



HAL
open science

Variability within a pea core collection of LEAM and HSP22, two mitochondrial seed proteins involved in stress tolerance

Marie-Hélène Avelange-Macherel, Nicole Payet, David Lalanne, Martine Neveu, Dimitri Tolleter, Judith Burstin, David Macherel

► To cite this version:

Marie-Hélène Avelange-Macherel, Nicole Payet, David Lalanne, Martine Neveu, Dimitri Tolleter, et al.. Variability within a pea core collection of LEAM and HSP22, two mitochondrial seed proteins involved in stress tolerance. *Plant, Cell and Environment*, 2015, 38 (7), pp.1299-1311. 10.1111/pce.12480 . hal-01392646

HAL Id: hal-01392646

<https://hal.science/hal-01392646>

Submitted on 28 Mar 2023

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.

Original Article

Variability within a pea core collection of LEAM and HSP22, two mitochondrial seed proteins involved in stress tolerance

Marie-Hélène Avelange-Macherel¹, Nicole Payet², David Lalanne², Martine Neveu², Dimitri Tolleter³, Judith Burstin⁴ & David Macherel⁵

¹Agrocampus Ouest, ²INRA and ⁵Université d'Angers, UMR 1345 Institut de Recherche en Horticulture et Semences, Angers F-49045, France, ³ANU College of Medicine, Biology and Environment, Acton 2601, Australia and ⁴GEAPSI, INRA, UMR 1347 Agroécologie, centre de Dijon F-21065, France

ABSTRACT

LEAM, a late embryogenesis abundant protein, and HSP22, a small heat shock protein, were shown to accumulate in the mitochondria during pea (*Pisum sativum* L.) seed development, where they are expected to contribute to desiccation tolerance. Here, their expression was examined in seeds of 89 pea genotypes by Western blot analysis. All genotypes expressed LEAM and HSP22 in similar amounts. In contrast with HSP22, LEAM displayed different isoforms according to apparent molecular mass. Each of the 89 genotypes harboured a single LEAM isoform. Genomic and RT-PCR analysis revealed four LEAM genes differing by a small variable indel in the coding region. These variations were consistent with the apparent molecular mass of each isoform. Indels, which occurred in repeated domains, did not alter the main properties of LEAM. Structural modelling indicated that the class A α -helix structure, which allows interactions with the mitochondrial inner membrane in the dry state, was preserved in all isoforms, suggesting functionality is maintained. The overall results point out the essential character of LEAM and HSP22 in pea seeds. LEAM variability is discussed in terms of pea breeding history as well as LEA gene evolution mechanisms.

Key-words: gene evolution; indel; LEA protein; mitochondria; pea accessions; small HSP.

INTRODUCTION

Most seeds can withstand the almost total loss of cellular water during the last stage of their development. Such desiccation tolerance (or anhydrobiosis) can be found across all biological kingdoms and relies on an array of mechanisms, including the accumulation of non-reducing sugars and stress proteins (Hoekstra *et al.* 2001). Mitochondria are critical for energy supply during germination and because respiratory activity is known to resume very rapidly during

imbibition, it was inferred that mitochondria should remain functional throughout the dry state (Macherel *et al.* 2007). Accordingly, pea seed mitochondria were shown by proteomic analysis to contain high levels of LEAM, a late embryogenesis abundant (LEA) protein, and of HSP22, a small heat shock protein (sHSP) (Bardel *et al.* 2002; Grelet *et al.* 2005). LEA proteins are highly hydrophilic proteins which were highlighted more than 30 years ago as the major proteins expressed during late seed maturation (Dure & Galau 1981). LEA proteins are prominent seed proteins, but were also found later in other anhydrobiotes (eubacteria, rotifers, nematodes, tardigrades and arthropods), suggesting an important role with respect to desiccation tolerance (Tunnacliffe *et al.* 2010). LEA protein sequences are very diverse and several classifications have been proposed based on their properties. More recently, several hundreds of LEA protein sequences have been gathered in a dedicated database (<http://forge.info.univ-angers.fr/~gh/Leadb>) and subjected to computational and statistical analyses (Hunault & Jaspard 2010; Jaspard *et al.* 2012). The distinctive features of LEA proteins are their high hydrophilicity, the presence of repeated motifs and their lack of well-defined tertiary structure in the hydrated state (Tunnacliffe & Wise 2007). In spite of their widely recognized importance for desiccation tolerance, the molecular function of LEA proteins is only starting to emerge, with a variety of effects in agreement with their diversity (Tunnacliffe *et al.* 2010; Battaglia & Covarrubias 2013).

LEAM is a pea LEA protein that belongs to protein family LEA_4 (PFAM). It accumulates in the matrix of pea seed mitochondria and was shown to be a fully disordered protein in the hydrated state (Tolleter *et al.* 2007). However, upon drying, the protein was shown to fold into an amphipathic helical form (class A α -helix) that could insert laterally within membrane, parallel to the plane, being stabilized by electrostatic interactions with phosphate groups of phospholipids (Tolleter *et al.* 2007, 2010). Hence, LEAM is able to reversibly insert into the protein-free regions of the inner mitochondrial membrane upon drying, providing protection in the dry state. Interestingly, a protein orthologous to

Correspondence: M.-H. Avelange-Macherel. Fax: +33 241 225 549; e-mail: marie-helene.macherel@agrocampus-ouest.fr

LEAM was also identified in the mitochondria from the encysted embryos of the brine shrimp *Artemia franciscana*, which are anhydrobiotic organisms (Menze *et al.* 2009). HSP22 is a sHSP that was originally shown to be expressed in pea leaf mitochondria in response to heat stress (Lenne & Douce 1994). sHSPs, which are members of the protein chaperone family, share a highly conserved α -crystalline motif and are especially abundant in plants (Siddique *et al.* 2008). They generally exhibit a dynamic oligomeric structure and, in contrast with other heat shock proteins, are not constitutive, being highly induced under stress conditions (Wang *et al.* 2004). sHSPs have been shown to prevent aggregation of misfolded polypeptides and/or protect membranes in stress conditions (Torok *et al.* 2001; Nakamoto & Vigh 2007; Benesch *et al.* 2008). In plants, sHSPs are expressed in response to heat, oxidative and water stress (Sun *et al.* 2002). Interestingly, several sHSPs are also expressed during seed development (Derocher & Vierling 1994; Wehmeyer *et al.* 1996; Wehmeyer & Vierling 2000). The role of the mitochondrial HSP22 has not been clarified, but its high level of expression in seeds suggests it could be involved in desiccation tolerance and contribute to heat stress tolerance during germination (Bardel *et al.* 2002; Stupnikova *et al.* 2006). In order to gain more insight into the representation and genetic diversity of the mitochondrial stress proteins LEAM and HSP22, we explored their expression patterns in seeds of 89 pea accessions. This core collection from INRA CRG (Dijon, France) comprises cultivars, breeding lines, local populations, wild species and germplasm from the genus *Pisum* (*Pisum fulvum*, *P. abyssinicum*, *P. sativum*) and covers a wide genetic diversity (Supporting Information Table S1; Baranger *et al.* 2004; Smýkal *et al.* 2012). LEAM and HSP22 were found to be abundant in seeds of all accessions, but the results highlighted the variability of LEAM, with several isoforms that were characterized at the protein and genomic level.

MATERIALS AND METHODS

Plant materials

The seeds of 89 *Pisum* accessions were provided by INRA (Institut National pour la Recherche Agronomique, CRG Protéagineux, INRA, 17 rue de Sully, 21065 Dijon, France; <https://urgi.versailles.inra.fr/siregal/siregal/grc.do>). Plants were grown simultaneously in a greenhouse (22 °C day/16 °C night, 16 h photoperiod). Two independent seed harvests were made.

Western blot analysis

Ten seeds per genotype were ground with steel beads, using a tissue lyser (Qiagen GmbH, Hilden, Germany). One hundred fifty milligrams of powder were mixed with 1.5 mL extraction buffer [50 mM NaPO₄ pH 8.0, 10 mM ethylenediaminetetraacetic acid (EDTA), 0.1% v/v Triton X100, 0.1% v/v lauryl sarcosine, 10 mM β -mercaptoethanol, protease inhibitor cocktail (complete EDTA free tablets,

Roche Diagnostic GmbH, Mannheim, Germany)] for 30 min on ice and then clarified by centrifugation at 12 000 g, for 20 min at 4 °C. To select for heat soluble proteins, the supernatants were heated at 95 °C for 10 min and centrifuged again. Protein concentrations were measured according to the Bio-Rad protein assay (Bio-Rad Laboratories, Inc., Hercules, CA, USA) based on the method of Bradford.

Proteins (40 μ g) were separated on a 13.5% polyacrylamide gel using standard sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) conditions. Electroblothing and immunodetection were performed according to Stupnikova *et al.* (2006), using chemiluminescence for detection. The secondary anti-rabbit IgG antibody conjugated to horseradish peroxidase (Sigma, St. Louis, MO, USA) was used at a dilution of 1:50 000. Blots were incubated for 5 min in Clarity™ Western ECL substrate (Bio-Rad) and chemiluminescence was monitored with a molecular imager Chemidoc TM XRS system (Bio-Rad).

