RESEARCH PAPER

Identification of new gene expression regulators specifically expressed during plant seed maturation

Laurent Gutierrez1, Geneviève Conejero3, Mathieu Castelain1, Stéphanie Guénin2, Jean-Luc Verdeil3, Brigitte Thomasset4 and Olivier Van Wuytswinkel1,*

1 Laboratoire de Génomique Fonctionnelle des Plantes, Université de Picardie Jules Verne, Faculté des Sciences, 33 rue Saint Leu, F-80039 Amiens cedex, France
2 Centre de Ressources Régionales en Biologie Moléculaire, Université de Picardie Jules Verne, Faculté des Sciences, 33 rue Saint Leu, F-80039 Amiens cedex, France
3 Plateau d’Histocytologie et d’Imagerie Cellulaire Végétale, IFR 127, TA/40/02, CIRAD, Avenue Agropolis, F-34398 Montpellier cedex 5, France
4 Laboratoire de Génie Enzymatique et Cellulaire, UMR 6022 CNRS, Université de Technologie de Compiègne, BP 20529, F-60205 Compiègne cedex, France

Received 25 July 2005; Accepted 25 January 2006

Abstract

A cDNA-AFLP approach on Linum usitatissimum (flax) was used to identify genes specifically expressed during the seed maturation process. Among the 20 000 cDNA-AFLP tags produced, 486 were selected for their seed-specific expression during maturation. When compared with the publicly available databases, half of them presented some significant similarity with known plant sequences. The results obtained confirmed the accuracy of the approach as numerous genes previously described as being expressed exclusively in plant seeds were identified in this screen. The focus was on sequences similar to plant regulators involved in the control of gene expression, either at the transcriptional, post-transcriptional, or post-translational levels. Using a real-time RT-PCR approach, seed-specific expression kinetics were confirmed for 13 of these regulators that were never characterized for being expressed during seed maturation. Among these, a flax gene of the non-LEC1-like HAP3 family and a flax MYB factor were shown to be expressed in specialized tissues of flax embryo using an in situ hybridization approach. By expression kinetic comparison between these flax genes and their Arabidopsis counterparts, it was found that the new HAP3 gene should be related to a ubiquitous seed maturation mechanism, while a new MYB factor appears to be related to a more seed-specific maturation mechanism. These results demonstrate the utility of the flax database in not only identifying new genes expressed during seed maturation but also in being able to highlight the distinction between conserved and non-conserved seed maturation mechanisms.

Key words: cDNA-AFLP, flax, gene expression, maturation, seed.

Introduction

Seed development is a critical process in the life cycle of higher plants, allowing the connection between two distinct sporophytic generations leading to the maintenance of the species. According to several plant models, this process can be divided into morphogenesis and maturation stages (Goldberg et al., 1994). During morphogenesis, the 2n zygote will divide following a stereotyped cell division set leading to the formation of the heart-stage embryo (Laux et al., 2004). Maturation begins from early torpedo stage, proceeds through a period of embryo growth and seed filling, and ends with a desiccation phase after which the embryo enters into a quiescent state, thereby permitting its storage and survival in various environmental conditions (Bewley and Black, 1994). The storage compounds found in most seeds consist of carbohydrates, oils, or specialized storage proteins. These reserves are of major importance
because of their involvement in a successful early seedling growth but also because they represent the economic value of many crops. Whereas the main biosynthesis pathways involved in seed filling are well known, their regulation still remains widely uncharacterized. Indeed, while pathways for starch (James et al., 2003) or oil (Schmid and Ohlrogge, 1996) biosynthesis are well described, their accurate regulations are not understood (Hills, 2004).

