



**HAL**  
open science

## **A role for an endosperm-localized subtilase in the control of seed size in legumes**

Isabelle d'Erfurth, Christine Le Signor, Gregoire Aubert, Myriam Sanchez, Vanessa Vernoud, Brigitte Darchy, Jeannine Lherminier, Virginie Bourion, Nathalie Bouteiller, Abdelhafid Bendahmane, et al.

### ► **To cite this version:**

Isabelle d'Erfurth, Christine Le Signor, Gregoire Aubert, Myriam Sanchez, Vanessa Vernoud, et al.. A role for an endosperm-localized subtilase in the control of seed size in legumes. *New Phytologist*, 2012, 196 (3), pp.738-751. 10.1111/j.1469-8137.2012.04296.x . hal-01209890

**HAL Id: hal-01209890**

**<https://hal.science/hal-01209890v1>**

Submitted on 29 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A role for an endosperm-localized subtilase in the control of seed size in legumes

I. D'Erfurth<sup>1\*</sup>, C. Le Signor<sup>1\*</sup>, G. Aubert<sup>1</sup>, M. Sanchez<sup>1</sup>, V. Vernoud<sup>1</sup>, B. Darchy<sup>1</sup>, J. Lherminier<sup>1</sup>, V. Bourion<sup>1</sup>, N. Bouteiller<sup>2</sup>, A. Bendahmane<sup>2</sup>, J. Buitink<sup>3</sup>, J. M. Prosperi<sup>4</sup>, R. Thompson<sup>1</sup>, J. Burstin<sup>1</sup> and K. Gallardo<sup>1</sup>

<sup>1</sup>INRA (Institut National de la Recherche Agronomique), UMR1347 Agroécologie, BP 86510, F-21000, Dijon, France; <sup>2</sup>INRA/CNRS (Centre National de la Recherche Scientifique), Unité de Recherche en Génomique Végétale, CP5708, 91057, Evry, France; <sup>3</sup>INRA, UMR1345 Institut de Recherche en Horticulture et Semences, SFR 4207 QUASAV, 49045, Angers, France; <sup>4</sup>INRA, UMR1334 Amélioration Génétique et Adaptation des Plantes, 34060, Montpellier, France

## Summary

Author for correspondence:

Karine Gallardo

Tel: +33 (0)3 80 69 33 91

Email: gallardo@dijon.inra.fr

Received: 20 July 2012

Accepted: 25 July 2012

*New Phytologist* (2012) **196**: 738–751

doi: 10.1111/j.1469-8137.2012.04296.x

**Key words:** endosperm, *Medicago truncatula* (barrel medic), *Pisum sativum* (pea), seed size, subtilase, TILLING.

- Here, we report a subtilase gene (*SBT1.1*) specifically expressed in the endosperm of *Medicago truncatula* and *Pisum sativum* seeds during development, which is located at a chromosomal position coinciding with a seed weight quantitative trait locus (QTL).
- Association studies between *SBT1.1* polymorphisms and seed weights in ecotype collections provided further evidence for linkage disequilibrium between the *SBT1.1* locus and a seed weight locus. To investigate the possible contribution of *SBT1.1* to the control of seed weight, a search for TILLING (Targeting Induced Local Lesions in Genomes) mutants was performed.
- An inspection of seed phenotype revealed a decreased weight and area of the *sbt1.1* mutant seeds, thus inferring a role of *SBT1.1* in the control of seed size in the forage and grain legume species. Microscopic analyses of the embryo, representing the major part of the seed, revealed a reduced number of cells in the *MtP330S* mutant, but no significant variation in cell size.
- *SBT1.1* is therefore most likely to be involved in the control of cotyledon cell number, rather than cell expansion, during seed development. This raises the hypothesis of a role of *SBT1.1* in the regulation of seed size by providing molecules that can act as signals to control cell division within the embryo.

## Introduction

Subtilisin-like proteases (subtilases, SBTs) are serine proteases characterized by an aspartate, histidine and serine catalytic triad. The first SBT was isolated from *Bacillus subtilis* (Strongin *et al.*, 1978). In plants, SBTs belong to multigene families: 56 SBT genes have been identified in the *Arabidopsis* genome and have been divided into six distinct subfamilies (Rautengarten *et al.*, 2005). These SBTs have been shown to control diverse developmental processes. For example, SDD1 (STOMATAL DENSITY AND DISTRIBUTION 1) controls stomatal distribution and density (Berger & Altmann, 2000; Von Groll *et al.*, 2002) and the overexpression of *AtSBT5.4* produces a clavata-like, loss-of-function phenotype with profound defects in the inflorescence stem and shoot apex (Liu *et al.*, 2009). Some SBTs exhibit broad substrate specificity and are predicted to be involved in nonselective protein turnover, whereas others exhibit much more stringent substrate requirements (Schaller *et al.*, 2012; and references therein). For example, the AMON (amontillado) SBT of *Drosophila*, required for pupal development, has been proposed to process and activate a diverse suite of peptide hormones

(Rayburn *et al.*, 2009). The ability of some plant SBTs to process specific substrates (Berger & Altmann, 2000; Vartapetian *et al.*, 2011) suggests that, as in *Drosophila* and animals, plant SBTs act in development via the cleavage of pro-hormones leading to the activation of peptide hormones. For instance, the SBT ATSBT1.1 of *Arabidopsis* processes a pre-pro-peptide of phytosulfokine (PSK) to the active peptide hormone (Srivastava *et al.*, 2008).

In *Arabidopsis*, soybean, barley and rice, several SBTs are highly expressed in distinct seed tissues, suggesting different roles in seed development and germination (Batchelor *et al.*, 2000; Yamagata *et al.*, 2000; Tanaka *et al.*, 2001; Beilinson *et al.*, 2002; Fontanini & Jones, 2002; Rautengarten *et al.*, 2008; Yang *et al.*, 2008). For example, *AtSBT1.7* is implicated in the release of mucilage from the seed coat during rehydration (Rautengarten *et al.*, 2008) and *ALE1* (abnormal leaf shape 1) controls embryo cuticle formation in *Arabidopsis* (Tanaka *et al.*, 2001; Yang *et al.*, 2008). Moreover, in a previous study of proteome and transcriptome changes during seed development in the model legume *Medicago truncatula*, two SBT genes were identified, one being specifically expressed in the seed coat during seed filling and the other in the endosperm at the pre-storage phase (Gallardo *et al.*, 2007). Because seeds of legumes, such as pea and soybean, are a rich source of proteins for animal and human nutrition, the identification of the molecular mechanisms

\*These authors contributed equally to this work.

regulating seed development is important for strategies of improvement of seed quality and yield. In this article, after showing that the locus harboring the endosperm-specific SBT from Gallardo *et al.* (2007) influences seed weight in *M. truncatula* and in pea (*Pisum sativum*), we identified ethyl methanesulfonate (EMS)-induced single nucleotide polymorphisms (SNPs) in the TILLING (Targeting Induced Local Lesions in Genomes) mutant populations of *M. truncatula* (Le Signor *et al.*, 2009) and pea (Dalmais *et al.*, 2008) to test the possible contribution of this SBT in the control of seed weight. The phenotypic characterization of the mutant seeds revealed a role for this endosperm-localized protease in the control of seed size in the wild legume and the cultivated grain legume species, which is likely to occur by modulation of cell division within the embryo.

## Materials and Methods

### Plant growth conditions

The barrel medic (*Medicago truncatula* Gaertn.) and pea (*Pisum sativum* L.) plants were grown under glasshouse conditions in pots filled with pouzzolane (inert medium, light volcanic grit). Pea plants were grown in 7-l pots with three plants per pot and *M. truncatula* plants were grown in 1.5-l pots with one plant per pot. Temperature was controlled to be 30°C during the day for both species and above 19 and 14°C during the night for *M. truncatula* and pea, respectively. Artificial lighting was supplied to reach 16 h light per day. The plants were not inoculated with *Sinorhizobium* sp. bacteria and nitrogen supply was not limited: plants were automatically supplied with 3.5N/3.1P/8.6K.

### Isolation of *MtSBT1.1* and *PsSBT1.1* sequences, phylogenetic analysis and mapping

*Medicago truncatula* sequences were retrieved from the Mt3.5 version of the *M. truncatula* genome. Phylogenetic analyses were performed using MEGA 5.05 (Tamura *et al.*, 2011). The phylogenetic tree was built using the neighbor-joining method with 1000 bootstrap generations. Multiple alignments were realized using Clustalw software. Mapping of *SBT1.1* on the LR4 genetic map of *M. truncatula* was carried out using a cleaved amplified polymorphic sequence (CAPS) marker, polymorphic between the two parental lines Jemalong 6 and DZA315.16. The SBT fragment was amplified (melting temperature of 55°C) from genomic DNA using the primers 5'-GTTCCAAGGTATTCAAGCCTACCC-3' (forward) and 5'-CCTAGAGCTCTCTCCACGATCAC-3' (reverse), and digested by *Cla*I overnight at 37°C. *PsSBT1.1* has been sequenced and mapped as 'Subt' using an SNP marker from the Cameor × China mapping population (Deulvot *et al.*, 2010). The *PsSBT1.1* sequence has been deposited in GenBank under accession number JX402205.

### Detection of *SBT1.1* polymorphism in natural collections

Four seeds of each of the 346 genotypes of the *M. truncatula* core collection (INRA Medicago Stock Center; <http://www.montpellier.inra.fr/BRC-MTR/>) were sown. DNA was extracted from 2-wk-old plantlets using the Plant DNeasy 96 kit (QIAGEN, Courtaboeuf, France) following the manufacturer's instructions. DNA pools with a 1 : 1 weight ratio of genotypes vs reference line A17 were made. Amplicons for the 346 pools of the *MtSBT1.1* gene were obtained by nested PCR with the same primers as for the TILLING experiment (see the following section). After endonuclease digestion, DNA fragments were separated on acrylamide gels and both images (700 and 800 nm) were acquired using a 4300 DNA analyzer (LiCOR, Lincoln, NE, USA). Polymorphic sites were recorded for each genotype. A sample of amplicons representative of each polymorphism was sequenced (Millegen, Toulouse, France). The mean seed weight of ecotypes from the INRA Medicago Stock Center was recorded from glasshouse productions in the years 2002 (346 genotypes) and 2011 (192 genotypes). Polymorphism in the *P. sativum* core collection (373 accessions) was revealed by a 384-SNP set of an Illumina GoldenGate assay (Deulvot *et al.*, 2010). Two of these SNPs were located within the *PsSBT1.1* gene and are analyzed in this article. Pea seed weight data of field trials from the years 2003 and 2007 were available for association analyses. The structure of the *M. truncatula* collection based on molecular markers is from Ronfort *et al.* (2006), and the structure of the pea collection was based on 28 simple sequence repeat markers. For both species, analysis of variance of seed traits was performed using a model with two factors, structure + SNP, with sequential-type sum of squares (SAS, 1999).