The primary antibodies (Neosystem, Strasbourg, France) were raised against synthetic peptides designed from LEAM (UniProtKB Q5NJJ5, Baccara cultivar) and HSP22 (UniProtKB P46254, Douce Provence cultivar) protein sequences. L1 and L2 (working dilution 1:5000) were raised against LEAM repeated domain (EKAKDYAYDAK ERTKD) and C-terminal end (KGYGENKGYDQNRGY), respectively. H1 and H2 (working dilution 1:1000) were raised against MVDLLTDNPVLSAAS and YKIDVIKAEMKN GVL of HSP22, respectively.

For two-dimensional (2-D) electrophoresis, heat soluble proteins (400 μ g) were precipitated with 10% w/v trichloroacetic acid in acetone and resuspended in 200 μ L IEF (isoelectric focusing) buffer (6M urea, 2M thiourea, 4% w/v CHAPS, 20 mM DTT and 1% v/v Biolytes from Biorad). Isoelectric focusing was performed with the Protean IEF Cell (Bio-Rad) using 7 cm strips (linear pH gradient 4.7–5.9, Bio-Rad) according to Bio-Rad recommendations. The second dimension (SDS-PAGE) was carried out using a 13.5% polyacrylamide gel.

Molecular masses were estimated using pre-stained Precision Plus Protein™ Standards (Bio-Rad). Seed mitochondria, purified according to Benamar *et al.* (2003), were used as positive control in Western blot analysis.

Bioinformatic analyses

The theoretical molecular mass and isoelectric point were determined from cDNA sequences using the ProtParam program (<http://expasy.org>). Helical projections of α -helices were built with the HeliQuest webserver (<http://heliquest.ipmc.cnrs.fr>).

cDNA synthesis

Total RNA was extracted from seed powder as described by Verwoerd *et al.* (1989) and quantified with a Nanodrop (Nanodrop Technologies Inc., Wilmington, DE, USA). RNA quality was checked by separation on a 1.2% w/v agarose gel.

After RQ1 RNase-free DNase (Promega Corporation, Madison, WI, USA) treatment, reverse transcription was performed with M-MLV reverse transcriptase (Promega) according to the manufacturer recommendations.

PCR amplifications

The genomic DNA used for PCR amplification was extracted from 3-week-old leaves using the Nucleospin Plant II kit from Macherey-Nagel (Düren, Germany).

Specific PCR primers (Supporting Information Table S2) were designed using the free Primer3 software (Rozen & Skaletsky 2000). Oligonucleotides were provided by Eurogentec (Liege, Belgium). PCR reactions were carried out in 15 μ L using 2 μ L genomic DNA or cDNA and GoTaq[®] Flexi DNA polymerase according to the manufacturer recommendations (Promega). Amplifications were performed in a DNA thermal cycler (iCycler, Biorad): 95 °C, 2 min; 35 \times (95 °C, 30 s; 65 °C, 30 s; 72 °C, 50 s); 72 °C, 5 min. PCR products were sequenced by Eurofins MWG Operon (Ebersberg, Germany).

Non-radioactive Southern blot

A two-step genomic DNA extraction (nuclei isolation followed by CTAB (Cetyl Trimethyl Ammonium Bromide) extraction) was performed on leaves (Hanania *et al.* 2004). DNA (10 μ g) was digested overnight at 37 °C, in a 100 μ L final volume using 100 U of restriction enzyme (Thermo Fisher Scientific Inc., Waltham, MA, USA) in the recommended buffer. DNA was separated on 0.8% Resophor[®] agarose (Eurobio, Courtaboeuf, France) gel, using 1X TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA pH 8). A total of 50 ng of DNA Molecular Weight Marker III, digoxigenin labelled (Roche Diagnostic GmbH), were used as a size ladder. A pCR[®]4-TOPO[®] plasmid (Invitrogen, Carlsbad, CA, USA) containing the LEAM cDNA (accession GenBank AJ628940) was linearized by NcoI digestion and 5 ng were run on the gel as a positive control. After depurination and denaturation, gels were blotted on Hybond N+ positively charged nylon membrane (GE Healthcare, Piscataway, NJ, USA) with the VacuGene[™] XL vacuum blotting System (GE Healthcare) for 1 h. Nucleic acids were fixed using a Stratalinker[®] UV cross-linker (Stratagene, Cedar Creek, TX, USA; 50 s at 0.120 J cm⁻²). After 30 min prehybridization, the membrane was hybridized with a digoxigenin-labelled probe overnight, at 42 °C in the DIG Easy Hyb solution (Roche Diagnostic GmbH). The probe (369 bp) was prepared with the PCR DIG Probe Synthesis Kit (Roche Diagnostic GmbH) using the Baccara genomic DNA as a template, primers PsLEAm-F1 and PsLEAM-R1 (Supporting Information Table S2) and the Phusion[®] DNA Polymerase (Finnzymes, Espoo, Finland). The immunodetection was carried out by adding CDP-Star reagent (Roche Diagnostic GmbH) and monitoring the chemiluminescence as previously described.

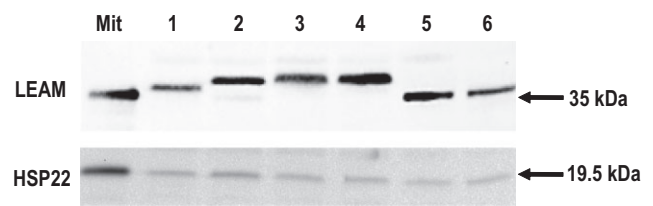


Figure 1. LEAM and HSP22 representation among a selection of pea genotypes. Dry seed protein extracts (40 μ g of proteins) were analysed by Western blot using anti-LEAM or anti-HSP22 antibodies (L2 and H2 antibodies). Mit: seed mitochondria from Baccara (reference genotype), lane 1 to 6: pea genotypes from the core collection (1: Cote d'or; 2: Mistral; 3: WNC23Z; 4: WNC Z61Z; 5: PI273279; 6: PI180693).

RESULTS

Systematic occurrence of LEAM and HSP22 proteins in pea seeds

Antibodies were raised against C-terminal and N-terminal peptides selected in the *P. sativum* LEAM (Baccara cultivar) and HSP22 (Douce Provence cultivar) sequences. Peptides were selected on the basis of their antigenic properties and specificity, that is their sequences being unlikely to be present in other LEA proteins or sHsps. Both anti-HSP22 antibodies detected a polypeptide with an apparent mass of 19.5 kDa and anti-LEAM antibodies recognized a 35 kDa protein in the mitochondria isolated from imbibed Baccara seeds (Fig. 1 and Supporting Information Figs S1 and S2). These apparent molecular masses correspond to the processed forms of HSP22 and LEAM after cleavage of their respective mitochondrial targeting peptides (Bardel *et al.* 2002). Antibodies were sensitive enough to detect the corresponding polypeptides in extracts from whole pea seeds (Fig. 1). We did not observe differences in the abundance of either protein in extracts from individual Baccara seeds (data not shown).

The abundance of LEAM and HSP22 was determined in seeds of the 89 pea genotypes by Western blot analysis performed on total protein extracts. A first screening was performed with antibodies raised against peptides in the C-terminal regions of LEAM and HSP22. Typical results for a series of representative genotypes are presented in Fig. 1. A protein with an apparent molecular mass of 20 kDa was detected with a similar intensity by the anti-HSP22 antibody in all genotypes, suggesting that HSP22 does not show any significant variability in pea. In contrast, the antibody raised against LEAM revealed up to three proteins differing by their apparent molecular mass (Fig. 1). For the sake of clarity, these proteins will be referred hereafter as isoforms. Isoform 1 (35 kDa) corresponds to LEAM in our reference genotype Baccara. The isoforms 2 and 3 exhibited slightly higher apparent molecular masses of, respectively, 37 and 39.5 kDa. While the distribution of isoforms was variable, the overall abundance of the proteins remained similar among all but one genotype; LEAM protein was not detected in Torsdag genotype. Because small variations of amino acid sequence could

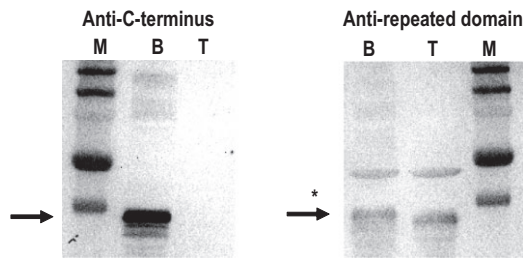


Figure 2. LEAM isoform pattern in Torsdag and Baccara genotypes seeds using two anti-LEAM antibodies for immunodetection. Dry seeds extracts (40 μ g) from Torsdag and Baccara genotypes were analysed by Western blot using two distinct anti-LEAM antibodies (L1 anti-repeated domain and L2 anti-C-terminus). LEAM is indicated by arrows, non-specific signals by asterisks (*). B, Baccara; T, Torsdag; M, molecular mass markers (100, 75, 50 and 37 kDa from top to bottom).

weaken or prevent antibody binding, a second screening was performed with another antibody raised against a peptide in the repeat domain of LEAM. The presence of the protein was confirmed in Torsdag seeds using this antibody, which clearly revealed isoform 1 (Fig. 2). The same results were obtained with several seed lots of Torsdag and the presence of isoform 1 was confirmed using isolated mitochondria from imbibed Torsdag seeds (data not shown).