Linum usitatissimum (flax) is an interesting crop regarding the diversity of its storage products. Indeed, flax seeds accumulate oil particularly rich in linoleic and linolenic acid which is quite unusual among oleaginous plants (Morris, 2003). Moreover, 2S, 7S, and 11S types of storage proteins (Marcone et al., 1994, 1998) are expressed in flax seeds. This diversity is also quite unusual. Generally, seeds of other plant species accumulate preferably only two of the three types of storage proteins. Finally, flax seeds are a source of fibre, mucilage carbohydrates (Cui, 2001), and remarkable secondary metabolites like phytooestrogen lignans (Axelson et al., 1982). All these components contribute to the interest shown in flax seed in the areas of nutrition and therapeutics, in which flax seed, for example, has been implicated in reducing metastasis formation (Dabrosin et al., 2002), cardiovascular risks (Bloedon and Szapary, 2004), and hypercholesterolemy (Lemay et al., 2002; Bhathena et al., 2003). As a consequence, flax appeared to be an extremely informative model for seed maturation study regarding the variety of biosynthesis pathways in its seeds, implying a higher diversity of genes involved in the seed maturation phase when compared with Arabidopsis. It must also be mentioned that the flax seed maturation phase extends for a longer time period when compared with Arabidopsis, allowing the harvest of seed batches at precise stages of maturation. Thus, this model also appeared to be attractive because of the possibility it gives to study the maturation process accurately at several successive development kinetic points. The last interesting point about the flax model is that flax is a dicot belonging to the Malpighiales order that contains the two major oilseed models, Arabidopsis and rapeseed (Savolainen and Chase, 2003). This phylogenetic distance should allow, by comparison with results obtained with the other plant models, identification of biological mechanisms involved in seed maturation which are (or are not) ubiquitous.

The main objective of this work was to set up an original database of seed maturation-specific genes, using a new plant model in order to gain insight into the diversity of biological mechanisms involved during this stage. With this in mind, a cDNA-AFLP analysis of flax seed maturation stage was performed using kinetic points covering the whole seed maturation stage, from embryo growth phase to mature seed. cDNA-AFLP (Bachem et al., 1996) is a valid alternative for quantitative genome-wide expression analysis, especially for species, like flax, for which genomic resources are lacking (Breyne et al., 2003; Volkmuth et al., 2003). Moreover, rather than just being an alternative, cDNA-AFLP analysis was shown to be an accurate tool for transcriptome study, allowing stronger specificity and sensitivity than microarrays analysis (Reijans et al., 2003). About 20 000 independent flax ESTs were analysed, among which 486 showed a specific expression during seed maturation. About 50% of these ESTs shared significant similarities with sequences identified from the publicly available databases. When translated, some of them were similar to proteins which had already been shown to be specifically expressed in plant seeds, indicating the accuracy of the cDNA-AFLP approach and allowing identification of the different phases of flax seed maturation. This work focuses on regulator genes in order to illustrate one area of interest in the information to be found in the database described here. Among the sequences characterized, 30 of them show some similarity with plant regulators involved in the control of gene expression, either at the transcriptional, post-transcriptional, or post-translational levels. Using a real-time RT-PCR approach, seed-specific expression kinetics were confirmed for all 13 of these regulators that had never been characterized for expression during seed maturation. Among these, a flax gene of the non-LEC1-like HAP3 family and a flax MYB factor were shown to be expressed in flax embryo-specialized tissues using an in situ hybridization approach. Expression kinetic comparison between these flax genes and their Arabidopsis counterparts was performed and interest in this transversal analysis in the distinction between ubiquitous and non-ubiquitous seed maturation mechanisms is discussed.

Materials and methods

Plant material

Linum usitatissimum plants (Barbara ecotype) were grown under a 16 h light period and day/night temperatures of 18/15 °C. Arabidopsis thaliana plants (Columbia ecotype) were grown under the same light period conditions and day/night temperatures of 22/20 °C. To harvest seeds (flax) or siliques (Arabidopsis) of defined age, individual flowers were tagged with coloured threads on the day of flower opening. Flax seeds, 10, 20, 30, 40, and 50 d after flowering (DAF), were dissected from the capsule, immediately frozen in liquid nitrogen, and stored at −80 °C. In the same way, Arabidopsis siliques were harvested and grouped into eight samples according to the number of days after flowering: 1–2, 3–4, 5–6, 7–8, 9–11, 12–14, 15–17, and 18–20. Flax seed used for in-situ hybridization experiments were dissected from capsules and immersed in fixation solution.

Histology

Lipids were stained using Nile red dye on fresh tissues and visualized on an LSM 510 Meta confocal microscope (Zeiss, Oberkochen, Germany), using an argon 488 nm laser. Proteins and polysaccharides were visualized on sections of flax embryos. Flax embryos were fixed overnight in freshly prepared 4% paraformaldehyde in phosphate buffer, then dehydrated in a graded ethanol series, and embedded in paraffin. Sections (8 μm thick) were double stained in periodic acid–Schiff (PAS) and naphthol blue-black (NBB). Polysaccharides were stained red with
cDNA-amplified fragment length polymorphism