Individual flowers were tagged and seeds were collected according to Gallardo *et al.* (2007). Seed coat, endosperm and embryo of the freshly harvested seeds were manually separated at 4°C under a magnifying glass (×3.5). All seed and tissue samples were ground in liquid nitrogen using a mortar and pestle. The powder was stored at -80°C until mRNA extraction and quantitative reverse transcription-polymerase chain reaction (RT-PCR) were performed according to Gallardo *et al.* (2007). Normalization of relative cDNA quantity was performed for each template using the *msc27* and *histone H1* gene controls for *M. truncatula* and pea seed samples, respectively, according to the relative standard curve method ( $\Delta C_T$ ). The expression stability of the control genes in the different samples was verified by comparison with two other constitutively expressed genes (Gallardo *et al.*, 2007). In addition, *PsSBT1.1* mRNA abundance in 12 plant organs (including developing and germinating seeds) at five developmental stages and under two nutrition conditions was estimated from RNA-Seq data. Details on the methods used for plant sample preparation, RNA extraction, cDNA library construction and sequencing are provided in Supporting Information Table S1.

### Transcript abundance determination

*In situ* hybridization

Freshly harvested *M. truncatula* seeds (at 10 and 12 d after pollination (dap)) were fixed in 4% paraformaldehyde. A 650-bp probe was defined as specific for the *MtSBT1.1* sequence at the

3' end of the gene (340 bp in 3' cDNA and 310 bp in 3' untranslated region (UTR)) with the forward primer 5'-GTTC TACTACTGCATTTCTCACATAAC-3' and the reverse primer 5'-AAATATGGGCATCTGCCACGG-3'. Sense and antisense probes were first amplified using T3 polymerase oligonucleotides with a 5' transcription T3 promoter fusion for template production by PCR, and then synthesized using the T3 polymerase. The method for digoxigenin labeling of RNA probes, tissue preparation and *in situ* hybridization was described by Coen *et al.* (1990) with modifications described by Bradley *et al.* (1993).

### TILLING of *SBT1.1*

Pea mutants were identified from an EMS population containing 4800 M2 lines from the cultivar Cameor by TILLING according to Dalmais *et al.* (2008). PCR amplification was based on nested PCR and universal primers. A first fragment of 1 kb was amplified and served as template for the second PCR (annealing temperature 58°C) with primers PsSbt2Ftag 5'-CCAATCAGCTGAAATGC-3' and PsSbt2Rtag2 5'-GTTCCATCACCTTTTCC-3' carrying a universal M13 tail (5'-CACGACGTTGTAACACGAC-3' for the forward primer; 5'-GGATAACAATTTTCACACAGG-3' for the reverse primer) labeled at the 5' end with infra-red dyes IRD700 and IRD800 (LiCOR). To confirm mutations, PCR amplification products (978 bp) were sequenced (GATC Biotech, Konstanz, Germany) and analyzed (Chromas v.1.4523 software, Technelysium Pty., South Brisbane, Australia). Twenty-five missense mutations were revealed. *Medicago truncatula* mutants were identified from an EMS population containing 4600 M2 lines (cultivar Jemalong A17) by TILLING according to Le Signor *et al.* (2009). TILLING was performed from two amplicons of *c.* 1 kb to cover the entire coding sequence. Nested PCR was conducted using the following inner primers labelled with IRD-700 and IRD-800 dyes: MtSBT-F2 5'-CTGCATACATTAGTCTTGAAATGG-3' and MtSBT-R2 5'-CTTACAACAGTTTTTCCGTCAGACC-3'. Twelve missense mutants were identified and sequenced (MilleGen, Toulouse, France). For both species, because of the large number of alleles identified, the possible impact of the missense mutations on protein function was assessed using SIFT software (Sorting Intolerant From Tolerant; Ng & Henikoff, 2003). The mutants retained were backcrossed twice with the wild-type lines (genotypes Cameor in pea and Jemalong A17 in *M. truncatula*). All progenies BC1, BC2, BC1S1 and BC2S1 (selfing of BC1 or BC2) were genotyped with specific dCaps markers designed according to Neff *et al.* (1998). The primers, PCR conditions and enzymes are available in Table S2.

### Seed trait phenotyping

The plants analyzed were progenies issued from one to three generations of selfing (S1, S2 or S3) of one backcross (BC1S2) or two backcrosses (BC2S2, BC2S3) of the EMS mutant lines with the wild-type A17 (*M. truncatula*) or Cameor (*P. sativum*) lines. For each mutation, comparisons were made between the seeds from the wild-type and mutant plants (four to seven plants of each genotype

were studied) coming from the same progeny in two independent experiments (2009 and 2010). Pods were harvested at maturity, weighed and then manually threshed. The number of seeds per pod was recorded in both species for a sample of 30 pods, and each seed was weighed individually. To estimate the seed surface area, seed samples from four to six plants per genotype for *M. truncatula* and seven plants per genotype for pea were scanned as digital images with an A3 scanner (Epson, Tokyo, Japan). For *M. truncatula*, a sample of 100 seeds per plant was scanned as color images. For pea, all samples of 5–10 seeds per plant were scanned using the transparency unit to avoid the above shadow effect and to produce high-quality contrasted images of these thick seeds. The average seed projected area (two-dimensional area measurement by projecting the seed shape on to an arbitrary plane), seed length for *M. truncatula* and seed diameter for pea were further determined by image analysis based on thresholding to create binary images using ImageJ software (Wayne Rasband, National Institute of Mental Health, Bethesda, MD, USA). All data were subjected to variance analysis (genotype effect) using the SAS system (Cary, NC, USA).

### Light and transmission electron microscopy

Mature and 16-dap seeds from the wild-type and *MtP330S* *M. truncatula* genotypes were vacuum infiltrated overnight at 4°C with a fixative mixture containing 3% (v/v) glutaraldehyde and 2% (w/v) paraformaldehyde in 0.1 M sodium phosphate-buffered medium (pH 7.2). After washing, seeds were fixed in 0.5% OsO<sub>4</sub> solution in phosphate buffer for 1 h, washed again and dehydrated in ethanol and propylene oxide before embedding in Epon (Spi-Chem, Neyco, Paris, France) according to the standard procedure (Luft, 1961). For fixation, a sample of 16-dap seeds was washed and dehydrated in ethanol before embedding in Histo-resin (Leica, Rueil-Malmaison, France) following the manufacturer's instructions. Thick sections (0.5 µm for Epon-embedded seeds and 1 µm for Histo-resin-embedded seeds) were cut on a Reichert ultramicrotome (Leica), mounted on glass slides and stained with 0.1% (w/v) toluidine blue plus 0.5% (w/v) methylene blue, pH 9, before examination by bright field microscopy with a DMRB microscope (Leica). At least three biological replicates of *MtP330S* and wild-type seeds at the 16-dap and mature stages were analyzed. Twenty to fifty images from independent sections per seed were acquired with a Sony 3CCD color camera driven by Visilog 6 (Noesis, Gif-sur-Yvette, France). Areas were measured with ImageJ (Rasband, 1997–2008).

The ultrastructures of the cotyledons of mature and 16-dap seeds from the wild-type and *MtP330S* alleles were studied by transmission electron microscopy. Ultrathin sections (80 nm) were collected on grids and counterstained with 3% (w/v) uranyl acetate in ethanol and lead citrate. Sections were examined with a Hitachi H7500 transmission electron microscope (Hitachi Scientific Instruments Co., Tokyo, Japan) operating at 80 kV and equipped with an AMT camera driven by AMT software (AMT, Danvers, MA, USA). Two mature seeds from two independent plants per genotype were observed (wild-type: five grids, 198 images;

*MtP330S*: five grids, 250 images), and 16-dap seed sections were studied (wild-type: two grids, 228 images; *MtP330S*: one grid, 225 images). Statistical analysis of the data was performed using Statistica 7.0 software (StatSoft, Inc., Tulsa, Oklahoma, USA).

### Scanning electron microscopy

*Medicago truncatula* mature seeds were fixed with 3% (v/v) glutaraldehyde and 2% (v/v) paraformaldehyde in 0.1 M sodium phosphate (pH 7.2). After 2 h of imbibition, the seed coat was removed and the embryo was cut to improve fixation. The samples were dehydrated in a graded ethanol series, followed by a graded acetone series, and dried in a critical point dryer (Balzers CPD-030, Pfeiffer Vacuum GmbH, Asslar, Germany) using CO<sub>2</sub> as a transition fluid. Dried sections were mounted on clean aluminum stubs with double-sided adhesive graphite tabs. Mounted specimens were coated with gold–palladium (12–15 nm thick) using a Polaron SC 7620 Mini Sputter Coater (Quorum Technologies Ltd, Ashford, UK). Samples were observed with a Philips XL-30 ESEM LaB6 scanning electron microscope. Sections were photographed digitally using an EVO40 scanning electron microscope (Carl Zeiss, Jena, Germany). Areas were measured with ImageJ (Rasband, 1997–2008) from 9–10 seeds per genotype. Statistical analysis of the data was performed using Statistica software, version 7.0.