At a whole, these results demonstrate that both HSP22 and LEAM are expressed to similar levels in the 89 genotypes, the latter displaying three different isoforms according to their electrophoretic behaviour (Western blots for the 89 genotypes are presented in Supporting Information Figs S1 and S2).

Distribution of LEAM isoforms among the pea core collection

The two different antibodies were used to analyse the isoform pattern in seeds from the pea core collection. Except for Torsdag, both immunoscreenings provided similar results as shown for representative genotypes (Fig. 3)

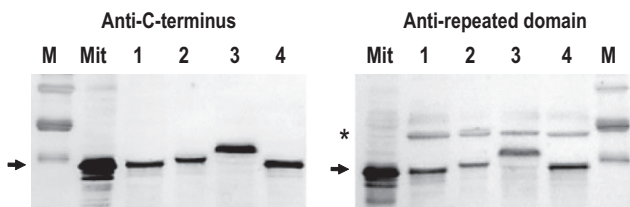


Figure 3. LEAM isoform pattern in selected pea genotypes seeds, using two anti-LEAM antibodies for immunodetection. Dry seeds extracts (40 μ g) from representative genotypes were analysed by Western blot, using L1 and L2 anti-LEAM antibodies. M, molecular mass markers 75, 50 and 37 kDa from top to bottom; Mit, mitochondria from *Pisum sativum* cv. Baccara seeds; lane 1, Baccara; lane 2, JI1794; lane 3, Glacier; lane 4, JI2605. Arrows indicate the isoform found in Baccara seed mitochondria and asterisks (*) non-specific signals.

enabling identification of the LEAM isoform pattern within the collection (Table 1). Testing seeds from an independent harvest confirmed that the variability in the LEAM isoform pattern was driven by the genotype without a strong environmental effect (data not shown). All genotypes harboured only one isoform. Isoform 1 was found in 49 genotypes and isoform 3 in 35 genotypes. In contrast with isoforms 1 and 3, isoform 2 was poorly represented (five genotypes).

Molecular characterization of LEAM isoforms

Apparent molecular mass variation of LEAM was unlikely to arise from post-translational modifications. Such differences of several kDa would have required large modifications such as glycosylation which is not expected for mitochondrial proteins. LEAM variability could rather have a genetic origin with isoforms encoded by different genes or resulting from alternative splicing events. If different *LEAM* genes coexist in the pea genome, regulation at the promoter level may also influence the distribution of isoforms in the core collection.

To explain the variability of LEAM at the molecular level, cDNA and gene sequences were determined for three genotypes, Baccara, Melrose and Glacier, which harbour isoforms 1, 2 and 3, respectively. The Western blot data shown in Fig. 4 confirm that seeds of each cultivar contain exclusively a single isoform, which clearly differ in apparent molecular mass as shown when the three protein extracts are mixed before electrophoretic separation.

PCR primers designed according to the full-length Baccara LEAM cDNA sequence (Grelet *et al.* 2005) were used to amplify by RT-PCR the corresponding LEAM precursor coding sequence for the three genotypes, using seed cDNAs as template, and amplicons were cloned and sequenced (Supporting Information Fig. S3). The comparison of the deduced amino acid sequences is presented in Fig. 5. No difference was found between genotypes in the targeting peptide sequence (amino acid 1 to 37). The mature LEAM protein sequence appeared very similar between the three genotypes, except that Melrose and Glacier exhibited a single insertion at a similar site (amino acid 168, 504 bp from the ATG) of 16 and 22 amino acids, respectively. Interestingly, the 16 amino acid insertion of the Melrose cultivar is shared by the Glacier cultivar, which exhibits an additional stretch of six residues. Outside of the indel region, a single amino acid insertion and 11 substitutions were found, almost all of them concerning residues with similar physicochemical features (Fig. 5).

We also identified by PCR the LEAM genomic sequence for the three genotypes, which confirmed the indel event and the other variations in the coding region (data not shown). Small differences were observed in the nucleotide sequence of the single intron conserved between the three *LEAM* genes. Melrose and Glacier intron sequences (114 nucleotides) were identical but differed from the corresponding Baccara *LEAM* intron sequence (113 bp) by three bases substitutions and an additional nucleotide (Supporting

Name	LEAM isoform	Name	LEAM isoform
90–2131	1	COTED'OR	2
ABYSSINICUM VAVILOVIANUM	1	DP	2
BACCARA	1	J11794	2
Bul179	1	J11844	2
CAMEOR	1	MELROSE	2
CAROUBY DE MAUSSANE	1	AFGHANISTAN ASIATICUM	3
CERISE ce CR*	1	ALASKA*	3
CHAMPION D'ANGLETERRE	1	AMINO	3
CHEYENNE	1	BALLET	3
CHINA	1	BINGEFORS	3
CORNE DE BELIER	1	CAPSICUM	3
COSTARICA	1	CE101 = FP	3
DESIREE	1	CHAMPAGNE	3
E11	1	CLAMART HATIF	3
E175	1	CUZCO1	3
E65*	1	DSP*	3
HOLLY11	1	ENGLISH	3
J11075	1	FIN DE LA BIEVRE	3
J11089	1	GLACIER	3
J11267	1	HAITI COLORE	3
J11703	1	HATIVER	3
J12202	1	J11831	3
J12376	1	J1228	3
J12473	1	J12383	3
J12523	1	J1241	3
J12546	1	K4269*	3
J12605	1	K8290	3
J1261*	1	KARNOBAT	3
J1296	1	KAZAR	3
K1666	1	KOROZA	3
K4088	1	MICHAUX DE PARIS	3
KIRIN40	1	MISTRAL*	3
LIVIOLETTA	1	MONGOLIA*	3
MESSIRE	1	MULTIRESISTANT	3
NEPALA	1	SERPETTE D'AUVERGNE	3
NFG KRUPP PEL.	1	SERPETTE GUILLOTEAUX	3
PI180693*	1	SOMMETTE	3
PI212112	1	WINTERBERGER	3
PI273279	1	WNC23Z*	3
PLEIN LE PANIER*	1	WNCZ61Z	3
SHRAT	1		
TELEPHONE A RAMES	1		
TERESE	1		
TORS DAG	1		
Vavilov-D265	1		
WIRAIG	1		
YANGWAN*	1		
ZP126	1		
ZP141	1		

Table 1. LEAM isoforms distribution in the 89 pea genotypes of the core collection

Pea genotypes were classified according to the LEAM electrophoretic mobility determined by Western blot analysis (1, isoform 1; 2, isoform 2; 3, isoform 3). The *LEAM* gene could be amplified by genomic PCR as described in the Materials and Methods section and further sequenced, except for 12 genotypes (*).

Information Fig. S4). Such minor differences are unlikely to interfere with a RNA splicing event.

We investigated, using Southern blot analysis, whether the *LEAM* gene was present as a single copy or multiple copies in the three genotypes, Baccara, Melrose and Glacier.

Genomic DNA was digested using three restriction enzymes that do not cut within any of the known *LEAM* gene sequences. Whatever the enzyme used, the pattern shown in Fig. 6 revealed more than one hybridized band, suggesting the occurrence of several *LEAM* genes in each genotype. Up

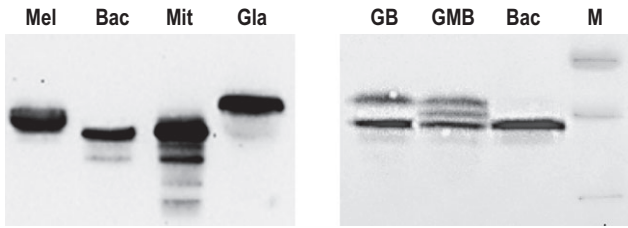


Figure 4. LEAM isoform pattern for the three genotypes selected for molecular characterization. Seed extracts from Baccara (Bac), Melrose (Mel) and Glacier (Gla) genotypes were analysed by Western blot using the anti-LEAM antibody L2. Baccara, Melrose and Glacier seeds contain the isoform 1, 2 and 3, respectively. The analysis also includes mixed seed extracts: GB (Gla + Bac) and GMB (Gla + Mel + Bac). Mit, mitochondria from Baccara seeds; M, molecular mass markers (50, 37 and 25 kDa from top to bottom).

to five DNA fragments hybridizing with the LEAM probe were detectable (see Fig. 6, NcoI lane). Southern blot profiles were similar between genotypes with Hind III digestion but some differences appeared upon EcoRI and NcoI digestion. Interestingly, the patterns obtained after NcoI digestion displayed one band whose size differed between genotypes. In Melrose DNA, this NcoI fragment was estimated to be approximately 135 bp longer than in Baccara, while in the case of Glacier DNA, this NcoI fragment displayed a 307 bp increase compared with Baccara. This polymorphism could be consistent with indels in Melrose and Glacier *LEAM*

LEAM-Bac	MAAMFTTRNAIFRFSKSFNPNVPSLSLKPSPRVFVASASHQSDWRNAADGKRNSMDWAYN	60
LEAM-Mel	MAAMFTTRNAIFRFSKSFNPNVPSLSLKPSPRVFVASASHQSDWRNAADGKRNSMDWAYN	60
LEAM-Gla	MAAMFTTRNAIFRFSKSFNPNVPSLSLKPSPRVFVASASHQSDWRNAADGKRNSMDWAYN	60