Two sets of total RNA were prepared from two independent batches of flax seeds, of 6-month-old mature seeds, and of vegetative tissues including stem, leaves, and apical meristems. Tissues were ground in liquid nitrogen and, after a preliminary polyphenol and polysaccharide precipitation (Gehrig et al., 2000), RNA was extracted from the supernatant using a hot phenol purification protocol (modified after Verwoerd et al., 1989). For each condition, experiments were independently performed on the two RNA sets. Poly (A)+ RNA was independently performed on the two RNA sets. Poly (A)+ RNA was amplified step was performed using the primers 5'-CAGGACTGGC-5' and 3'-CTGACGCATGGAT-3'. The selective PCR fragments, and a PCR preamplification step was performed using the adaptor primers without selective nucleotides. The selective PCR reactions were systematically performed with both cDNAs from the two independent RNA sets, and all the 256 AFLP combinations were done. Polymorphism was analysed as described by Bachem et al. (1996). The bands of interest, appearing in duplicate, were cut from the gel with a surgical blade, eluted, and reamplified with the same primer set used for the initial amplification. The reamplified cDNAs were derived fragment sequence was compared against all sequences in the derived fragment sequence was compared against all sequences in the Arabidopsis network service (Altschul et al., 1996). Double-stranded cDNA was digested with the restriction enzymes TaqI and Asel. The adaptor primers 5'-GACGATGATCCTGAC-3' and 5'-TACT-CAGGACTGGC-5' (TaqI) and 5'-CTCTGACGCTGACC-3' and 3'-CTGACGCATGGAT-5' (Asel) were ligated to the restriction fragments, and a PCR preamplification step was performed using the adaptor primers without selective nucleotides. The selective PCR amplification step was performed using the primers 5'-GATGAGTGCTGACC-3' (5'-GACGATGATCCTGAC-3') and 3'-CTGACGCATGGAT-5' (Asel) (N represents A, T, G, or C). AmpliTaq DNA Polymerase (Applied Biosystems, Foster City, CA, USA). For each condition, reactions were systematically performed with both cDNAs from the two independent RNA sets, and all the 256 AFLP combinations were done. Polymorphism was analysed as described by Bachem et al. (1996). The bands of interest, appearing in duplicate, were cut from the gel with a surgical blade, eluted, and reamplified with the same primer set used for the initial amplification. The reamplified cDNAs were subcloned using the pGEM®-T Easy Vector system (Promega, Madison, WI, USA) and sequenced (MWG Biotech AG, Ebersberg, Germany). Database searches were performed at the NCBI World Wide Web server using the Basic Local Alignment Search Tool (BLAST) network service (Altschul et al., 1997). Each transcript-derived fragment sequence was compared against all sequences in the non-redundant database using the BLASTX program. Default parameters were used for all analyses. The same BLASTX analysis was performed on the FLAdb++ database (Samson et al., 2004), allowing an increased accuracy of the determination of the homology scores.