### Seed composition analysis, zymogram protease assay and heterologous expression of SBT1.1

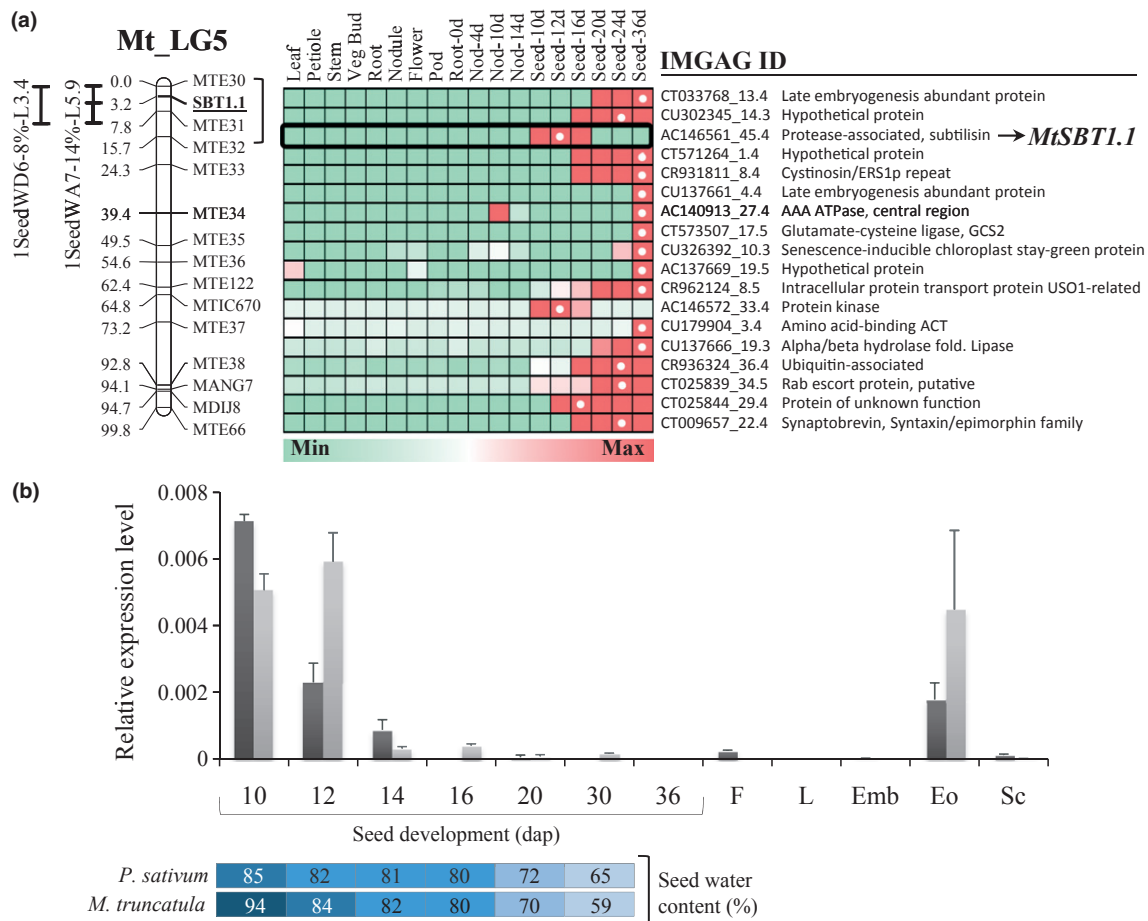
Three biological replicates of 50 mature seeds for the wild-type and *MtP330S* genotypes were ground to a fine powder. Nitrogen and carbon contents were determined twice from 5 mg of dried seed powder (encapsulated) using an elemental analyzer (NC2500 Thermo Scientific; CE Instruments, Milan, Italy). Further details on the method are provided in Supporting Information Fig. S1. To study the protein composition of mature seeds, total proteins were extracted twice in 50 µl mg<sup>-1</sup> of thiourea/urea lysis buffer according to Gallardo *et al.* (2007), and protein concentration was measured according to Bradford (1976). Protein extracts (10 µg proteins) were then separated in 12% sodium dodecyl-sulfate (SDS)-polyacrylamide gels in the presence of Laemmli buffer without heat denaturation. For the zymogram protease assay, proteins were extracted according to Wang *et al.* (2004) in a lysis buffer containing 1 M Tris-HCl, 1 mM dithiothreitol (DTT) and 1 mM EDTA. Protein extracts (20 µg proteins) were dissolved in Laemmli buffer without heat denaturation and loaded onto 12% SDS-polyacrylamide gels co-polymerized with 0.1% (w/v) gelatin. Following electrophoresis, the gels were treated according to Wang *et al.* (2004) for protease renaturation. The gels were pre-immersed in a 25% isopropanol (v/v) aqueous solution at room temperature for 30 min, and then transferred to 100 mM sodium phosphate (pH 7) or 100 mM sodium acetate (pH 5 or 6) containing a final concentration of 10 mM DTT and kept at 30°C overnight for proteolysis. Additional information on the in-gel protease activity test and details of the heterologous expression of SBT1.1 are summarized in Fig. S2.

## Results

### An SBT gene specifically expressed in seeds is associated with a seed weight locus

Seed weight measurements in the recombinant inbred line population LR4, derived from a cross between Jemalong 6 and DZA315.16, revealed a quantitative trait locus (QTL) on chromosome 5 of *M. truncatula* accounting for up to 14% of the variation in seed weight with a significant logarithm of odds (LOD) score above 3.0 (Fig. 1a and Vandecasteele *et al.*, 2011). This QTL is located between the two genetic markers MTE30 (BAC ID AC146852) and MTE32 (BAC ID CR931809) which are anchored to the physical map of *M. truncatula* (<http://www.medicago.org/>). The genes in this region, for which expression profiles were available in the Gene Expression Atlas of *M. truncatula* (Benedito *et al.*, 2008), were retrieved (753 genes) and transcript distribution in the various plant tissues was studied (Table S3). Eighteen genes that fall within the MTE30 and MTE32 interval were preferentially expressed during seed development (Fig. 1a), including a gene encoding an SBT previously shown to be localized in the endosperm of developing seeds (Gallardo *et al.*, 2007). This gene was then genetically mapped in the vicinity of the seed weight QTL. We named this gene, with accession number Medtr5g016120 (3.5 genome version; Young *et al.*, 2011), *MtSBT1.1*. This gene was a good candidate for the control of seed weight because it was the only gene in this region that was expressed during the early stages of seed development and specifically expressed in seeds relative to the other plant tissues (Fig. 1a; Benedito *et al.*, 2008). An orthologous pea gene was identified by the amplification of pea DNA by PCR, using primers based on the *M. truncatula* sequence. Sequencing of the PCR product confirmed the high similarity (94%) of the two genes. This gene (named *PsSBT1.1*) was mapped as 'Subt' (Deulvot *et al.*, 2010; Bordat *et al.*, 2011) close to the *rgp* marker from Aubert *et al.* (2006). Interestingly, this chromosomal region, which is syntenic with the *M. truncatula* region of chromosome 5 that includes the two markers MTE30 and MTE32, also harbors a seed weight QTL (Burstin *et al.*, 2007).

To further confirm that this syntenic chromosomal region may harbor a locus controlling seed weight in both species, the polymorphism of the *MtSBT1.1* gene was analyzed by EcoTILLING (Comai *et al.*, 2004) in 346 accessions of the *M. truncatula* core collection (Ronfort *et al.*, 2006). Five SNPs were recorded, two of which, at positions 598 (A598C) and 1020 (C1020T), showed a significant association with variations in seed and pod weight recorded in two independent experiments (years 2002 and 2011; Table 1). Sequencing of the amplicons for eight different accessions per SNP showed that SNP A598C induced an amino acid change at position 200 (MtI200L; Fig. 2) in the MtSBT1.1 protein sequence, whereas SNP C1020T was a synonymous change (MtL340L), which may be in linkage disequilibrium with the causal polymorphism. In pea, polymorphism data for two SNPs were available in 384 accessions of the INRA reference collection (Deulvot *et al.*, 2010), and seed weight was recorded in the field



**Fig. 1** A gene encoding a seed-specific subtilisin colocalizes with a seed weight quantitative trait locus (QTL) in *Medicago truncatula*. (a) Position of the one-seed weight QTL (detected for two independent experiments in the glasshouse: in Dijon in 2006, 1SeedWD6; in Angers in 2007, 1SeedWA7) explaining between 8 and 14% of the variation on the linkage group 5 (Mt\_LG5) of *M. truncatula* (LR4 population derived from a cross between Jemalong 6 and DZA315.16; Thoquet *et al.*, 2002), and mapping position of the subtilisin gene with BAC ID AC146561 (*SBT1.1*) in the QTL interval. The QTL labels indicate the name of the experiment, the corresponding  $r^2$  (%) followed by the logarithm of odds (LOD) scores (L). For both QTLs, the favorable allele is derived from DZA315.16. Of the 754 genes with expression profiling data available between the two genetic markers MTE30 (BAC ID AC146852) and MTE32 (BAC ID CR931809) anchored to the physical map, 18 are at least six-fold preferentially expressed in seeds (gene expression data from Benedito *et al.*, 2008). For each gene, the International Medicago Genome Annotation Group (IMGAG-Mt2-gene) is indicated. The *MtSBT1.1* gene is specifically expressed in seeds when compared with other plant organs. White dots on the heat map indicate maximum expression levels. The corresponding Affymetrix IDs are, from top to bottom: Mtr.36679.1.S1\_s\_at, Mtr.11561.1.S1\_at, Mtr.28698.1.S1\_at, Mtr.9557.1.S1\_at, Mtr.43942.1.S1\_s\_at, Mtr.2728.1.S1\_at, Mtr.9028.1.S1\_at, Mtr.35817.1.S1\_at, Mtr.2735.1.S1\_at, Mtr.13231.1.S1\_s\_at, Mtr.51717.1.S1\_at, Mtr.49623.1.S1\_at, Mtr.5494.1.S1\_at, Mtr.4879.1.S1\_at, Mtr.5806.1.S1\_s\_at, Mtr.14679.1.S1\_at, Mtr.2612.1.S1\_at, Mtr.39357.1.S1\_s\_at. (b) Relative expression levels of *SBT1.1* in *Pisum sativum* (dark grey bars) and *M. truncatula* (light grey bars) during seed development on a time scale in days after pollination (dap), in flowers (F), leaves (L) and seed tissues (embryo, Emb; endosperm, Eo; seed coat, Sc) collected at 12 dap. To compare the progress of seed development between the two species, the mean seed water content (expressed as a percentage on a fresh weight basis) from 10 to 30 dap was included as a blue scale. Error bars, + SE.

trials in the years 2003 and 2007. One of the two SNPs showed a significant association with seed weight (Table 1). This SNP is at position 612 (G612A) and is synonymous (PsK204K; Fig. 2). The association of two *SBT1.1* SNPs (C1020T in *M. truncatula* and G612A in pea) with variations in seed and/or pod weight remained highly significant after introduction in the model of the population structure effect (Table 1). These data support the evidence that the locus harboring *SBT1.1* influences seed weight. Because this region may exhibit extended linkage disequilibrium, future experiments are needed, including the exploration of the flanking regions of the *SBT1.1* gene, to specify the link between the *SBT* locus and the causal polymorphism.