LEAM-Bac	STSKARQDADEIADRERKTLNGDVSDVSKQVVRDAKERTKEAANRAAENADSAGVKS	120
LEAM-Mel	STSKARQDADEIADRERKTLNGDVSDVSKQVVRDAKERTKEAANRAAENADSAGVKS	120
LEAM-Gla	STSKARQDADEIADRERKTLNGDVSDVSKQVVRDAKERTKEAANRAAENADSAGVKS	120

LEAM-Bac	YAYDAKERTKDAANRAAENVESAGEKAKDYAYDAKERTKDAANRAAEN-----	168
LEAM-Mel	YAYDAKERTKDAANRAAENVESAGEKAREYAFDAKERTKDAANRAAE-----TARDYA	173
LEAM-Gla	YAYDAKERTKDAANRAAENVESAGEKAREYAFDAKERTKDAANRAAENVESAGETARDYA	180

LEAM-Bac	-----AESVGEKARDYAYDAKERTKEAAQAGETAKDYAYGAKERTKEAAESAGE	218
LEAM-Mel	FDKAKERTKEAAESAGEKVKDYAYDAKERTKEAAQAGETAKDYAYGAKERTKEAAESAGE	233
LEAM-Gla	FDKAKERTKEAAESAGEKVKDYAYDAKERTKEAAQAGETAKDYAYGAKERTKEAAESAGE	240

LEAM-Bac	TARDYAYDADTKTEAVGTVADTKTEGAKKTAEMTKEGAETAEITTEGVAGAATEALKSA	278
LEAM-Mel	TARDYAYDADTKTEAVGTVADTKTEGAKKTAEMTKEGAETAEITTEGVAGAATEALKSA	293
LEAM-Gla	TARDYAYDADTKTEAVGTVADTKTEGAKKTAEMTKEGAETAEITTEGVAGAATEALKSA	300

LEAM-Bac	GEMAKRTAQGAWETAKDQTKIKETVVGKDDDDNDRGGVGVAVVDEYDVELKRRKGYGESK	338
LEAM-Mel	GEMAKRTAQGAWETAKDQTKIKETVVGKDDDDNDRGGVGVAVVDEYDVELKRRKGYGENK	353
LEAM-Gla	GEMAKRTAQGAWETAKDQTKIKETVVGKDDDDNDRGGVGVAVVDEYDVELKRRKGYGENK	359

LEAM-Bac	GYDKSKGYGENKGYDQNRGY 358	
LEAM-Mel	GYDKSKGYGENKGYDQNRGY 373	
LEAM-Gla	GYDKSKGYGENKGYDQNRGY 379	

Figure 5. Alignment of the deduced amino acid sequences of LEAM precursors from Baccara (LEAM-Bac), Melrose (LEAM-Mel) and Glacier (LEAM-Gla) genotypes. The coding sequence for the three LEAM isoforms was amplified by PCR from the cDNA of Baccara, Melrose and Glacier genotypes. The alignment of the three sequences was generated using the ClustalW2 program. The degree of similarity between amino acids occupying the same position is indicated by the following symbols: (*) identical amino acids; amino acids with strong similar properties, (.) amino acids with weak similar properties. Grey boxes indicate the indel region.

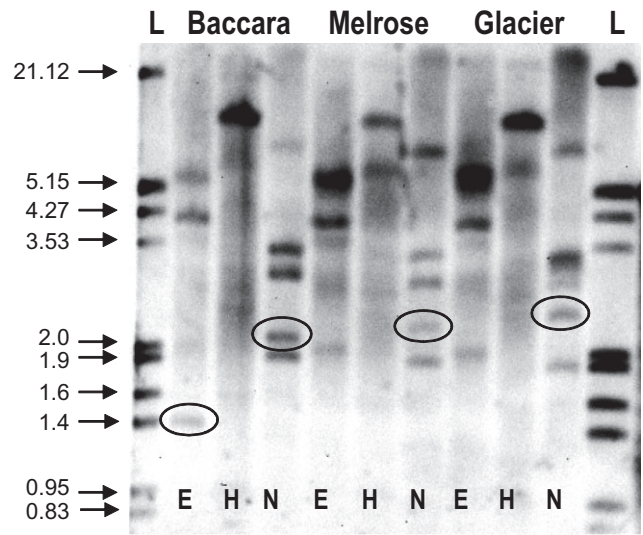


Figure 6. Southern blot analysis of the LEAM gene. Genomic DNA from Baccara, Melrose and Glacier genotypes was digested by EcoRI (E), HindIII (H) or NcoI (N). DNA fragments were separated by electrophoresis on a 0.8% Resophor® agarose gel and then transferred on a Hybond N + positively nylon membrane. The DIG-labelled probe used for hybridization corresponds to the nucleotides 26 to 394 of the Baccara LEAM coding sequence. The probe overlaps the intron but not the indel region. Sizes of the DNA ladder standards (L) are indicated in kbp. The main differences observed between the three genotypes are indicated by open circles.

genes, although fragment lengths were higher than expected as indels represent 48 bp in Melrose and 66 bp in Glacier. The occurrence of more than one *LEAM* gene on this variable NcoI fragment could account for these discrepancies. Restriction length polymorphism might explain the other divergences between genotypes observed by Southern blot analysis.

The pea core collection was further analysed by genomic PCR, using specific primers designed to match the three *LEAM* isoform gene sequences. The region corresponding to approximately 75% of the *LEAM* coding sequence, overlapping both the intron and indel, could be amplified and sequenced in 77 genotypes out of 89 (Supporting Information Fig. S5). Non-sequenced genotypes are indicated in Table 2. The overall molecular data were in good agreement with the Western blot results. The sequence analysis of the PCR products confirmed that the *LEAM* gene variability results mainly from an indel event. Indel size, sequence and position were the same as described for Melrose and Glacier except for one genotype (JI1794). Indeed, in this genotype, a 33 bp indel occurred at 787 bp from the ATG (Baccara coding sequence as reference) and corresponded to the duplication of a MTKEGAETAE motif (Fig. 7). The presence of this additional segment was consistent with the electrophoretic mobility of the JI1794 LEAM protein. According to the Western blot data, this genotype was originally classified in the isoform 2 group (Fig. 3), but genomic data revealed sequence encoding for a fourth LEAM isoform (referred as isoform 4).

Table 2. Distribution of LEAM isoforms in a subset of winter- and spring-sown pea cultivars, not included in the core collection

Cultivar	Sowing	LEAM isoform
Assas	Winter	2
Cartouche	Winter	3
Enduro	Winter	3
Isard	Winter	1
James	Winter	1
Lucy	Winter	3
Bluemoon	Spring	1
Crackerjack	Spring	1
Goldfinger	Spring	1
Hardy	Spring	1
Lumina	Spring	1
Metaxa	Spring	1
Standard	Spring	1

LEAM isoforms were identified in dry seed protein extracts by Western blot analysis, using the two different anti-LEAM antibodies.

Isoform 1 genes displayed few nucleotide variations between genotypes while isoform 2 and isoform 3 genes were perfectly conserved. The isoform 4 gene shared more similarities with the isoform 1 gene; however, eight SNP (Single Nucleotide Polymorphism) were specific and not found in neither of the three other isoforms genes (data not shown). Except for indels, the 77 LEAM nucleotide sequences were highly conserved among the pea core collection, as well as the deduced amino acid sequences.

Biochemical features of LEAM isoforms

Relevant genotypes were chosen to further compare the properties of LEAM isoforms. As shown in Fig. 8, the four isoforms were found to be fully heat soluble, a property which is a typical feature of LEA proteins (Boudet *et al.* 2006) and which was previously used to purify Baccara recombinant LEAM (Grelet *et al.* 2005). Protein extracts were separated by 2D-PAGE and subjected to Western blot analysis to examine the protein profiles of LEAM isoforms. All isoforms displayed a similar rosary-type pattern with several spots sharing the same apparent molecular mass but with distinct isoelectric points ranging from 4.7 to 5.2 (Fig. 9).

```

ISO1/ISO4 KERTKDAANRAAEN-----AESVGEKARDYAYDAKERTKEAAQ 192
          KERTKDAANRAAE-----TARDYAFDAKERTKEAAESAGEKVKDYAYDAKERTKEAAQ 207
          KERTKDAANRAAENVESAGETARDYAFDAKERTKEAAESAGEKVKDYAYDAKERTKEAAQ 214

ISO1/ISO2/ISO3 AKKTAEMTKEGAEKTAE-----TTGEVAGAATE 273
          AKKTAEMTKEGAEKTAEMTKEGAEKTAETTGEVAGAATE 284

```

Figure 7. Alignment of the four LEAM isoforms in the indel region. LEAM precursors amino acid sequences were aligned using the ClustalW2 program. Indels are indicated in bold, repeated motifs are underlined in amino acid sequences. Numbers indicate the position of the last amino acid in the displayed sequences.