Real-time reverse transcriptase-PCR analysis

Total RNA was treated with amplification-grade DNase I (Invitrogen). First-strand cDNA was synthesized using Superscript II™ Reverse Transcriptase (Invitrogen). These cDNAs were used for PCR experiments using gene-specific primers designed with the LC Probe Design® software (Roche, Basel, Switzerland) from identified flax EST sequences or from available Arabidopsis sequences. All primers were optimized for being used with an annealing temperature of 60°C during PCR. Forward and reverse primers were used for producing a single amplicon from flax genes LuC1 (5'-ATATAAAGGGCGGCA-GTACTAGC-3' and 5'-TCCAACAACTACAACCGAGC-3'), LuC2 (5'-GAGAGGATTGTTCTCCGTGCA-3' and 5'-AAGCTCT-TCATTACCTCGTGC-3'), Lu0114 (5'-ACCTCTGGATGTAGG-3' and 5'-ATCCTGAGGTTTCCA-3'), Lu0146 (5'-CTGACCTTGGCCATGGGCA-3' and 5'-GTAGACCCGGAAAG-TGG-3'}, Lu09021 (5'-AGATGAGGGGTCACTG-3' and 5'-TCAC-GGCTTAAAGCACA-3'), Lu09022 (5'-AATGAACACTCTGGTG-3' and 5'-CGAAGCTTGATCATC-3'), Lu09105 (5'-AATGAACACTTGTTAGAGG-3' and 5'-CGAAGCTTGATCATC-3'), Lu10015 (5'-ACCGACAGAAGTTTTTCC-3' and 5'-GGAATTGGAATGTTCAACTGT-3'), Lu10143 (5'-ACACGGGTTGAA-3' and 5'-GGCCGCTCTGAA-3'), Lu1144 (5'-ATGGCGGATTTGTGTC-3' and 5'-GACCGCTCTCCCATAGT-3'), and Lu12076 (5'-TCCAAACTCCGCCAGC-3' and 5'-GGACGCTCTGGTGAC-3'). Real-time PCR was performed on a Roche LightCycler® using the FastStart DNA MasterPLUS SYBR Green I kit (Roche) in a final volume of 20 μl according to the manufacturer’s protocol. All reactions were prepared in duplicate and performed twice. For each sample and calibrator (vegetative control), the relative amount of a target gene and a reference gene allowing the normalization of small differences in template amounts, were determined. Crossing-point values, which are the PCR cycle numbers at which the accumulated fluorescent signal in each reaction crosses a threshold above background, were obtained with the LightCycler® software 3.5 (Roche) using the second derivative maximum method. Crossing-point values are a function of the amplification efficiency of the respective PCR. These data were then exported into the ReLQuant® software (Roche). This software provides efficiency-corrected, calibrator-normalized quantification results. Results are expressed as the target/reference ratio of the sample, divided by the target/reference ratio of the calibrator, and, therefore, are corrected for sample heterogeneities and detection-caused variances. The efficiency-corrected quantification performed by ReLQuant® is based on relative standard curves describing the PCR efficiencies of each target and the reference gene. The relative standard curves are determined and are used for each analysis. For each gene of interest, the same melting point temperature and size were observed for PCR products obtained by real time RT-PCR using cDNAs or the corresponding plasmid-cloned sequence as the template, indicating a low probability of false priming.

RNA in situ hybridization

One antisense Lu1146, riboprobe 240 bases long, one antisense Lu10015 riboprobe, 176 bases long, and one antisense GUS riboprobe, 200 bases long, were labelled with biotin (Invitrogen), MaxiScript, and BrightStar Psoralen-Biotin kit; Ambion, Austin, TX, USA). A probe concentration of 2.5 μg ml⁻¹ was used in an overnight hybridization of tissue sections 8 μm thick. Detection was performed with NBT-BCIP for 24 h (Vectastain ABC-AP kit; Vector Laboratories, Burlingame, CA, USA).

Results and discussion

Microscopic observations of flax embryo development

In order to visualize the evolution of its structure and determine the time points at which to gather seeds used in the cDNA-AFLP study described here, observations on flax embryos harvested at different developmental stages were performed using a microscope. The torpedo stage appears in...
embryos 10 DAF (Fig. 1A) still surrounded by endosperm. The 20 DAF embryo (Fig. 1B) fills the seed sac and reaches its full size. Between 20 and 40 DAF, cotyledon thickness increases (Fig. 1C). At this stage, the green colour of the embryo softens and pigmentation appears in the peripheral layer of the inner integument (Fig. 1D) to give the characteristic brown colour of the desiccated seed at 50 DAF. Therefore, 10 DAF is the hinge step, when embryo morphogenesis is over and the embryo growth phase indicates the beginning of the seed maturation stage. Then, during flax embryogenesis, maturation seems to extend from 10 to 50 DAF.

The diversity and cellular organization of flax seed storage products were revealed by using specific dyes on seed sections (Fig. 1). Mature seeds (Fig. 1E, F) accumulate all the three major classes of storage compounds: lipids, proteins, and polysaccharides. The complex evolution of polysaccharide stock (pink colour in Fig. 1A–C) during embryo maturation and how it is stored in the mature seed are of particular interest (Fig. 1E). At 10 DAF, polysaccharides are stored in two main forms: starch (few amyloplasts) in the endosperm; mucilage in the torpedo embryo. It must be noticed that the polysaccharide-specific pink dye is almost absent at 20 DAF, indicating that starch and mucilage storage forms are transitory (Fig. 1B). Moreover, this stage presents an important synthesis of storage proteins (blue colour). However, while showing an increased storage protein content, the 40 DAF embryo contains polysaccharides mainly as mucilage, starch being completely absent (Fig. 1C, D). Finally, in the mature embryo (Fig. 1E), stored polysaccharides are still present as mucilage but also in a hollow structure in the centre of the two or three storage protein bodies present in each embryo cell. The nature of these protein body polysaccharides is unknown. Such structures, as far as is known, have not been described before and potentially constitute an unidentified polysaccharide storage form.