### Tissue-specific expression patterns of *SBT1.1*

Expression analyses of the identified *SBT* genes, *MtSBT1.1* and *PsSBT1.1*, by quantitative RT-PCR revealed similar specific expression patterns during seed development and among seed tissues for both the *M. truncatula* and pea genes (Fig. 1b). We also examined RNA-Seq data obtained from 12 different pea organs, including leaves, stems, pods, roots, nodules, flowers, developing seeds at 12 dap and germinating seeds. This confirmed the specificity of expression of the *PsSBT1.1* gene in developing seeds, with an RPKM (reads per kilobase of exon model per million mapped reads, i.e. expression level) value of 7.9 in 12-dap seeds and

**Table 1** Mean values and standard errors of pod and seed weights from pea and *Medicago truncatula* ecotypes showing polymorphisms in *SBT1.1*

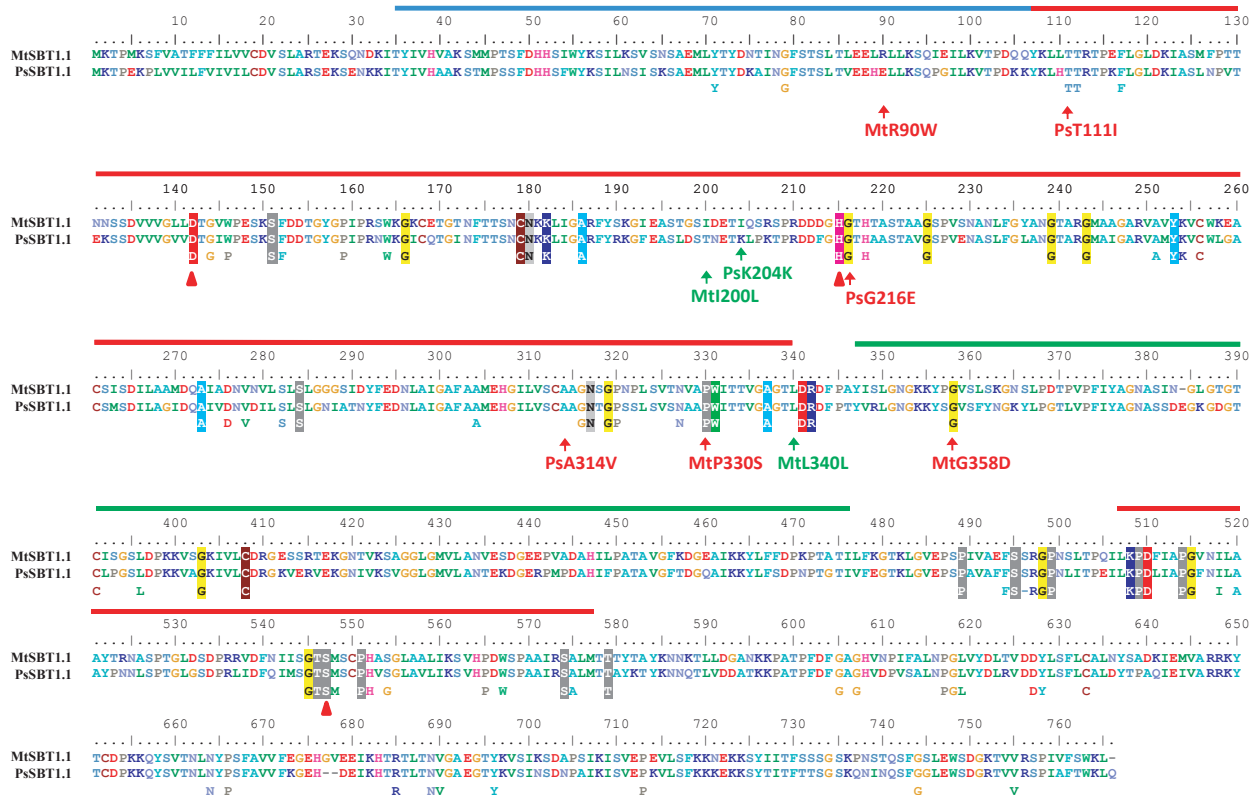
Species	SNP	AA change	Trait	Year <sup>a</sup>	SNP_0		SNP_1		SNP effect ( <i>P</i> value) <sup>c</sup>	
					Mean value	SE <sup>b</sup>	Mean value	SE <sup>b</sup>	Without structure	With structure
<i>M. truncatula</i>	A598C	I200L	Seed weight (mg)	2002	4.68	0.06	4.40	0.09	0.012	0.116
				2011	4.86	0.08	4.55	0.15	0.064	0.281
	C1020T	L340L	Pod weight (g)	2002	9.46	0.23	7.41	0.38	< 0.0001	0.039
				2011	10.4	0.32	8.36	0.58	0.002	0.070
	C1020T	L340L	Seed weight (mg)	2002	4.52	0.04	5.84	0.21	< 0.0001	0.003
				2011	4.73	0.07	5.78	0.31	0.001	0.110
C1020T	L340L	Pod weight (g)	2002	8.63	0.18	14.37	0.85	< 0.0001	< 0.0001	
			2011	9.79	0.26	14.96	1.21	< 0.0001	0.008	
<i>P. sativum</i>	G612A	K204K	Seed weight (g)	2003	0.19	0.004	0.16	0.007	0.0005	< 0.0001
				2007	0.18	0.004	0.15	0.006	0.001	< 0.0001

For each single nucleotide polymorphism (SNP) and trait analyzed, SNP\_1 refers to plants showing a polymorphism in *SBT1.1* when compared with the reference lines used for EcoTILLING (Cameor in pea and Jemalong A17 in *M. truncatula*), and SNP\_0 refers to plants with no polymorphism when compared with the reference lines. AA, amino acid.

<sup>a</sup>Years of production of the ecotypes. In *M. truncatula*, 346 and 192 genotypes were phenotyped from the 2002 and 2011 harvests, respectively. In *P. sativum*, 373 genotypes were evaluated in two independent experiments (2003 and 2007).

<sup>b</sup>SE, standard error.

<sup>c</sup>Probability (*P* value) of the Fisher test for genotype (SNP) effect with and without population structure as cofactor (Ronfort *et al.*, 2006; J. Burstin *et al.*, unpublished).



**Fig. 2** Alignment of the *SBT1.1* protein sequences from *Medicago truncatula* (MtSBT1.1) and pea (PsSBT1.1). The ethyl methanesulfonate (EMS) mutations are indicated in red and natural polymorphisms are indicated in green for both species. The inhibitory pro-domain, protease and protease-associated domains are indicated by blue, red and green lines, respectively. They were retrieved from the National Center for Biotechnology Information (NCBI) Conserved Domain Database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>). The amino acids of the catalytic triad (D/H/S) are indicated by red triangles. Residues that are highly conserved among the sequences used to obtain the phylogenetic tree (see Fig. 4) are noted below the alignment (the amino acids conserved at 100% in the sequences from Fig. 4 are highlighted).

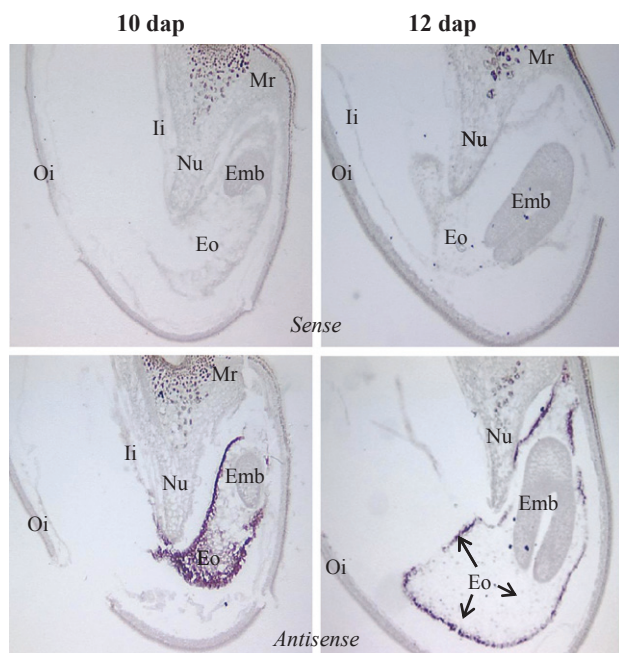
RPKM values of < 0.8 in other tissues (Table S1). The *MtSBT1.1* and *PsSBT1.1* genes were highly expressed in the endosperm compared with other seed tissues (i.e. the embryo and the external envelopes or seed coat) and showed a peak of expression at 10–12 dap (Fig. 1b). In our experimental conditions, 10 dap in pea and 12 dap in *M. truncatula* correspond to a late embryogenesis stage, which is characterized by a seed water content of c. 85% on a fresh weight basis (see Fig. 1b). This stage marks the transition between embryogenesis and seed filling, as embryo cells continue to divide and prepare for reserve deposition (Gallardo *et al.*, 2007). *In situ* hybridizations were performed on *M. truncatula* seed sections at 10 and 12 dap. This confirmed the endosperm-specific expression of *SBT1.1* and revealed a stronger signal in the outer cell layers of the endosperm (Fig. 3).

### Phylogenetic analysis and gene expression profiling of SBTs

The phylogenetic relationship between 55 SBT genes was established (Fig. 4). The phylogenetic tree included *MtSBT1.1*, *PsSBT1.1* and protein sequences from *M. truncatula* and *Arabidopsis* showing the highest homologies with these two SBTs (from 53% to 75%). Four additional SBTs known to be expressed in seeds were included: RSP1 (rice serine protease; Yamagata *et al.*, 2000) of *Oryza sativa*, SCS1 (seed coat-specific; Batchelor *et al.*, 2000), SLP1, SLP2 (subtilisin-like proteases; Beilinson *et al.*, 2002) of soybean (*Glycine max*) and ALE1 of *Arabidopsis* (Tanaka *et al.*, 2001). It is important to note that we recently discovered, in the last *M. truncatula* genome version, a second gene that we named *MtSBT1.2*, showing 90% similarity to *MtSBT1.1* at the protein

level, suggesting a duplication event in the ancestor. This protein sequence was also included in Fig. 4. This analysis revealed that the closest *Arabidopsis* SBTs related to *MtSBT1.1* and *PsSBT1.1* belong to the SBT1 subgroup described by Rautengarten *et al.* (2005) (Fig. 4).

