Arabidopsis thaliana* accessions. Plant J. 2012;69(6): 1094–1101.

Smith LM, Bomblies K, Weigel D. Complex evolutionary events at a tandem cluster of *Arabidopsis thaliana* genes resulting in a single-locus genetic incompatibility. PLoS Genet. 2011;7(7):e1002164.

Sweigart AL, Brandvain Y, Fishman L. Making a murderer: the evolutionary framing of hybrid Gamete-Killers. Trends Genet. 2019; 35(4):245–252.

Tan YF, O'Toole N, Taylor NL, Millar AH. Divalent metal ions in plant mitochondria and their role in interactions with proteins and oxidative stress-induced damage to respiratory function. Plant Physiol. 2010;152(2):747–761.

Tang H, Zhang X, Miao C, Zhang J, Ming R, Schnable JC, Schnable PS, Lyons E, Lu J. ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biol. 2015;16:3.

Taylor NL, Heazlewood JL, Millar AH. The *Arabidopsis thaliana* 2-D gel mitochondrial proteome: refining the value of reference maps for assessing protein abundance, contaminants and post-translational modifications. Proteomics. 2011;11(9):1720–1733.

The 1001 Genome Consortium. 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. Cell. 2016;166:481–491.

Turner BC. Geographic distribution of neurospora spore killer strains and strains resistant to killing. Fungal Genet Biol. 2001;32(2): 93–104.

Vaid N, Laitinen RAE. Diverse paths to hybrid incompatibility in *Arabidopsis*. Plant J. 2019;97(1):199–213.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11):e112963.

Wang GW, He YQ, Xu CG, Zhang Q. Identification and confirmation of three neutral alleles conferring wide compatibility in inter-subspecific hybrids of rice (*Oryza sativa* L.) using near-isogenic lines. Theor Appl Genet. 2005;111(4):702–710.

Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 2005; 21(9):1859–1875.

Xie Y, Tang J, Xie X, Li X, Huang J, Fei Y, Han J, Chen S, Tang H, Zhao X, *et al.* An asymmetric allelic interaction drives allele transmission bias in interspecific rice hybrids. Nat Commun. 2019;10(1):2501.

Yang J, Zhao X, Cheng K, Du H, Ouyang Y, Chen J, Qiu S, Huang J, Jiang Y, Jiang L, *et al.* A killer-protector system regulates both hybrid sterility and segregation distortion in rice. Science. 2012; 337(6100):1336–1340.

Yu Y, Zhao Z, Shi Y, Tian H, Liu L, Bian X, Xu Y, Zheng X, Gan L, Shen Y, *et al.* Hybrid sterility in rice (*Oryza sativa* L.) involves the tetratricopeptide repeat domain containing protein. Genetics. 2016; 203(3):1439–1451.

Yu X, Zhao Z, Zheng X, Zhou J, Kong W, Wang P, Bai W, Zheng H, Zhang H, Li J, *et al.* A selfish genetic element confers non-Mendelian inheritance in rice. Science. 2018;360(6393):1130–1132.

Zhang C, Wang D, Wang J, Sun Q, Tian L, Tang X, Yuan Z, He H, Yu S. Genetic dissection and validation of chromosomal regions for transmission ratio distortion in intersubspecific crosses of. Front Plant Sci. 2020;11:563548.

Zhang C, Wang J, Marowsky NC, Long M, Wing RA, Fan C. High occurrence of functional new chimeric genes in survey of rice chromosome 3 short arm genome sequences. Genome Biol Evol. 2013; 5(5):1038–1048.

Zhang Y, Swart C, Alseekh S, Scossa F, Jiang L, Obata T, Graf A, Fernie AR. The extra-pathway interactome of the TCA cycle: expected and unexpected metabolic interactions. Plant Physiol. 2018;177(3):966–979.

Communicating editor: A. Larracuente