A cDNA-AFLP screen identified 486 flax ESTs specific to seed maturation

cDNA-AFLP analysis was performed as described in the Materials and methods, using developing flax seeds 10, 20, 30, 40, and 50 DAF. Vegetative tissues (stem, leaves, apical meristems) and mature seeds were also included as controls. It must be mentioned that the 10 DAF samples contain a large amount of maternal tissues, including the endosperm. The cDNA expression profiles were determined by PCR-selective amplification using 256 different primer combinations, and ~20 000 cDNA fragments were screened. For each condition, two cDNA-AFLP reactions were performed from two sets of independently prepared mRNAs samples. Comparison of fingerprints obtained from the five seed populations and the two controls allowed identification of transcript-derived fragments, called ESTs in the following text, from mRNAs that accumulated specifically during flax seed maturation (Fig. 2). Most ESTs showed similar levels of accumulation in each of the two independently repeated reactions.

A total of 486 ESTs specific to seed maturation were extracted from the polyacrylamide gels, reamplified, and subcloned for sequencing. The sequence obtained for each

---

Fig. 1. Distribution of storage compounds in flax embryo. (A–C) Longitudinal sections of flax embryos at different developmental stages: 10 DAF (A), 20 DAF (B), and 40 DAF (C). (D) Enlargement of the area framed in (C). (E, F) Cotyledon storage cells of mature seed. (A–E) Light micrographs using bright-field. (F) Confocal micrograph. Localization of storage products using specific dyes: (A–E) for proteins (naphthol Blue Black: blue dye), for polysaccharides (Schiff: pink dye), and (F) for lipids (Nile Red: yellow fluorescence). cot, Cotyledon; cx, cortex; en, endosperm; in, inner integument; mi, micropyle; mu, mucilage; pc, procambium; pl, pigment layer; sam, shoot apical meristem; sl, storage lipid; sp, storage proteins; spo, storage polysaccharides; st, starch. Scale bars in D–F=20 μm.
tag was compared with the GenBank non-redundant public sequence database using the BLASTX program (Altschul et al., 1997). Among these, 256 (52.7%) showed close matches [BLASTX expectation values (E) of $<10^{-3}$] to database entries with assigned identities (see supplementary data at www.jxb.oupjournals.org). This result is in agreement with previous cDNA-AFLP studies in which around 50% of the ESTs identified had no or poor identity with sequences available so far (Durrant et al., 2000; Milioni et al., 2002; Kwon et al., 2004; Mao et al., 2004). These sequences are classified into 11 groups based on functional categories established for Arabidopsis (Arabidopsis Genome Initiative, 2000). Figure 3 gives the proportion of ESTs in each category.

**Marker genes cDNA-AFLP patterns identify the different phases of flax seed maturation**

Seed maturation goes through three successive phases: embryo growth phase, storage phase, and late maturation phase, including acquisition of desiccation tolerance and dormancy induction. Expression profiles of marker genes identified by cDNA-AFLP show the right staging of flax seed maturation. The homology scores for some of these sequences and their corresponding cDNA-AFLP patterns are given in Fig. 4.

The flax EST Lu06073 similar (E=1e-14) to ABI3 (Giraudat et al., 1992) confirms that the five kinetic points used in the present cDNA-AFLP study are covering the entire flax seed maturation stage. Indeed, ABI3 is one of the major transcription factors controlling maturation in Arabidopsis seeds (Ooms et al., 1993; Parcy et al., 1994, 1997; Nambara et al., 1995), expressed from the end of the embryo growth phase and remaining at a high expression level during the entire maturation stage (Raz et al., 2001).

In the microscope observations (Fig. 1), the flax seed maturation stage was identified as extending from 10 to 50 DAF. These observations are sustained by the cDNA-AFLP analysis since the Lu06073 expression pattern indicates that this gene is expressed at a low level in seed at 10 DAF, reaches its highest expression level from 20 to 50 DAF, and shows no expression in mature seeds. Flax ESTs sequences Lu16014, Lu07021, Lu12051, Lu11041, and Lu04051 are similar to APETALA2 (E=3e-47), NAM-like (E=2e-06), Cyc1B (E=7e-13), CYSTEINE PROTEINASE (E=5e-33), and HSP60 CHAPERONIN (E=2e-09), respectively, from Arabidopsis. As shown in Fig. 1A, the 10 DAF flax embryo is in the growth phase that marks out the beginning of maturation. This timing is confirmed by these five genes known for their involvement during this phase in embryo or in maternal tissues (Jofuku et al., 1994; Souer et al., 1996; Aida et al., 1997; Apuya et al., 2001; Maes et al., 2001; Raz et al., 2001; Takada et al., 2001; Wan et al., 2002; Vroemen et al., 2003; Dong et al., 2004; Weir et al., 2004) and for which expression is specifically found in 10 DAF flax seeds (Fig. 4). Many ESTs in the database share similarities with legumin 11S (23 ESTs), vicillin 7S (four