The expression profiles of the 10 *M. truncatula* SBT genes from the SBT1 subgroup were examined by exploiting the Gene Expression Atlas from Benedito *et al.* (2008) available at bar.utoronto.ca (Winter *et al.*, 2007). A comparative analysis of gene expression for the eight *Arabidopsis* genes from the SBT1 subgroup was performed by examining the GeneChip data from Le *et al.* (2010) also available at bar.utoronto.ca. This revealed that only three *M. truncatula* genes were specifically expressed in seeds (Fig. S3, see also Fig. 4). Interestingly, of these, only *MtSBT1.1* and *MtSBT1.2* exhibited a pattern of expression restricted to the early stages of seed development. *MtSBT1.1* and *PsSBT1.1* cluster in a branch comprising the *Arabidopsis* SBT *AtSBT1.7*, also named *ARA12* (*At5g67360* gene). However, *ARA12* is unlikely to be the functional homologue of *MtSBT1.1* and/or *PsSBT1.1* because it is expressed in the seed coat with a specific role in mucilage release from the seed coat on rehydration (Rautengarten *et al.*, 2008). Of the eight *Arabidopsis* genes from the SBT1 subgroup, only one (*At1g01900*, also named *AtSBT1.1*) has a similar expression profile to *MtSBT1.1* with a specific and strong expression in seeds (Fig. S3, see also Fig. 4). Interestingly, the expression of this *Arabidopsis* gene was restricted to the endosperm and peaked, like *MtSBT1.1*, at the linear cotyledon stage that precedes storage protein accumulation (Fig. S4), suggesting that the *At1g01900* gene could be the functional homologue of the legume *SBT1.1* genes.

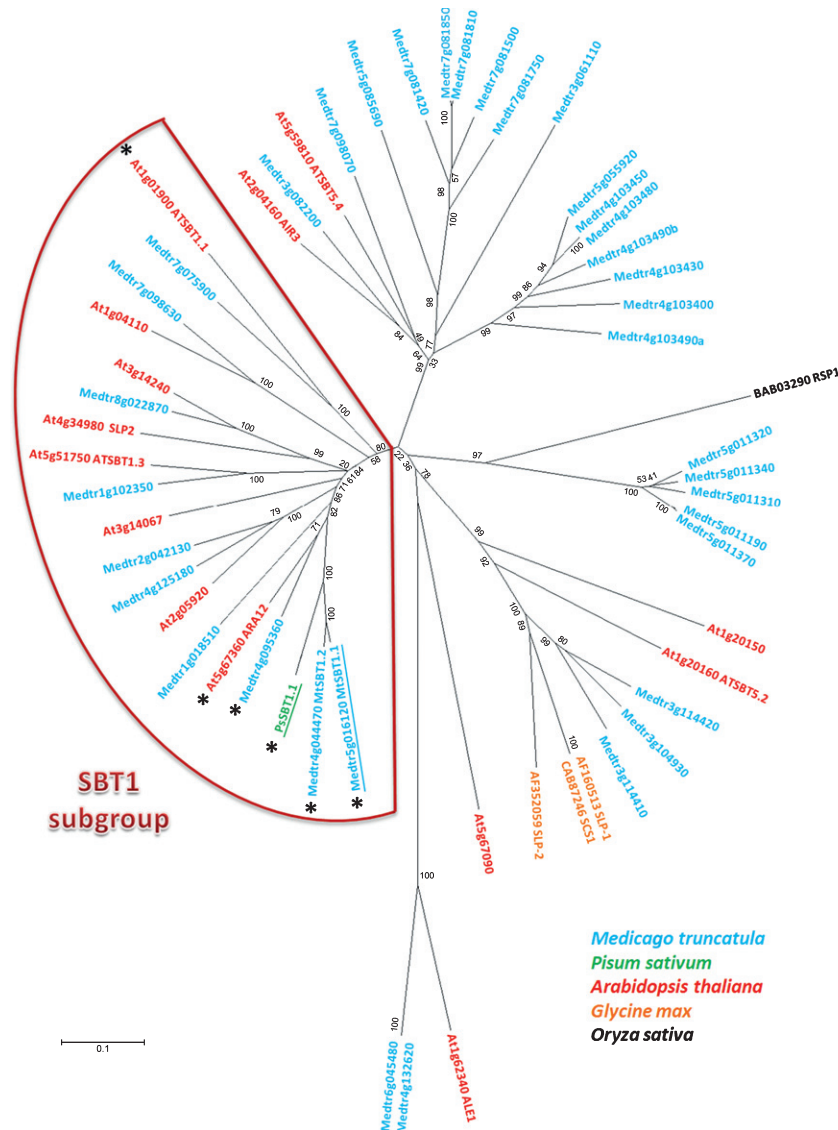


**Fig. 3** *In situ* hybridization experiments of 10 and 12-d after pollination (dap) *Medicago truncatula* seeds with *MtSBT1.1* sense (upper) and antisense (lower) probes. Emb, embryo; Eo, cellular endosperm; li, inner integuments; Mr, micropylar region of integuments; Nu, nucellus; Oi, outer integuments.

### Isolation of point mutations in *MtSBT1.1* and *PsSBT1.1*

The TILLING method applied to the *M. truncatula* and *P. sativum* EMS collections allowed us to identify 12 and 25 missense mutations in *MtSBT1.1* and *PsSBT1.1*, respectively. Based on their position in the sequence and on the resulting amino acid substitutions, six were retained for further analyses: three in *M. truncatula* (*MtR90W*, *MtP330S* and *MtG358D*) and three in pea (*PsT111I*, *PsG216E* and *PsA314V*; Fig. 2). All six point mutations induced nonconservative amino acid substitutions (see Table S4). Based on the 178 SBT sequences found in plant databases, the SIFT tool (<http://sift.jcvi.org/>; Ng & Henikoff, 2003) predicted that all of these substitutions would affect protein function. All six mutations were located in one of the main functional domains of the SBT proteins: the inhibitory pro-domain (*MtR90W*), the protease domain (all pea mutations and *MtP330S*) or the protease-associated domain (*MtG358D*). They were located in highly conserved sites (up to 97%), except for *MtR90W* (24%, see Table S4). It is worth noting that *MtP330S* and *PsG216E* are mutations in the most conserved sites, and that *PsG216E* is located adjacent to the catalytic site His<sup>215</sup> (Fig. 2). As no nonsense mutations were found in the *MtSBT1.1* and *PsSBT1.1* genes, a screening of the *M. truncatula* Tnt1 library (Tadege *et al.*, 2008) was performed, but failed to detect any insertion in the *MtSBT1.1* gene.





**Fig. 4** Unrooted phylogenetic tree of 55 subtilisin-like proteases related to SBT1.1 from *Medicago truncatula* (*MtSBT1.1*) and pea (*PsSBT1.1*). *Medicago truncatula* protein sequences were retrieved from the 3.5 genome version. The phylogenetic tree was constructed using the neighbor-joining method with 1000 bootstraps (Mega5 software; Tamura *et al.*, 2011). The stars indicate the genes specifically expressed in seeds when compared with other plant organs (see Figs S3, S4 for detailed information on gene expression). The SBT1 subgroup described in Rautengarten *et al.* (2005) is indicated. The *MtSBT1.1* and *PsSBT1.1* genes are underlined.

### EMS mutants for SBT1.1 produce smaller seeds

To reduce the number of background mutations, the six retained EMS mutants were backcrossed twice to the wild-type A17 line in *M. truncatula* and the Cameor line in pea. Single seed weights were then measured for mutant and wild-type lines in the same segregating genetic background in two independent experiments (years 2009 and 2010, Table 2 and Fig. S5). In the 2009 experiment, the one-seed weight decreased highly significantly ( $P < 0.01$ ) for three mutants: by 13% for *MtP330S*, 23% for *PsG216E* and 11% for *PsA314V*. In the 2010 experiment, seed weight reductions were highly significant ( $P < 0.005$ ) for all six EMS mutants compared with the wild-type. Seed weight reductions were thus highly significant ( $P < 0.01$ ) in both experiments

for the *MtP330S*, *PsG216E* and *PsA314V* mutations, indicating a stable effect of these mutations. The other mutations (*MtR90W*, *MtG358D* and *PsT111I*) showed a significant decrease in seed weight (by 6–10%) only in the 2010 experiment, suggesting a possible contribution of a genotype by environment effect. In both species, the seed weight of the heterozygous lines was intermediate between the homozygous wild-type and mutant plants, indicating a dosage-dependent effect of the wild-type allele (Fig. S5). Variations in other traits were also seen (Table 2), but none were consistently observed for the mutant alleles. Single pod weight was reduced significantly for the *MtG358D* mutant in the 2010 experiment and for the *PsG216E* mutant in both experiments. Variations in the number of seeds per pod were also found for the *MtP330S* and *PsG216E* mutants, but only in 2010.

**Table 2** Phenotyping data for the six ethyl methanesulfonate (EMS) mutations selected in *SBT1.1* from *Medicago truncatula* (M5BT1.1) and pea (*PisSBT1.1*)

Amino acid substitution	Trait	Expt 1 (year 2009)						Expt 2 (year 2010)						Geno. effect
		Seed progeny	Wild-type		Mutant		Geno. effect	Seed progeny	Wild-type		Mutant		Geno. effect	
			Mean value	SE	Mean value	SE			Mean value	SE	Mean value	SE		
MTR90W	Pod weight (g)	BC1S2	0.095	0.013	0.110	0.014	0.470	BC2S2	0.100	0.002	0.097	0.002	0.315	
	Seed no. per pod	BC1S2	7.575	0.695	8.626	0.761	0.334	BC2S2	9.033	0.145	9.072	0.158	0.797	
	Seed weight (mg)	BC1S2	3.097	0.000	2.999	0.000	0.310	BC2S2	4.094	0.054	3.900	0.058	0.003	
MTP330S	Pod weight (g)	BC2S2	0.089	0.004	0.082	0.003	0.119	BC2S2	0.086	0.002	0.086	0.002	0.902	
	Seed no. per pod	BC2S2	8.323	0.252	8.777	0.196	0.157	BC2S2	8.707	0.152	9.240	0.139	0.010	
	Seed weight (mg)	BC2S2	3.730	0.049	3.235	0.038	0.000	BC2S2	3.806	0.052	3.548	0.047	0.000	
MtG358D	Pod weight (g)	BC2S2	0.077	0.003	0.083	0.003	0.162	BC2S2	0.092	0.002	0.083	0.002	0.006	
	Seed no. per pod	BC2S2	6.736	0.260	7.027	0.282	0.448	BC2S2	7.575	0.212	7.106	0.178	0.092	
	Seed weight (mg)	BC2S2	3.620	0.049	3.642	0.046	0.083	BC2S2	3.964	0.064	3.720	0.054	0.004	
PsT111	Pod weight (g)	BC2S2	0.812	0.104	1.008	0.095	0.170	BC2S3	1.130	0.038	1.054	0.045	0.193	
	Seed no. per pod	BC2S2	3.107	0.375	3.618	0.341	0.318	BC2S3	3.879	0.136	3.750	0.161	0.543	
	Seed weight (g)	BC2S2	0.231	0.005	0.235	0.004	0.471	BC2S3	0.248	0.002	0.223	0.002	0.001	
PsG216E	Pod weight (g)	BC2S2	1.122	0.060	0.798	0.118	0.017	BC2S3	1.434	0.045	1.375	0.066	0.043	
	Seed no. per pod	BC2S2	3.932	0.220	3.467	0.436	0.344	BC2S3	4.695	0.152	4.408	0.222	0.044	
	Seed weight (g)	BC2S2	0.229	0.003	0.177	0.007	0.000	BC2S3	0.264	0.002	0.239	0.002	0.001	
PsA314V	Pod weight (g)	BC2S2	1.152	0.123	1.106	0.223	0.858	BC2S3	1.488	0.099	1.453	0.072	0.698	
	Seed no. per pod	BC2S2	3.758	0.408	3.900	0.741	0.867	BC2S3	5.000	0.329	4.926	0.239	0.951	
	Seed weight (g)	BC2S2	0.245	0.005	0.218	0.008	0.006	BC2S3	0.264	0.003	0.236	0.002	0.001	