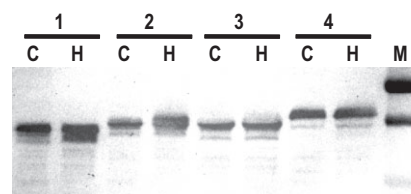


Figure 8. Heat solubility of the four LEAM isoforms. Seed protein extracts were heated for 10 min at 95 °C then centrifuged at 4 °C for 30 min at 14 000 g. Supernatants from heated (H) or non-heated samples (C) were analysed by Western blot using the anti-LEAM antibody L2. Results are shown for representative genotypes expressing each LEAM isoform. 1, Baccara (ISO1); 2, Melrose (ISO2); 3, J11794 (ISO4); 4, Glacier (ISO3); M, molecular mass markers (50, 37 kDa from top to bottom).

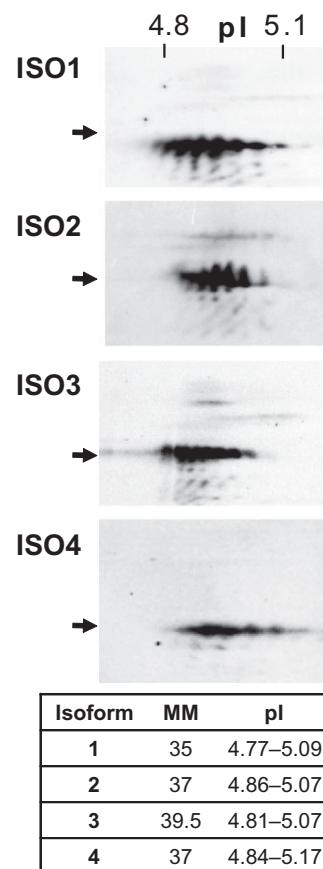


Figure 9. 2D-PAGE profiles of the four LEAM isoforms. Heat soluble seed proteins from Baccara (ISO1), Melrose (ISO2), Glacier (ISO3) and J11794 (ISO4) genotypes were separated by 2D-PAGE. A narrow pH range (4.7 to 5.9) was used for IEF and the second dimension separation was performed by SDS-PAGE using a 13.5% polyacrylamide gels. After blotting, LEAM isoforms were immunodetected with the anti-LEAM C-terminal antibody (L2) and revealed by chemiluminescence. Arrows indicate the 37 kDa molecular mass marker. Apparent molecular mass (MM) and pI (isoelectric point) were determined experimentally for each isoform and are indicated in the table.

Such a profile, which was previously observed in the case of Baccara LEAM, results from intrinsic chemical post-translational modifications (Tolletter *et al.* 2007). The apparent molecular mass and isoelectric points, which were experimentally determined for each isoform, were consistent with the bioinformatics predictions (Fig. 9). The additional segment of 11, 16 and 22 amino acids, displayed by the isoform 4, 2 and 3, respectively, have very little impact on the isoelectric point.

For isoforms 2 and 3, the insertion occurs in a repeated region that is predicted to fold into class A amphipathic α -helix, a structure which allows the protein to interact with membranes in the dry state (Tolletter *et al.* 2007, 2010). Class A helices display a peculiar arrangement of the polar face, with positively charged residues aligned on both sides of the interface with the non-polar face, while negatively charged residues are aligned in the middle of the polar face (Fig. 10). This arrangement allows the protein to insert laterally in the phospholipid bilayer, with the non-polar face exposed to the fatty acid core, while positively charged residues interact with the phosphate groups, stabilizing the protein (Tolletter *et al.* 2010). The isoform 4 indel occurs in another repeated region that is also predicted to fold into α -helix structures. When helical projections of isoforms were examined in the region corresponding to the indels, the class A motif appeared strikingly conserved in the four isoforms (Fig. 10). Therefore, the different indels are unlikely to interfere with this essential structural feature of the LEAM protein.

Correlation between LEAM isoforms and genotype characteristics

Seeds of the core collection display contrasting colour, size and shape but we did not find any correlation between seed phenotypic features and their LEAM isoform content. Correlations between LEAM isoform distribution and the genotypes' geographical origin and cultivation status were investigated using available data from the Plant Genetic Resources Information System of the National Institute for Agronomical Research (INRA); <https://urgi.versailles.inra.fr/siregal/siregal/grc.do>. As shown in Fig. 11, isoform 1 was the most represented in Europe, Africa, Middle and Far East accessions. In contrast, in North and South America, isoform 3 and 2 were most common. Isoform 2 was not present in Africa, Middle and Far East genotypes. JI1794 (isoform 4) originated from the Middle East. The overrepresentation of isoform 1 was also noticeable among wild species, which mostly come from Middle East and Africa (Fig. 12). The isoform 3 was more associated with cultivated genotypes (Fig. 12).

Because LEA proteins, especially dehydrins, have been associated with cold and freezing tolerance (Puhakainen *et al.* 2004), we have also investigated whether the distribution of isoforms could be associated with winter hardiness. The pea core collection contains 41% and 20% of spring and winter varieties, respectively. Results shown in Fig. 13 underlined the overrepresentation of isoform 1 in spring-sown peas (61%). The isoforms 2 and 3 were preponderant

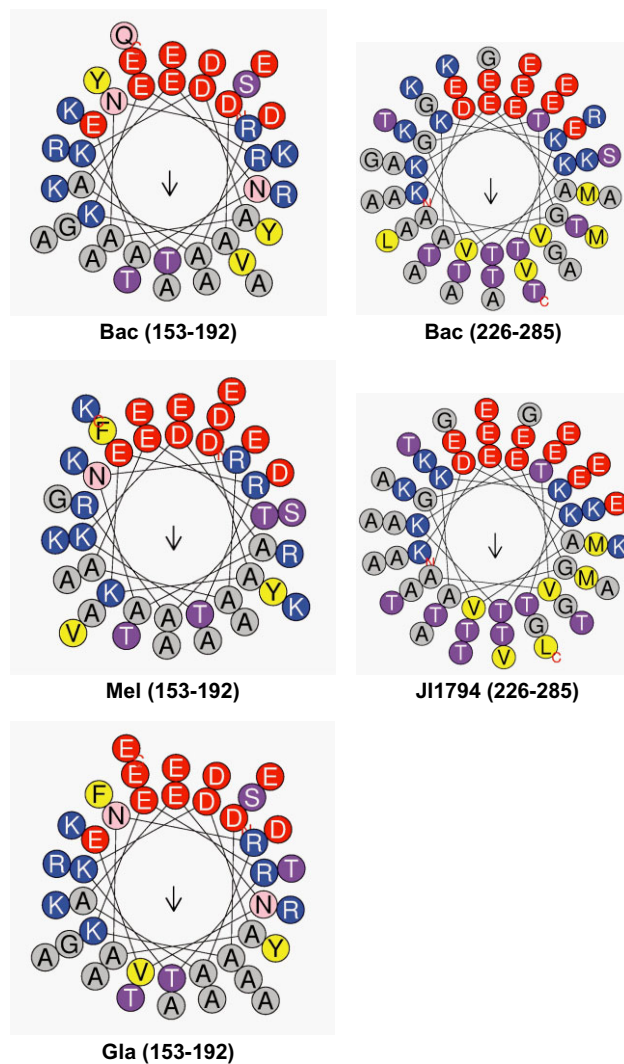


Figure 10. Helical wheel projections of the indel regions. Helical projections of α -helices were obtained using the HeliQuest web server (<http://heliquest.ipmc.cnrs.fr/>). Residues are colour coded in blue (K, R) for positively and red (D, E) for negatively charged residues. Non-polar residues are shown in yellow or grey and others in purple, light blue or light pink colours. The arrow shows the hydrophobic moment. Bac, isoform 1 from Baccara; Mel, isoform 2 from Melrose; Gla, isoform 3 from Glacier; JI1794, isoform 4 from JI1794. The numbers between brackets correspond to residue positions in LEAM precursors.

in winter-sown genotypes. Furthermore, the four genotypes displaying isoform 2 were all winter-sown peas. Because data concerning winter hardiness was missing for 39% of the core collection, we have analysed an additional set of winter and spring-sown cultivars. The seven spring cultivars displayed only isoform 1, while the isoform distribution was heterogeneous in the six winter-sown genotypes, with two occurrences for isoform 1, one for isoform 2 and three for isoform 3 (Table 2). This additional analysis confirms that isoform 1 is overrepresented in spring cultivars, while the isoform distribution appears more heterogeneous in winter varieties.