![Fig. 2. cDNA-AFLP autoradiogram showing the accumulation patterns of ESTs from flax seed during seed development. Templates were derived from vegetative tissues (V), from seeds 10, 20, 30, 40, and 50 DAF and from mature seed (MS). For each condition, two independent cDNA-AFLP reactions were loaded side by side. Lanes are in groups of 14; each group was amplified using a different combination of primers with two selective nucleotides. Various expression patterns can be detected: constitutive expression (i), seed specific expression (ii), and vegetative expression (iii).](http://jxb.oxfordjournals.org/)

![Fig. 3. Classification of 486 seed-specific flax ESTs. On the basis of BLASTX expectation (E) values of $<10^{-3}$, 210 sequences were assigned to functional categories. Among these, 30 sequences (6.2%) share similarity with gene expression regulators.](http://jxb.oxfordjournals.org/)
ESTs), and albumin 2S (eight ESTs) confirming the presence of these three kinds of storage proteins in flax seed as previously shown (Marcone et al., 1994, 1998). Two additional ESTs similar to oleosins are found, as well as the Lu12074 EST similar to the FAD2 gene (E=1e-25). Oleosins and FAD2 are required for lipid synthesis and storage in seeds (Heppard et al., 1996). For each kind of storage protein and for oleosins, a representative cDNA-AFLP pattern is shown (Fig. 4). Expression kinetics of these genes shows that the storage phase of flax seed maturation lasts from 20 to 30 DAF. Flax ESTs sequences Lu13141 and Lu10071 are similar to LATE EMBRYOGENESIS ABUNDANT (LEA) (E=1e-31) and HSP70 (E=1e-28), respectively, from Arabidopsis (Fig. 4). These genes, expressed during the desiccation phase of seed maturation (DeRocher and Vierling, 1995; Sung et al., 2001; Wise and Tunnaciffe, 2004), allow identification of 30–50 DAF as the late maturation phase of flax seed. Different expression patterns are found among these 13 regulators, from the embryo growth phase to the late maturation phase.

Seed-specific expression is validated for each of them using a real-time RT-PCR approach performed with a third set of independently prepared mRNA samples (Fig. 5A–C), showing a good correlation with the cDNA-AFLP pattern. To normalize each gene expression level between the different samples, amplification of a sequence from a reference constitutive gene was performed. The flax LuC1 and LuC2 EST sequences which showed a constitutive cDNA-AFLP pattern were used (Fig. 5D). LuC1 is similar to an Arabidopsis NADH:UBIQUINONE OXIDOREDUCTASE (E=1e-24) and LuC2 is similar to an Arabidopsis DSBA OXIDOREDUCTASE family cDNA (E=6e-23). A constant level of LuC1 expression was validated in real-time RT-PCR when normalized against the LuC2 expression level.

Identification of new gene expression regulators expressed during flax seed maturation and validation of their expression patterns by real-time RT-PCR

Many of the 256 seed maturation-specific ESTs identified in the database (see supplementary data at www.jxb.oupjournals.org) were never characterized as either playing a role in plant seed development or having a seed-specific expression. This amount of new expression data implies the existence of many mechanisms involved in seed maturation that still remain unknown. In order to validate information contained in the database described here and to highlight its interest, the focus was on ESTs similar to gene expression regulators. Among the 30 ESTs identified in this functional class, 13 share similarities with Arabidopsis regulators, known for their involvement in gene expression regulation at the transcriptional, post-transcriptional, or post-translational levels, and that had never been shown to be expressed in a plant seed. Different expression patterns are found among these 13 regulators, from the embryo growth phase to the late maturation phase.