Mean values and standard errors (SEs) are shown for pod weight, seed number per pod and one-seed weight measured in the EMS mutant and wild-type seeds in two independent experiments. Probability (*P* value) of the Fisher test for genotype effect is given, and data with *P* < 0.01 are highlighted in grey. BC1S2, BC2S2, BC2S3 stand for progenies issued from one (two) backcross(es) with the reference line followed by two or three generations of selfing.

**Table 3** Average seed projected area and length or diameter of seeds containing one of the six ethyl methanesulfonate (EMS) mutations selected in *SBT1.1* from *Medicago truncatula* (*MtSBT1.1*) and pea (*PsSBT1.1*), estimated from image acquisitions

Allele	Trait	Wild-type		Mutant		Genotype effect, <i>P</i> value
		Mean value	SE	Mean value	SE	
MtR90W	Seed area (cm <sup>2</sup> )	0.058	0.000	0.056	0.000	< 0.0001
	Seed length (cm)	0.330	0.002	0.325	0.001	0.00942
MtP330S	Seed area (cm <sup>2</sup> )	0.054	0.000	0.051	0.000	< 0.0001
	Seed length (cm)	0.320	0.002	0.310	0.002	< 0.0001
MtG358D	Seed area (cm <sup>2</sup> )	0.058	0.000	0.055	0.000	0.00029
	Seed length (cm)	0.330	0.002	0.320	0.002	< 0.0001
PsT111I	Seed area (cm <sup>2</sup> )	0.434	0.002	0.406	0.002	< 0.0001
	Seed diameter (cm)	0.747	0.002	0.723	0.002	< 0.0001
PsG216E	Seed area (cm <sup>2</sup> )	0.446	0.003	0.407	0.003	< 0.0001
	Seed diameter (cm)	0.757	0.003	0.724	0.003	< 0.0001
PsA314V	Seed area (cm <sup>2</sup> )	0.442	0.003	0.413	0.003	0.00020
	Seed diameter (cm)	0.753	0.003	0.727	0.003	0.00020

Image acquisitions were performed from 100 individual seeds per plant for *MtSBT1.1* and three samples of 5–10 seeds per plant for *PsSBT1.1*. The data correspond to BC2S2 and BC2S3 progenies from Experiment 2 (see Table 2). Probability (*P* value) of the Fisher test for genotype effect is given.

Single seed weight was therefore the most significantly varying trait, with *P* values below 0.01, between the wild-type and mutant plants in both species for the two experiments. To investigate whether seed weight variations were caused by differences in seed size or in the amounts of dry matter per seed, a fine analysis of seed surface (projected area, see Materials and Methods) and seed length or diameter was performed for the six mutant alleles from the 2010 experiment. For all six mutants, a significant decrease in seed projected area by 4–9% and length by 2–5% was found for the homozygous mutant seeds when compared with wild-type seeds (Table 3), thus strongly implicating a role for *SBT1.1* in the control of seed size in both species.

### *SBT1.1* is likely to control embryo cell number

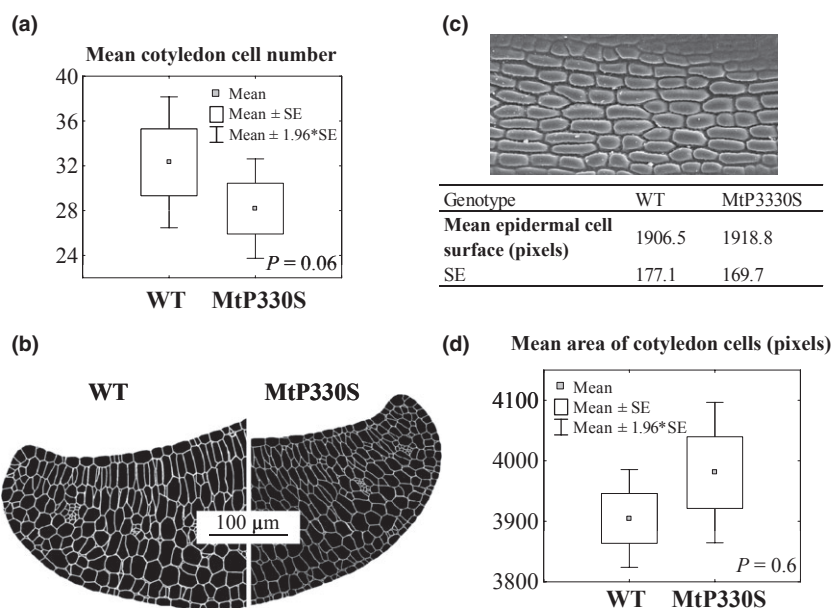
To determine the cause of seed size variation, we performed a microscopic analysis of seeds from the *MtP330S* allele, which exhibits a strong and stable phenotype in *M. truncatula* (Table 2). First, an optical microscopy analysis on seed sections collected at the beginning of seed filling (16 dap in Fig. 1b) did not reveal any significant differences in the morphology and structure of the endosperm, seed coat and embryo tissues compared with the wild-type (data not shown). However, the difficulty of collecting seeds that are synchronized in their development renders a fine comparison of individual seed tissues/cells problematic early in seed development, especially for weak phenotypes caused by missense point mutations. Indeed, pods derived from simultaneously pollinated flowers contain seeds at varying developmental stages according to their position within the pod, and that of the pod on the plant. We therefore focused on the dry mature stage to determine whether the variation in seed size was caused by changes in cotyledon cell number and/or cell size. At maturity, the major tissue component of legume seeds is the embryo, constituents of the endosperm being absorbed by the latter during seed filling and

maturation. By estimating the number of cotyledon cells from optical/light microscopy sections, we found a decrease of *c.* 12%, with a *P* value of 0.06, in the number of cotyledon cells in the *MtP330S* mutant seeds compared with the wild-type (Fig. 5a,b). Scanning electron microscopy did not reveal any significant differences in the mean surface area of cotyledon epidermal cells between the *MtP330S* and wild-type alleles (Fig. 5c). Then, the sizes of epidermal cotyledon cells, round inner and long inner cotyledon cells were estimated from light microscopy images. Again, no significant differences (*P* = 0.6) in cell surface area for the three types of cell were observed between the *MtP330S* mutant and wild-type seeds, although there was a trend towards an increased surface area of all cotyledon cells in the *MtP330S* mutant (Table S5, Fig. 5d).

Seed nitrogen and carbon contents, together with the seed protein composition of the mutants, were unaltered when compared with the wild-type (Fig. S1), suggesting little influence of cotyledon cell number variations on seed composition. A high-resolution transmission electron microscopy study of 16-dap and mature seeds did not show any obvious differences in the organization of seed reserves; lipid droplets and protein storage vacuoles were similar in number and shape in wild-type and mutant seed cotyledons (data not shown). Together, these data indicate that the decrease in weight and size of the *SBT1.1* mutant seeds is caused by a decrease in the number of cells in the mature embryo, rather than a decrease in cell size. *SBT1.1* is therefore more likely to be involved in the control of cell division than in cell expansion during seed development.

### Discussion

In most flowering plants, the triploid endosperm comprises cells containing two maternally derived nuclear genomes and one paternal genome. In Arabidopsis, an increased dose of maternal



**Fig. 5** Surface area and number of cotyledon cells in mature *Medicago truncatula* seeds of the homozygous *MtP330S* mutant vs wild-type. (a) Box plot for comparison of mean cell number per cotyledon section in the *MtP330S* and wild-type mature seeds after reconstruction of cotyledons from mature seeds of the homozygous *MtP330S* mutant and wild-type lines. (b) Typical binary image showing the reconstruction of cotyledons from mature seeds of the homozygous *MtP330S* mutant and wild-type lines made by merging the different images obtained after light microscopy of cotyledons embedded in Epon resin (magnification,  $\times 20$ ) and stained with toluidine blue. (c) Mean epidermal cell surface of mature cotyledons in pixels for at least 600 cells determined from scanning electron microscopy images (from 9–10 seeds analyzed per genotype). SE, standard error. (d) Mean comparison of cell size (in pixels) measured on 0.5- $\mu\text{m}$ -thick sections from cotyledon embedded tissues observed under light microscopy (data acquired using ImageJ from three biological replicates). The cells were divided into three types before analysis (epidermal cells, inner round cells and inner long cells). The box plot corresponds to the global analysis of all cells. Data on the size for each cell type separately are available in Table S5.

genomes in the endosperm with respect to the paternal contribution inhibits endosperm development and causes a dramatic reduction in seed size, whereas an increased paternal contribution has the opposite effect (Scott *et al.*, 1998). These data, together with the analysis of the *HAIKU* and *TITAN* Arabidopsis mutants (Tzafrir *et al.*, 2002; Garcia *et al.*, 2003), indicate a central role of the endosperm in the control of seed size, but many of the genes involved in the determination of this character are unknown. We report here an SBT gene, named *SBT1.1*, which is specifically expressed in legume seeds during the early stages of seed development (10–12 dap; Fig. 1) at the transition stage at which embryo cells are still continuing to divide and are starting to initiate cellular expansion (Abirached-Darmency *et al.*, 2005). *SBT1.1* is expressed in the endosperm (Figs 1, 3). Its exact role during seed development is unknown, but possible functions include a role in controlling embryo growth by providing developmental signals (Berger *et al.*, 2006) or by regulating metabolic fluxes (Melkus *et al.*, 2009).