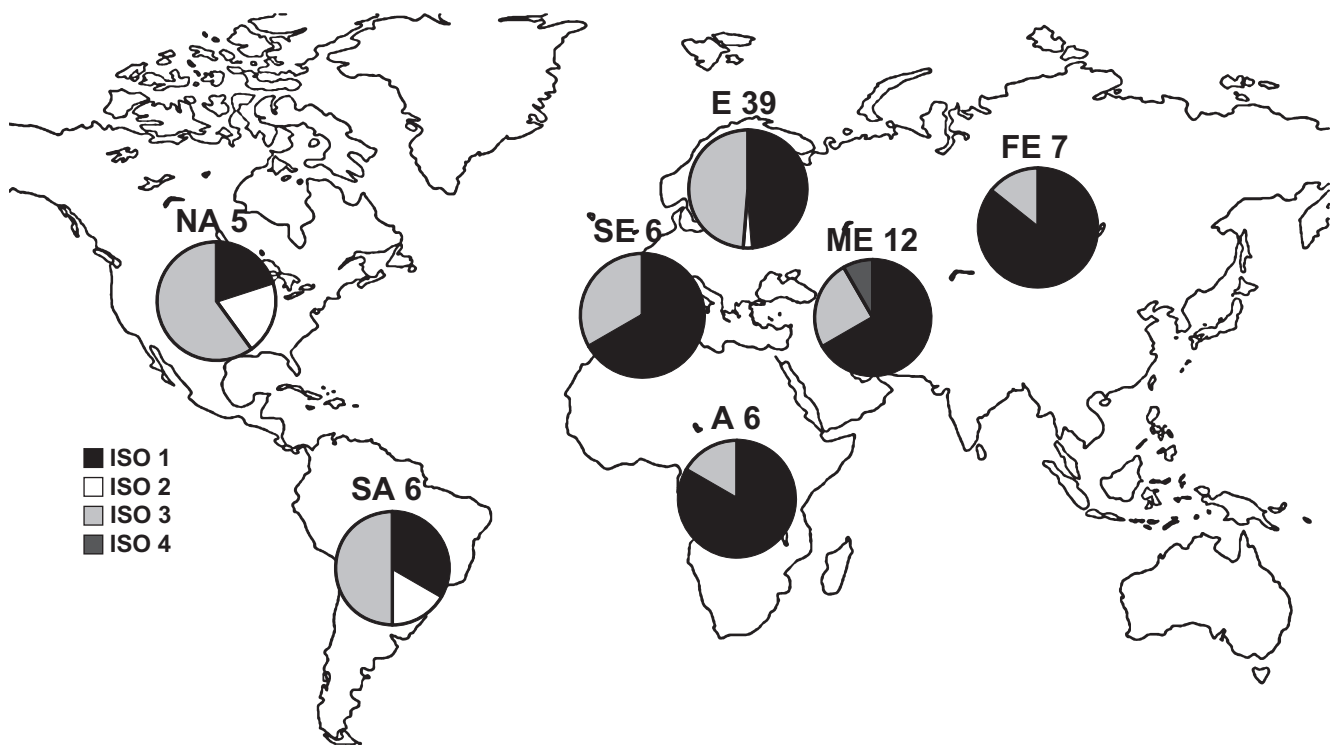


Figure 11. Geographical distribution of LEAM isoforms.

Pea accessions originate from North America (NA), South America (SA), South Europe (SE), Europe and North Europe (E), Africa (A), Middle East (ME) and Far East (FE). The origin of eight genotypes was unknown. Pie charts display the distribution of the LEAM isoforms in each geographical zone. Figures above each chart correspond to the genotype number in each geographical zone.

DISCUSSION

LEAM and HSP22 proteins accumulate to high levels in the mitochondria during pea seed development (Bardel *et al.* 2002) and are thus expected to contribute to seed desiccation tolerance as well as to germination vigour (Macherel *et al.* 2007). Here, the Western blot analysis revealed that both

proteins were systematically detected at almost identical levels in the seeds of 89 pea genotypes. This reinforces the idea that these two mitochondrial proteins perform important functions in pea seeds. Biochemical and biophysical

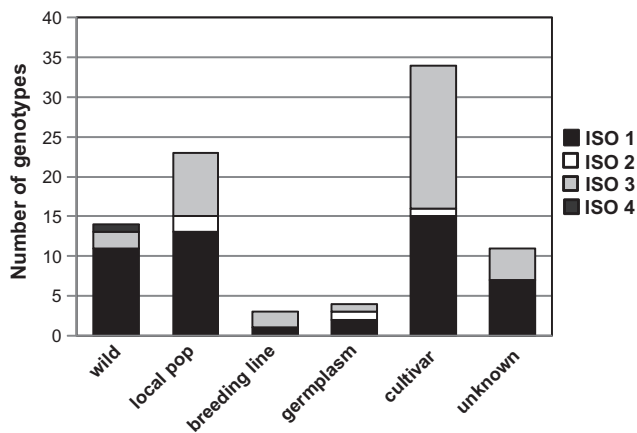


Figure 12. LEAM isoforms distribution as a function of the culture status of pea genotypes. The pea core collection harbours wild genotypes, local populations, breeding lines, germplasm and cultivars. Eleven genotypes have unknown culture status. LEAM isoform distribution was determined for each culture status.

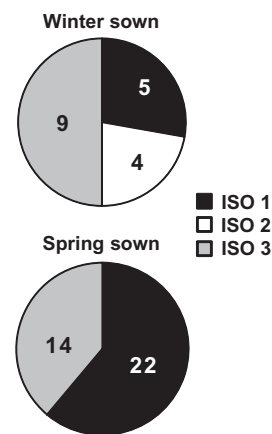


Figure 13. LEAM isoforms distribution in spring- and winter-sown pea genotypes. Pie charts show the LEAM isoform distribution in the pea core collection as a function of winter hardiness. The number of genotypes for each isoform class is indicated on chart pies. Winter hardiness data were missing for 39% of genotypes from the core collection, including JI1794 (ISO4).

analyses strongly argue for a protective role of LEAM in the protection of mitochondrial inner membrane upon desiccation (Toltelet *et al.* 2007, 2010). A side function could be to preserve the activity of mitochondrial enzymes under stress (Grelet *et al.* 2005). Another LEA protein from the LEA_1 family, ATEM6, was shown to be required for normal *Arabidopsis thaliana* seed development, but not for seed germination (Manfre *et al.* 2006). The protein was suggested to buffer the water loss occurring during embryo maturation. Kalembe & Pukacka (2008) have proposed that dehydrins, together with sugars and possibly sHSPs, sustain the viability of beech seeds during storage through their contribution to the glassy state formation. The role of HSP22 in seed mitochondria is not yet elucidated. Although the expression of sHSPs is generally stress-dependent, they have been shown to be developmentally expressed during seed and pollen development (Waters 2012). The onset of sHSP accumulation in seeds occurs in the mid and late maturation phases, as demonstrated for cytosolic sHSPs in pea and other species (Wehmeyer *et al.* 1996; Wehmeyer & Vierling 2000). After germination and early seedling growth, sHSPs abundance decreases rapidly (Derocher & Vierling 1994). In *Arabidopsis* mutants with reduced seed dormancy, sHSPs abundance was not affected, while expression was undetectable in a desiccation-sensitive mutant (Wehmeyer *et al.* 1996). These results support the idea that sHSPs might be essential for desiccation tolerance (Wehmeyer *et al.* 1996) and/or seed yield and quality as shown for a plastid sHSP (Chauhan *et al.* 2012). sHSPs may also accumulate in developing seeds to protect embryos from heat stress that could occur during seed formation or germination. The remarkable temperature tolerance of isolated pea seed mitochondria suggests that HSP22 could contribute to heat tolerance (Stupnikova *et al.* 2006). While it is becoming clear that both LEAM and HSP22 play important roles as stress response proteins in seeds, knockout mutants will be required to determine the developmental stages in which they are essential.

By combining immunoblot analysis, cDNA and genomic analysis, we could demonstrate the occurrence of four different LEAM isoforms within the pea genotype collection. These isoforms, which are encoded by specific genes and differ by a variety of indels, do not result from RNA alternative splicing but rather from a genetic insertion/deletion polymorphism. Each genotype was shown to harbour only one isoform, both at the protein and gene level, although Southern blotting suggested the presence of multiple copies of the gene. Resolving the actual genomic organization of *LEAM* gene in pea will require a genome sequence, which is not yet available. In the model legume *Medicago truncatula*, the genome has been fully sequenced (<http://www.medicago-hapmap.org/>) and a single gene (Medtr2g014040.1) was shown to bear strong homologies with *LEAM* (BLAST2.0 <http://blast.wustl.edu> expect value = 2.4×10^{-90} , 54% identities) and it is indeed predicted to encode a mitochondrial protein. As 384 *Medicago* lines are being sequenced (*Medicago* Hapmap project) in order to study polymorphism (SNP, insertions/deletions and copy

number), it will be of interest to determine whether *LEAM* isoforms with indels also occurs in this species.

All LEAM isoforms shared common features such as thermal solubility and a rosary-shaped profile in 2D-PAGE. The latter was attributed to intrinsic chemical modifications (Asn and Gln deamidation and Trp oxidation) in the case of Baccara LEAM (Isoform 1), which may contribute to protein architecture by maintaining the amphipathic and charge properties of the helical structure (Toltelet *et al.* 2007). Although isoforms 2, 3 and 4 have not been analysed by mass spectrometry, their 2D-PAGE profiles suggest they bear similar post-translational modifications, highlighting the importance of these spontaneous chemical events to protein function. The genes encoding the isoforms were very similar except for the presence of indels in the central repeated motif region. Indels do not alter the class A helical motif which is the functional signature of the proteins (Toltelet *et al.* 2007). Hence, the four LEAM isoforms might play similar roles in protecting the mitochondrial inner membrane during desiccation. In *Arabidopsis*, two paralogues encode putative orthologs of LEAM and both are dually targeted to mitochondria and plastids (Candat *et al.* 2014). Interestingly, these two proteins differ by several indels which do not affect their class A helix motifs. The conservation of the class A helix motif in LEAM proteins in spite of the presence of indels agrees with the observation that short internal indels generally do not have an impact on protein structure (Kim & Guo 2010). This is a case study that shows that within species, gene evolution has driven the occurrence of small indels in LEAM proteins that allow some variation in the length of the protein without affecting its structural features. It is noteworthy that LEA proteins do not display complex structural folds or enzyme activity and might therefore withstand slight structural variations in contrast to sHSPs (such as HSP22), which exhibit greater structural constraints. Most LEA proteins, such as LEAM, are low complexity polypeptides with repeated motifs (Tunnacliffe & Wise 2007), in which indel events are likely to occur (Bjorklund *et al.* 2006). In contrast, sHSPs such as HSP22 do not exhibit such repeated sequences, and indel events have not been reported in this chaperone family.