Fig. 4. cDNA-AFLP patterns of growth phase or maturation stage marker genes. Duration of each of the flax seed development stages was deduced from the pattern of expression of listed genes and is indicated at the bottom of the figure (see text for details).
Fig. 5. Comparison of expression patterns of regulator genes obtained by cDNA-AFLP and real-time RT-PCR. (A–C) For each gene, a graph showing its expression kinetically assessed by real-time RT-PCR is placed above its cDNA-AFLP pattern. The level of each gene expression was normalized with LuCl and expressed with a value relative to the gene expression level in the vegetative control. The cDNA-AFLP patterns were confirmed for all these 13 genes. (A) Transcriptional regulators; (B) post-transcriptional regulators; (C) post-translational regulators; (D) constitutive cDNA-AFLP and real-time RT-PCR patterns of LuC1 and LuC2. Real-time RT-PCR data shown represent mean values obtained from four independent amplification reactions and the error bars indicate ± standard deviation.
Identification of new transcription factors expressed during flax seed maturation

The flax EST sequence Lu12076 is similar to a MYB factor (E=3e-06) and shows strong expression only during the embryonic growth phase (10 DAF) (Figs 4, 5A). When compared with the Arabidopsis MYB family, the Lu12076-translated sequence shares strong similarity with a subfamily of the R2R3 MYBs characterized by a C-terminal LNL(E/D)L motif. It must be mentioned that the similarity zone between the flax and the Arabidopsis sequences is located at the C-terminus and contains the LNLEL motif (data not shown). In Arabidopsis, this subfamily contains four members including AtMYB4 which was shown to be involved in plant UV protection by controlling the expression of sinapate ester sunscreens (Jin et al., 2000). The Lu12076 EST is detected very early in the flax seed maturation process and perhaps during the morphogenesis stage. Only one R2R3 MYB factor called TRANSPARENT TESTA 2 (AtMYB123) had been shown previously to be expressed during early embryo development, from zygote to early torpedo stage (Nesi et al., 2001). TT2 acts as a major factor for the determination of the BAN mRNA expression pattern, BAN being involved in tannin accumulation in seed (Debeaujon et al., 2003). However, TT2 seems to be of a distinct MYB subfamily rather than the one containing AtMYB4. This result could indicate the existence of a new class of MYB factors involved in early plant seed maturation.

Flax EST sequences Lu1142 and Lu1146 are similar to the same HAP3 subunit of the CCAAT-box binding transcription factor and show specific expression during the embryonic growth phase and up to the end of the storage phase (10–30 DAF) (Fig. 5A). HAP3 is one of the three subunits of the CCAAT-box binding transcription factor (CBF) (for reviews, see Maity and de Crombrugghe, 1998; Mantovani, 1999). The HAP3 protein contains three regions, the A, B, and C domains, with the central B domain conserved throughout eukaryotes (Li et al., 1992; Xing et al., 1993; Sinha et al., 1996). Arabidopsis LEC1 shares significant sequence similarity with the HAP3 subunit of the CCAAT-box binding transcription factor (Lotan et al., 1998; Lee et al., 2003). LEC1 is one of the major genes required for normal development during both morphogenesis and maturation phases in plant embryos (reviewed by Harada, 2001). The 10 Arabidopsis HAP3 genes (AHAP3) can be divided into two classes: the LEC1-type and non-LEC1-type (Kwong et al., 2003). At present, only LEC1-type AHAP3 genes were shown to be involved in embryogenesis. The two flax ESTs, Lu11142 and Lu11146, are both similar to non-LEC1-type AHAP3 (E=9e-22 and E=6e-23, respectively), sharing much less similarity with LEC1-type AHAP3 (E=1e-08, data not shown). This could indicate a role for the non-LEC1-type HAP3 factors in the plant embryo maturation process.

The flax Lu10015 EST sequence is similar to another MYB factor (E=2e-09) and is strongly expressed during the storage phase, but more precisely close to the end of this phase (30 DAF) (Figs 4, 5A). The translated Lu10015 sequence is highly similar to a subfamily of MYB-like SHAQKYF DNA-binding domain protein consisting of five members in Arabidopsis (data not shown). This DNA-binding domain, restricted to plant proteins, is often associated with a response regulator domain (Rose et al., 1999). These factors are characterized, in part, by a well-conserved motif SH(ALKYQY(RF) at the C-terminal end of the DNA-binding domain. The translated Lu10015 EST contains this domain (data not shown).