EMS-induced mutations were identified in the *SBT1.1* sequence of *M. truncatula* and pea by the TILLING method developed by McCallum *et al.* (2000) (Table 2). A detailed inspection of the seed phenotype revealed a decreased weight and surface area of the mutant seeds when compared with the wild-type, thus inferring a role of *SBT1.1* in the control of seed size in both pea and *M. truncatula* species. This is consistent with the evidence for linkage disequilibrium between the SBT locus and a seed weight locus syntenic between these two species, as shown through QTL and association studies (Fig. 1a and Table 1). Although the two

legume species have evolved differently, *M. truncatula* being a wild legume species producing small seeds and pea a crop species producing large seeds, our data suggest that they share similar genetic control of seed weight. In pea, the natural variant SNP G612A did not modify the amino acid properties, but is a promising marker correlated with the causal polymorphism for use in selection.

A large and stable effect on seed weight (a decrease of up to 23%, Table 2) was found for the EMS point mutations located in the protease domain (PsG216E, PsA314V and *MtP330S*, Fig. 2), highlighting these residues as determinant in the function of *SBT1.1*. By contrast, mutations in the inhibitory pro-domain (*MtR90W*), at the junction between the inhibitory pro-domain and the protease domain (PsT111I) and in the protease-associated domain (*MtG358D*) had little effect on seed weight and/or effects that were not reproducible between years (Table 2), suggesting genotype by environment effects and/or that these amino acid substitutions have less impact on protein function. Using PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>), we determined whether the amino acid substitutions in the protease domain were predicted to alter the protein secondary structure. The most prominent changes were observed for the *MtP330S* allele, for which the proline to serine substitution at codon 330 was predicted to result in a change from an extended or  $\beta$  strand to a coil structure at residues 336–337 (GA), 355–356 (KY), 358–360 (GVS) and 374 (F). With the objective of comparing the protease activity of the *MtP330S* and wild-type alleles of *SBT1.1*, we expressed both proteins in *Escherichia coli*. The proteins were clearly expressed after

induction (Fig. S2a), but we did not recover any in-gel protease activity using gelatin as a substrate, suggesting that the bacterially expressed protein is present in an inactive form, or possibly that this protein cleaves a specific substrate. SBTs are prone to extensive post-translational modifications that are required for protease maturation (e.g. processing, glycosylation), which renders it difficult to produce a functional SBT using heterologous expression systems compared with homologous systems using suspension cultures of transgenic cells (Srivastava *et al.*, 2008; Cedzich *et al.*, 2009). Among the studies that have succeeded in the heterologous expression of an active form of recombinant SBT is that of Janzík *et al.* (2000), who used the baculovirus/insect cell system to express an active SBT1 subtilase from tomato. In our study, only the seed extracts exhibited an in-gel protease activity, with no clear differences between the *MtP330S* and wild-type alleles (Fig. S2b), probably because of the presence of other proteases in the developing seeds. In *M. truncatula*, a second gene (*MtSBT1.2*), which is closely related to *MtSBT1.1*, was identified in the recently released Medicago 3.5 genome version (Fig. 4). This gene exhibits 90% similarity to *MtSBT1.1* at the protein level. *MtSBT1.2* is also expressed in the endosperm (data not shown), and may account for the invariant activity seen. *MtSBT1.2* is located on chromosome 4 and is not associated with a seed weight QTL (Vandecasteele *et al.*, 2011).

By focusing on the *MtP330S* homozygous mutant for MtSBT1.1, we found a significant decreased projected area of mature seed (Table 3). The surface area of individual epidermal and inner embryo cells did not decrease when compared with the wild-type (Fig. 5c,d, and Table S5), but there were fewer cells ( $P = 0.06$ ) in the cotyledons of the *MtP330S* mutant seeds (Fig. 5a,b). Although not significant, the trend towards an increased cell surface area of the *MtP330S* mutant seeds compared with the wild-type suggests an increase in cell expansion, which is not sufficient to compensate for the seed size defect. This alteration had no effect on seed nitrogen and carbon contents (Fig. S1). It should be noted that, at maturity, the embryo represents the major part of the legume seed, the endosperm being absorbed by the embryo during development. Therefore, the decrease in seed weight of the *SBT1.1* mutants is likely to be a result of a decrease in the number of cells in the mature embryo. This is consistent with previous data showing that the final cell number in the cotyledons may determine the capacity of the storage organ to accumulate dry matter (Munier-Jolain & Ney, 1998). It is noteworthy in this connection that some SBTs have previously been related to cell division, such as SDD1 and AtSBT5.4, involved in the regulation of stomatal distribution and meristem maintenance, respectively (Berger & Altmann, 2000; Liu *et al.*, 2009).

Substrates of specific SBTs remain mostly unknown. There are indications for a role for SBTs in the maturation of pro-hormones. In Arabidopsis, the SBT AtSBT1.1 participates in the maturation of PSKs (Srivastava *et al.*, 2008), which are sulfated peptide hormones implicated in the stimulation of cell division (Hanai *et al.*, 2000). Three sequences of putative PSKs were retrieved from the *M. truncatula* 3.5 genome version (Medtr5g015140, Medtr1g017760, Medtr4g016870) and, interestingly, one (Medtr5g015140) is located at a chromosomal position coinciding with

the seed weight QTL described in Fig. 1. This putative *PSK* gene (BAC clone ID CU013517, Table S3) is at *c.* 500 kb from *MtSBT1.1* between the MTE30 and MTE32 markers. An alignment of the protein sequence with the Arabidopsis PSKs revealed a best score with PSK4 (AT3G49780, 54% similarity). Interestingly, Srivastava *et al.* (2008) demonstrated the high specificity of AtSBT1.1 for the AtPSK4 precursor *in vitro* using an affinity-purified AtSBT1.1, and also during callus formation. Although there is no report for such a role of AtSBT1.1 in developing seeds, it is interesting to note that AtSBT1.1 exhibits a seed-specific pattern of expression restricted to the endosperm at a stage preceding storage protein deposition (i.e. the linear cotyledon stage corresponding to the 10–12-dap stages in *M. truncatula* and pea, see Fig. 1b) (Figs S3, S4). It is therefore tempting to speculate that *MtSBT1.1* and *PsSBT1.1* could play a role in the regulation of seed weight via the maturation of molecules in the endosperm, such as peptide hormones, which could act non-cell autonomously to control embryo cell division during late embryogenesis. In support of this hypothesis, a recent study identified a polypeptide signaling molecule promoting seed growth and overall seed size in Arabidopsis. This peptide ligand, which regulates embryo and suspensor proliferation, is encoded by the CLE8 (*CLAVATA3/EMBRYO SURROUNDING REGION-RELATED8*) gene, which is expressed in early embryos and in the endosperm (Fiume & Fletcher, 2012). Alternatively, certain SBT proteases are involved in programmed cell death in plants (Vartapetian *et al.*, 2011), and MtSBT1.1 could perhaps exert a similar function in the endosperm. However, we did not observe any structural difference between the wild-type and mutant endosperms at 16 dap, suggesting that degeneration of the endosperm occurs similarly in the *MtP330S* and wild-type seeds at 16 dap.

In conclusion, our results demonstrate a role for the endosperm-localized SBT1.1 in the control of seed size in legumes, which probably acts through the regulation of embryo cell division. Further studies are needed to identify the substrate of SBT1.1 in developing seeds, which will help to define the precise role of SBT1.1, and to assess the possible contribution of the seed-specific SBT recently identified, MtSBT1.2, to the control of seed weight in legumes. Moreover, the selection of favorable alleles in natural populations by exploiting polymorphisms in the SBT1 gene (e.g. G612A in Table 1) may help to breed pea and other crop legumes for high seed weight.

## Acknowledgements

We sincerely thank Thierry Huguet (ENSAT, Montpellier, France) for providing us with the genotyping set of the LR4 population, and Jérôme Gouzy, Marion Verdenaud (INRA, UMR441, Toulouse, France) and Vincent Savoie (INRA, UMR1347, Dijon, France) for providing us with the links between the Affymetrix probesets and the corresponding genomic region in *M. truncatula*. We are grateful to Marion Dalmais (INRA, URGV Evry, France) for design and optimization with regard to the TILLING screen in pea and to Joelle Ronfort (INRA, UMR1334, Montpellier, France) for providing us with the structure of the *M. truncatula* collection. We also thank our colleagues from

UMR1347 (INRA, Dijon, France): Christine Arnould, Aline Bonnotte and Joël Michel for their help with microscopic observations, Charles Schneider for creating scripts allowing for the analyses of the cell surface, Anne-Lise Santoni for elemental analyses, Delphine Aimé for setting up the in-gel activity test, Benjamin Vallée for initiating MtSBT1.1 cloning, and Mathieu Siol for helpful advice regarding the phylogenetic analysis. This work was financed by the Génoplatte 2010 QualityLegSeed project GPLA06036G.