In *Arabidopsis*, almost half of the 51 *LEA* genes result from ancient genome duplication with conservation of the gene pairs, often in direct tandem repeats (Bies-Etheve *et al.* 2008; Hundertmark & Hinch 2008). Tandem duplications have been more frequently observed for genes involved in biotic and abiotic stress as shown in rice and *Arabidopsis* (Rizzon *et al.* 2006). Duplicated genes are often maintained when they provide a selective advantage through functional diversification and/or positive gene dosage (Blanc & Wolfe 2004; Zou *et al.* 2009). Whole genome duplication and tandem duplications may also have occurred in pea, such as in the legume model *M. truncatula* (Young *et al.* 2011), which could explain the presence of multiple *LEAM* copies encoding the same isoform. LEAM, like many LEA proteins, comprises several repeats, which could have arisen through duplication, when a segment copy is recombined near its origin. Bjorklund *et al.* (2006) have reviewed the possible

mechanisms (exon shuffling, DNA slipping) involved in the generation and expansion of domain repeats, which are common in eukaryote proteins. Repeats generally expand in the middle of the repeated regions, which would be consistent with the position and structure of the indel expanding in genes encoding isoform 2, 3 and 4. Furthermore, it has been demonstrated that disordered regions are overrepresented among indels and that indel residues are mainly exposed to the protein surface (Kim & Guo 2010; Light *et al.* 2013). LEA proteins, which are generally disordered (Jaspard *et al.* 2012), could therefore be prone to indel evolution. Indel as a source of LEA variability has been reported for pea and other species. Grosseindemann *et al.* (1998) have studied two pea dehydrin allelic variants (Dhn1 and Dhn3) differing in their electrophoretic mobility. Differences in polypeptide length were shown to result mainly from tandem duplication of 21 (*Dhn1*) and 24 (*Dhn3*) amino acids. In *M. truncatula*, size variations of a cold-inducible dehydrin gene were attributed to the variable occurrence of large indels (Remus-Borel *et al.* 2010). Interestingly, such a polymorphism was associated with improved freezing tolerance. In barley, three isoforms of the B19 protein (Em protein) have been described and differed by the number of repetition of a 20-amino acid motif located in the middle of the sequence. The *B19* genes were shown to be differentially expressed in response to ABA (Abscisic Acid), salt and osmotic stress in seed embryos, suggesting different functions during seed desiccation and environmental stress (Espelund *et al.* 1992). In animals, an interesting case of LEA gene evolution has been described in a desiccation-tolerant bdelloid rotifer (Pouchkina-Stantcheva *et al.* 2007). In these asexual organisms, two LEA genes, which are expected to have arisen from former alleles, have diverged to play complementary roles in stress survival. The two proteins that differ by a long 44-residue indel displayed contrasting structural behaviour upon transition from the hydrated to the dry state and differed in their capacity to interact with membranes or protect proteins during drying. Expression of the fluorescent-tagged LEA proteins in mammalian cells revealed they were routed in the secretory pathway and might be involved in intra- and extracellular protection mechanisms (Tripathi *et al.* 2012).

The core collection used in this study is part of a larger set that has been classified using protein and PCR-based markers (Baranger *et al.* 2004). Distribution of LEAM isoforms could not be related to the molecular marker-based classification of the 89 pea accessions. This suggests that the isoforms could have arisen very early before or during pea domestication. The preponderance of isoform 1 and 3 may result from selective and non-selective pea breeding and suggest they have been the main progenitors. Middle East, and particularly the region between Turkey and Iraq, is considered as the cradle of pea domestication (Weeden 2007). Primitive land races have been found also in Ethiopia and Egypt. The observations that Middle East and African native peas display mainly isoform 1 suggest that the corresponding gene may be the ancestral *LEAM* gene. If true, insertions rather than deletions in the coding sequence would explain the variability of LEAM protein. This hypothesis is in agree-

ment with Bjorklund *et al.* (2010) who studied the evolution of muscle nebulin, a protein with numerous repeated units that would have evolved from an invertebrate precursor through duplications of repeated domains. The fact that the gene orthologous to LEAM in *M. truncatula* is closer to isoform 1 is also in favour of the ancestry of that LEAM isoform. However, as the isoform 1 gene also exhibits several repeated units, we cannot exclude the possibility that isoform 1 gene has derived from a smaller gene with fewer repeats. The genes encoding isoform 2, 3 and 4, which harbour insertions, may have emerged through tandem segmental duplications in the isoform 1 gene. Tandem segmental duplications have been well documented in keeping with eukaryotes evolution (Koszul & Fischer 2009) and repeated regions strongly favoured such events (Bjorklund *et al.* 2010). The preponderance of isoform 2 and 3 over isoform 4 suggests that the region in which this indel occurred was more favourable to segmental duplication.

LEAM indels might be ancient, having even appeared before duplication events (whole genome duplication, single gene duplication) as only one *LEAM* sequence is present in a given pea genotype. Another possibility is that the original isoform 1 gene has been gradually eliminated after indel events in peas bearing the isoform 2, 3 or 4. In this case, the indels might have provided some selective advantages that remain to be unravelled. None of the isoform 2 genotypes are wild species and they are not native from the areas of pea origin and domestication. This suggests that this indel event has arisen later from genotypes harbouring the isoform 3 as this indel is a truncated version of indel 3. The pea core collection comprises mainly European cultivars and wild species and local populations are underrepresented. This situation is often encountered with pea core collections (Smýkal *et al.* 2012) and may have biased our analysis. A larger collection of wild pea accessions would be required to strengthen these hypotheses about *LEAM* gene evolution.

In summary, *LEAM* and *HSP22* are expressed at similar levels in seeds of all pea genotypes, but in contrast to *HSP22*, *LEAM* displays four isoforms differing in variable indels. The presence of the indels does not affect the class A helix structural motif of the protein, which is crucial for its function. This shows that *LEA* genes can evolve as a consequence of gene and sequence duplication events, yielding proteins that can retain their original structural and functional properties, like in the case of *LEAM* in pea or opening the path to neofunctionalization in other cases.

ACKNOWLEDGMENTS

We are grateful to Mike Ambrose for providing us with the core 16 of the JIC and to X. Zong, M. Vishnyakova, C. Coyne and A. Ramos-Monreal for providing us with some accessions from their national collections.

REFERENCES

- Baranger A., Aubert G., Arnau G., Laine A.L., Deniot G., Potier J., . . . Burstin J. (2004) Genetic diversity within *Pisum sativum* using protein- and PCR-based markers. *Theoretical Applied Genetics* **108**, 1309–1321.

- Bardel J., Louwagie M., Jaquinod M., Jourdain A., Luche S., Rabilloud T., . . . Bourguignon J. (2002) A survey of the plant mitochondrial proteome in relation to development. *Proteomics* **2**, 880–898.
- Battaglia M. & Covarrubias A.A. (2013) Late embryogenesis abundant (LEA) proteins in legumes. *Frontiers in Plant Science* **4**, 190.
- Benamar A., Tallon C. & Macherel D. (2003) Membrane integrity and oxidative properties of mitochondria isolated from imbibing pea seeds after priming or accelerated ageing. *Seed Science Research* **13**, 35–45.
- Benesch J.L.P., Ayoub M., Robinson C.V. & Aquilina J.A. (2008) Small heat shock protein activity is regulated by variable oligomeric substructure. *Journal of Biological Chemistry* **283**, 28513–28517.
- Bies-Etheve N., Gaubier-Comella P., Debures A., Lasserre E., Jobet E., Raynal M., . . . Delseny M. (2008) Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Molecular Biology* **67**, 107–124.
- Bjorklund A.K., Ekman D. & Elofsson A. (2006) Expansion of protein domain repeats. *PLoS Computational Biology* **2**, e114.
- Bjorklund A.K., Light S., Sagit R. & Elofsson A. (2010) Nebulin: a study of protein repeat evolution. *Journal of Molecular Biology* **402**, 38–51.
- Blanc G. & Wolfe K.H. (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *The Plant Cell* **16**, 1679–1691.
- Boudet J., Hoekstra F.A., Rogniaux H., Larre C., Satour P. & Leprince O. (2006) Comparative analysis of the heat stable proteome of radicles of *Medicago truncatula* seeds during germination identifies late embryogenesis abundant proteins associated with desiccation tolerance. *Plant Physiology* **140**, 1418–1436.
- Candat A., Paszkiewicz G., Neveu M., Gautier R., Logan D.C., Avelange-Macherel M.-H. & Macherel D. (2014) The ubiquitous distribution of late embryogenesis abundant proteins across cell compartments in Arabidopsis offers tailored protection against abiotic stress. *The Plant Cell* **26**, 3148–3166.
- Chauhan H., Khurana N., Nijhavan A., Khurana J.P. & Khurana P. (2012) The wheat chloroplastic small heat shock protein (sHSP26) is involved in seed maturation and germination and imparts tolerance to heat stress. *Plant, Cell & Environment* **35**, 1912–1931.
- Derocher A.E. & Vierling E. (1994) Developmental control of small heat-shock protein expression during pea seed maturation. *The Plant Journal* **5**, 93–102.
- Dure L. & Galau G.A. (1981) Developmental biochemistry of cottonseed embryogenesis and germination. *Plant Physiology* **68**, 187–194.
- Espelund M., Saeboe-Larssen S., Hughes D.W., Galau G.A., Larsen F. & Jakobsen K.S. (1992) Late embryogenesis-abundant genes encoding proteins with different numbers of hydrophilic repeats are regulated differentially by abscisic acid and osmotic stress. *The Plant Journal* **2**, 241–252.
- Grelet J., Benamar A., Teyssier E., Avelange-Macherel M.-H., Grunwald D. & Macherel D. (2005) Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. *Plant Physiology* **137**, 157–167.
- Grosselindemann E., Robertson M., Wilmer J.A. & Chandler P.M. (1998) Genetic variation in pea (*Pisum*) dehydrins: sequence elements responsible for length differences between dehydrin alleles and between dehydrin loci in *Pisum sativum* L. *Theoretical and Applied Genetics* **96**, 1186–1192.
- Hanania U., Velcheva M., Sahar N. & Perl A. (2004) An improved method for isolating high-quality DNA from *Vitis vinifera* nuclei. *Plant Molecular Biology Reporter* **22**, 173–177.
- Hoekstra F.A., Golovina E.A. & Buitink J. (2001) Mechanisms of plant desiccation tolerance. *Trends in Plant Science* **6**, 431–438.
- Hunault G. & Jaspard E. (2010) LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics* **11**, 221.
- Hundertmark M. & Hincha D.K. (2008) LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* **9**, 118.
- Jaspard E., Macherel D. & Hunault G. (2012) Computational and statistical analyses of amino acid usage and physico-chemical properties of the twelve late embryogenesis abundant protein classes. *PLoS ONE* **7**, e36968.
- Kalemba E.M. & Pukacka S. (2008) Changes in late embryogenesis abundant proteins and a small heat shock protein during storage of beech (*Fagus sylvatica* L.) seeds. *Environmental and Experimental Botany* **63**, 274–280.
- Kim R. & Guo J.T. (2010) Systematic analysis of short internal indels and their impact on protein folding. *BMC Structural Biology* **10**, 24.
- Koszul R. & Fischer G. (2009) A prominent role for segmental duplications in modeling eukaryotic genomes. *Comptes Rendus Biologies* **332**, 254–266.
- Lenne C. & Douce R. (1994) A low molecular mass heat-shock protein is localized to higher plant mitochondria. *Plant Physiology* **105**, 1255–1261.
- Light S., Sagit R., Ekman D. & Elofsson A. (2013) Long indels are disordered: a study of disorder and indels in homologous eukaryotic proteins. *Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics* **1834**, 890–897.
- Macherel D., Benamar A., Avelange-Macherel M.-H. & Tolleter D. (2007) Function and stress tolerance of seed mitochondria. *Physiologia Plantarum* **129**, 233–241.
- Manfre A.J., Lanni L.M. & Marcotte W.R. Jr. (2006) The Arabidopsis group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development. *Plant Physiology* **140**, 140–149.
- Menze M.A., Boswell L., Toner M. & Hand S.C. (2009) Occurrence of mitochondria-targeted late embryogenesis abundant (LEA) gene in animals increases organelle resistance to water stress. *Journal of Biological Chemistry* **284**, 10714–10719.
- Nakamoto H. & Vigh L. (2007) The small heat shock proteins and their clients. *Cellular and Molecular Life Science* **64**, 294–306.
- Pouchkina-Stantcheva N.N., McGee B.M., Boschetti C., Tolleter D., Chakrabortee S., Popova A.V., . . . Tunnacliffe A. (2007) Functional divergence of former alleles in an ancient asexual invertebrate. *Science* **318**, 268–271.
- Puhakainen T., Hess M.W., Makela P., Svensson J., Heino P. & Palva E.T. (2004) Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. *Plant Molecular Biology* **54**, 743–753.
- Remus-Borel W., Castonguay Y., Cloutier J., Michaud R., Bertrand A., Desgagnes R. & Laberge S. (2010) Dehydrin variants associated with superior freezing tolerance in alfalfa (*Medicago sativa* L.). *Theoretical and Applied Genetics* **120**, 1163–1174.
- Rizzon C., Ponger L. & Gaut B.S. (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. *PLoS Computational Biology* **2**, 989–1000.
- Rozen S. & Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology* **132**, 365–386.
- Siddique M., Gernhard S., von Koskull-Doring P., Vierling E., & Scharf K.D. (2008) The plant sHSP superfamily: five new members in Arabidopsis thaliana with unexpected properties. *Cell Stress and Chaperones* **13**, 183–197.
- Smykal P., Aubert G., Burstin J., Coyne C.J., Ellis N.T.H. & Flavell A.J. (2012) Pea (*Pisum sativum* L.) in the genomic era. *Agronomy* **2**, 74–115.
- Stupnikova I., Benamar A., Tolleter D., Grelet J., Borovskii G., Dorne A.J. & Macherel D. (2006) Pea seed mitochondria are endowed with a remarkable tolerance to extreme physiological temperatures. *Plant Physiology* **140**, 326–335.
- Sun W., Van M.M. & Verbruggen N. (2002) Small heat shock proteins and stress tolerance in plants. *Biochimica et Biophysica Acta* **1577**, 1–9.
- Tolleter D., Jaquinod M., Mangavel C., Passirani C., Saulnier P., Manon S., . . . Macherel D. (2007) Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation. *The Plant Cell* **19**, 1580–1589.
- Tolleter D., Hincha D.K. & Macherel D. (2010) A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state. *Biochimica et Biophysica Acta* **1798**, 1926–1933.
- Torok Z., Goloubinoff P., Horvath I., Tsvetkova N.M., Glatz A., Balogh G., . . . Vigh L. (2001) Synechocystis HSP17 is an amphitropic protein that stabilizes heat-stressed membranes and binds denatured proteins for subsequent chaperone-mediated refolding. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3098–3103.
- Tripathi R., Boschetti C., McGee B. & Tunnacliffe A. (2012) Trafficking of bdelloid rotifer late embryogenesis abundant proteins. *The Journal of Experimental Biology* **215**, 2786–2794.
- Tunnacliffe A. & Wise M.J. (2007) The continuing conundrum of the LEA proteins. *Die Naturwissenschaften* **94**, 791–812.
- Tunnacliffe A., Hincha D.K., Leprince O. & Macherel D. (2010) LEA proteins: versatility of form and function. In *Sleeping Beauties – Dormancy and Resistance in Harsh Environments* (eds E. Lubzens, J. Cerda & M. Clark), pp. 91–108. Springer, Berlin.
- Verwoerd T.C., Dekker B.M.M. & Hoekema A. (1989) A small-scale procedure for the rapid isolation of plant RNAs. *Nucleic Acids Research* **17**, 2362.

- Wang W., Vinocur B., Shoseyov O. & Altman A. (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science* **9**, 244–252.
- Waters E.R. (2012) The evolution, function, structure, and expression of the plant sHSPs. *Journal of Experimental Botany* **64**, 391–403.
- Weeden N.F. (2007) Genetic changes accompanying the domestication of *Pisum sativum*: is there a common genetic basis to the 'Domestication syndrome' for legumes? *Annals of Botany* **100**, 1017–1025.
- Wehmeyer N. & Vierling E. (2000) The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. *Plant Physiology* **122**, 1099–1108.
- Wehmeyer N., Hernandez L.D., Finkelstein R.R. & Vierling E. (1996) Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. *Plant Physiology* **112**, 747–757.
- Young N.D., Debelle F., Oldroyd G.E.D., Geurts R., Cannon S.B., Udvardi M.K., ... Roe B.A. (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**, 520–524.
- Zou C., Lehti-Shiu M.D., Thomashow M. & Shiu S.H. (2009) Evolution of stress-regulated gene expression in duplicate genes of *Arabidopsis thaliana*. *PLoS Genetics* **5**, e1000581.

Received 1 September 2014; received in revised form 17 October 2014; accepted for publication 21 October 2014

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. LEAM immunodetection in the 89 pea accessions seeds.

Figure S2. HSP22 immunodetection in the 89 pea accessions seeds.

Figure S3. LEAM coding sequences for the three selected genotypes Baccara, Melrose and Glacier.

Figure S4. Comparison of the intron found in LEAM genes variants.

Figure S5. Partial sequence of the LEAM gene found in the genotypes of the pea core collection.

Table S1. References of the *Pisum* accessions.

Table S2. Primers used for PCR amplification of *LEAM* genes.