## References

- Abirached-Darmency M, Abdel-gawwad MR, Conejero G, Verdeil JL, Thompson R. 2005. *In situ* expression of two storage protein genes in relation to histo-differentiation at mid-embryogenesis in *Medicago truncatula* and *Pisum sativum* seeds. *Journal of Experimental Botany* 56: 2019–2028.
- Aubert G, Morin J, Jacquin F, Loridon K, Quillet MC, Petit A, Rameau C, Lejeune-Hénaut I, Huguët T, Burstin J. 2006. Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. *Theoretical and Applied Genetics* 112: 1024–1041.
- Batchelor AK, Boutlier K, Miller SS, Labbé H, Bowman L, Hu M, Johnson DA, Gijzen M, Miki BL. 2000. The seed coat-specific expression of a subtilisin-like gene, SCS1, from soybean. *Planta* 211: 484–492.
- Beilinson V, Moskalenko OV, Livingstone DS, Reverdatto SV, Jung R, Nielsen NC. 2002. Two subtilisin-like proteases from soybean. *Physiologia Plantarum* 115: 585–597.
- Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T *et al.* 2008. A gene expression atlas of the model legume *Medicago truncatula*. *Plant Journal* 55: 504–513.
- Berger D, Altmann T. 2000. A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in *Arabidopsis thaliana*. *Genes and Development* 14: 1119–1131.
- Berger F, Grini PE, Schnittger A. 2006. Endosperm: an integrator of seed growth and development. *Current Opinion in Plant Biology* 9: 664–670.
- Bordat A, Savoies V, Nicolas M, Salse J, Chauveau A, Bourgeois M, Potier J, Houtin H, Rond C, Murat F *et al.* 2011. Translational genomics in legumes allowed placing *in silico* 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. *Genes, Genomes, Genetics* 1: 93–103.
- Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Analytical Biochemistry* 72: 248–254.
- Bradley D, Carpenter R, Sommer H, Hartley N, Coen E. 1993. Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of *Antirrhinum*. *Cell* 72: 85–95.
- Burstin J, Marget P, Huat M, Moessner A, Mangin B, Duchene C, Desprez B, Munier-Jolain N, Duc G. 2007. Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. *Plant Physiology* 144: 768–781.
- Cedzich A, Huttenlocher F, Kuhn BM, Pfannstiel J, Gabler L, Stintzi A, Schaller A. 2009. The protease-associated domain and C-terminal extension are required for zymogen processing, sorting within the secretory pathway, and activity of tomato subtilase 3 (SISBT3). *Journal of Biological Chemistry* 284: 14068–14078.
- Coen ES, Romero JM, Doyle S, Elliott R, Murphy G, Carpenter R. 1990. *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* 63: 1311–1322.
- Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR *et al.* 2004. Efficient discovery of DNA polymorphisms in natural populations by EcoTILLING. *Plant Journal* 37: 778–786.
- Dalmais M, Schmidt J, Le Signor C, Moussy F, Burstin J, Savoies V, Aubert G, Brunaud V, de Oliveira Y, Guichard C *et al.* 2008. UTILLdb, a *Pisum sativum* *in silico* forward and reverse genetics tool. *Genome Biology* 9: R43.
- Deulvot C, Charrel H, Marty A, Jacquin F, Donnadiou C, Lejeune-Hénaut I, Burstin J, Aubert G. 2010. Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC Genomics* 11: 468.
- Fiume E, Fletcher JC. 2012. Regulation of Arabidopsis embryo and endosperm development by the polypeptide signaling molecule CLE8. *Plant Cell* 24: 1000–1012.
- Fontanini D, Jones BL. 2002. SEP-1 – a subtilisin-like serine endopeptidase from germinated seeds of *Hordeum vulgare* L. cv. Morex. *Planta* 215: 885–893.
- Gallardo K, Firnhaber C, Zuber H, Hélicher D, Belghazi M, Henry C, Küster H, Thompson R. 2007. A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. *Molecular and Cellular Proteomics* 6: 2165–2179.
- Garcia D, Saingery V, Chambrier P, Mayer U, Jurgens G, Berger F. 2003. Arabidopsis haiku mutants reveal new controls of seed size by endosperm. *Plant Physiology* 131: 1661–1670.
- Hanai H, Matsuno T, Yamamoto M, Matsubayashi Y, Kobayashi T, Kamada H, Sakagami Y. 2000. A secreted peptide growth factor, phytosulfokine, acting as a stimulatory factor of carrot somatic embryo formation. *Plant and Cell Physiology* 41: 27–32.
- Janzik I, Macheroux P, Amrhein N, Schaller A. 2000. LeSBT1, a subtilase from tomato plants. Overexpression in insect cells, purification, and characterization. *Journal of Biological Chemistry* 275: 5193–5199.
- Le BH, Cheng C, Bui AQ, Wagmaster JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S *et al.* 2010. Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences, USA* 107: 8063–8070.
- Le Signor C, Savoies V, Aubert G, Verdier J, Nicolas M, Pagny G, Moussy F, Sanchez M, Baker D, Clarke J *et al.* 2009. Optimizing TILLING populations for reverse genetics in *Medicago truncatula*. *Plant Biotechnology Journal* 7: 430–441.
- Liu JX, Srivastava R, Howell S. 2009. Overexpression of an *Arabidopsis* gene encoding a subtilase (AtSBT5.4) produces a clavata-like phenotype. *Planta* 230: 687–697.
- Luft JH. 1961. Improvements in epoxy resin embedding methods. *Journal of Cell Biology* 9: 409–414.
- McCallum C, Comai L, Greene E, Henikoff S. 2000. Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiology* 123: 439–442.
- Melkus G, Rolletschek H, Radchuk R, Fuchs J, Rutten T, Wobus U, Altmann T, Jakob P, Borisjuk L. 2009. The metabolic role of the legume endosperm: a noninvasive imaging study. *Plant Physiology* 151: 1139–1154.
- Munier-Jolain N, Ney B. 1998. Seed growth rate in grain legumes II. Seed growth rate depends on cotyledon cell number. *Journal of Experimental Botany* 49: 1971–1976.
- Neff MM, Neff JD, Chory J, Pepper AE. 1998. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. *Plant Journal* 14: 387–392.
- Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* 31: 3812–3814.
- Rasband WS. 1997–2008. *ImageJ*. Bethesda, MD, USA: US National Institutes of Health. <http://rsb.info.nih.gov/ij/>.
- Rautengarten C, Steinhauser D, Büssis D, Stintzi A, Schaller A, Kopka J, Altmann T. 2005. Inferring hypotheses on functional relationships of genes: analysis of the *Arabidopsis thaliana* subtilase gene family. *PLoS Computational Biology* 1: e40.
- Rautengarten C, Usadel B, Neumetzler L, Hartmann J, Büssis D, Altmann T. 2008. A subtilisin-like serine protease essential for mucilage release from *Arabidopsis* seed coats. *Plant Journal* 54: 466–480.
- Rayburn LYM, Rhea J, Jocoy SR, Bender M. 2009. The proprotein convertase *amontillado* (*amon*) is required during *Drosophila* pupal development. *Developmental Biology* 333: 48–56.
- Ronfort J, Bataillon T, Santoni S, Delalande M, David JL, Prospero JM. 2006. Microsatellite diversity and broad scale geographic structure in a model legume: building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula*. *BMC Plant Biology* 6: 28.
- SAS. 1999. *Statistical Analysis System software*. Cary, NC, USA: SAS Institute Inc.

- Schaller A, Stintzi A, Graff L. 2012. Subtilases – versatile tools for protein turnover, plant development, and interactions with the environment. *Physiologia Plantarum* 145: 52–66.
- Scott RJ, Spielman M, Bailey J, Dickinson HG. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* 125: 3329–3341.
- Srivastava R, Liu JX, Howell SH. 2008. Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in *Arabidopsis*. *Plant Journal* 56: 219–227.
- Strongin AY, Izotova LS, Abramov ZT, Gorodetsky DI, Ermakova LM, Baratova LA, Belyanova LP, Stepanov VM. 1978. Intracellular serine protease of *Bacillus subtilis*: sequence homology with extracellular subtilisins. *Journal of Bacteriology* 133: 1401–1411.
- Tadege M, Wen J, He J, Tu H, Kwak Y, Eschstruth A, Cayrel A, Endre G, Zhao PX, Chabaud M *et al.* 2008. Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. *Plant Journal* 54: 335–347.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.
- Tanaka H, Onouchi H, Kondo M, Hara-Nishimura I, Nishimura M, Machida C, Machida Y. 2001. A subtilisin-like serine protease is required for epidermal surface formation in *Arabidopsis* embryos and juvenile plants. *Development* 128: 4681–4689.
- Thoquet P, Gherardi M, Journet EP, Kereszt A, Ane JM, Proserpi JM, Huguet T. 2002. The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. *BMC Plant Biology* 2: 1.
- Tzafirir I, McElver JA, Liu CM, Yang LJ, Wu JQ, Martinez A, Patton DA, Meinke DW. 2002. Diversity of TITAN functions in *Arabidopsis* seed development. *Plant Physiology* 128: 38–51.
- Vandecasteele C, Teulat-Merah B, Morère-Le Paven MC, Leprince O, Ly Vu B, Viau L, Ledroit L, Pelletier S, Payet N, Satour P *et al.* 2011. QTL analysis reveals a correlation between the ratio of sucrose/raffinose family oligosaccharides and seed vigour in *Medicago truncatula*. *Plant, Cell & Environment* 34: 1473–1487.
- Vartapetian AB, Tuzhikov AI, Chichkova NV, Taliansky M, Wolpert TJ. 2011. A plant alternative to animal caspases: subtilisin-like proteases. *Cell Death and Differentiation* 18: 1289–1297.
- Von Groll U, Berger D, Altmann T. 2002. The subtilisin-like serine protease SDD1 mediates cell-to-cell signaling during *Arabidopsis* stomatal development. *Plant Cell* 14: 1527–1539.
- Wang YT, Yang CY, Chen Y-T, Lin Y, Shaw J-F. 2004. Characterization of senescence-associated proteases in postharvest broccoli florets. *Plant Physiology and Biochemistry* 42: 663–670.
- Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. 2007. An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* 2: e718.
- Yamagata H, Uesugi M, Saka K, Iwasaki T, Aizono Y. 2000. Molecular cloning and characterization of a cDNA and a gene for subtilisin-like serine proteases from rice (*Oryza sativa* L.) and *Arabidopsis thaliana*. *Bioscience, Biotechnology, and Biochemistry* 64: 1947–1957.
- Yang S, Johnston N, Talideh E, Mitchell S, Jeffree C, Goodrich J, Ingram G. 2008. The endosperm-specific ZHOUP1 gene of *Arabidopsis thaliana* regulates endosperm breakdown and embryonic epidermal development. *Development* 135: 3501–3509.
- Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H *et al.* 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480: 520–524.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Fig. S1** Composition of mature seeds from the *MtP330S* and wild-type genotypes.

**Fig. S2** Protease activity tests for the wild-type and *MtP330S* alleles.

**Fig. S3** Expression profiles across plant tissues of the *Medicago truncatula* and *Arabidopsis* subtilase genes from the SBT1 subgroup.

**Fig. S4** Expression data across seed tissues for the eight *Arabidopsis* subtilases from the SBT1 subgroup.

**Fig. S5** Seed weight variations between wild-type and the heterozygous (Het) and homozygous (HO) mutant plants.

**Table S1** RNA-Seq data for the *PsSBT1.1* gene in 12 organs at five developmental stages and under two nutrition conditions

**Table S2** List of dCaps primers used for genotyping the pea and *Medicago truncatula* mutants

**Table S3** *In silico* mapping of genes between the genetic markers MTE30 and MTE32 which are anchored to the physical map of *Medicago truncatula*

**Table S4** SIFT (Sorting Intolerant From Tolerant) predictions for the amino acid changes identified by TILLING (Targeting Induced Local Lesions in Genomes)

**Table S5** Microscopic observations of mature seeds from the *MtP330S* TILLING (Targeting Induced Local Lesions in Genomes) mutant and the wild-type form of *MtSBT1.1*

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