The flax Lu05113 EST sequence presents a clear similarity with the C-terminus of CONSTANS-LIKE1 (COL1) (E=2e-07) and is also strongly expressed during the storage phase, but more precisely close to the end of this phase (30 DAF) (Figs 4, 5A). COL1, which codes for an Arabidopsis zinc-finger transcription factor, was characterized as being a regulator of circadian rhythm, its expression being itself regulated by the circadian clock (Ledger et al., 2001). COL1 is part of the CONSTANS-like factor family which contains 16 members (Griffiths et al., 2003). The translated Lu05113 EST shows at its N-terminus a part of the CONSTANS factor-specific CCT domain while its C-terminus presents a high similarity with an extension only found in six of the Arabidopsis CONSTANS proteins [CO and COL(1–5); data not shown]. No gene of this family has ever been described as showing a seed-specific expression. Differences in Lu05113 expression levels are too low to claim this gene is seed specific, but the present result indicates that, at least in flax, a CONSTANS-like factor could also be involved in the maturation process. It can be proposed that seed-specific expression of CONSTANS-like genes in flax could be an adaptation to couple this process to the environment, especially to photoperiod. Indeed, flax seed development spans at least 8 weeks, showing why its comparison with a plant such as Arabidopsis, for which this process takes only 3 weeks, is of interest.

Finally, the Lu09021 and Lu06141 flax ESTs are expressed during the late maturation phase (Figs 4, 5A). Lu09021 is similar to a single Arabidopsis ZINC-FINGER TRANSCRIPTION FACTOR (E=6e-15) and is expressed from 30 to 50 DAF but its transcript levels are drastically decreased in mature seed. Lu06141, similar to a HISTONE2B (H2B) (E=3e-15), remains strongly expressed from 30 DAF to mature seed. It was shown that H2B, once ubiquitinylated, plays an important role in the trans-histone methylation of histone H3, a modification with close ties to the regulation of gene expression (reviewed by Osley, 2004).
Identification of new post-transcriptional regulators expressed during seed maturation

The Lu11064 flax EST is similar to AGO1-like (E=6e-22) and is slightly more expressed during the growth phase and the first half of the storage phase (10–20 DAF) (Figs 4, 5B). The Arabidopsis agol mutant phenotype is mainly characterized by several developmental defects (Bohmert et al., 1998). AGO1 was described as an important factor involved in miRNA production as it participates to the formation of the RISC complex (Vacheret et al., 2004). Concerning the role of AGO1 factors in plant development, data were mainly obtained on the leaf development model (Palatnik et al., 2003; Kidner and Martienssen, 2004). The present data suggest that production of miRNA through AGO1-like factors activity could also be involved in plant seed maturation.

The Lu09022 and Lu09105 flax ESTs, both similar to an mRNA CLEAVAGE FACTOR subunit (E=1e-20 and E=1e-28, respectively), are strongly expressed at the end of the storage phase (30 DAF) and then remain expressed, but more weakly, till the end of the embryo development process (Figs 4, 5B). These proteins are part of the cleavage factor complex involved in the 3′ splicing and polyadenylation of mRNAs. A development-specific expression of such factors looks surprising at first. However, factors involved in the same process were shown to play a role in floral induction through the autonomous pathway by controlling the 3′ processing of specific transcripts (Simpson et al., 2003). Moreover, mRNA cleavage-associated factors were recently shown to be expressed in the Arabidopsis flower and to be essential for its embryo development (Xu et al., 2004). The embryo expression pattern characterized in this study for the Lu09105 and Lu09022 flax ESTs could indicate that a related mechanism is involved during the late phase of plant embryo development. Following the same trend, the Lu08041 flax EST is expressed during the late maturation phase, from 40 to 50 DAF, and its transcript levels are drastically decreased in mature seed (Figs 4, 5B). This EST is similar to a U5 SMALL NUCLEAR RIBONUCLEOPROTEIN HELICASE (U5sn RNP) (E=2e-63). U5 snRNP, in association with other small ribonucleoprotein particles and non-snRNP proteins, is one element of the spliceosome (for a review, see Will and Lührmann, 2001).

Identification of new post-translational regulators expressed during seed maturation

Flax EST Lu05142 and Lu05146, similar to FLAVIN-BINDING KELCH REPEAT, F-BOX (FKF1) (E=3e-29 and E=1e-25, respectively), are expressed from 40 DAF, but their mRNA levels increased until reaching a maximum in mature seed (Figs 4, 5C). F-box proteins are part of the SCF complex involved in the targeted ubiquitylation process allowing degradation of specific proteins by the 26S proteasome (Vierstra, 2003). In the Arabidopsis genome, around 750 F-box coding genes were found, indicating the importance of post-translational regulations in plants. Involvement of an F-box-containing SCF complex in the auxin transduction pathway is now well demonstrated (Dharmasiri and Estelle, 2002) as well as their participation in the response to other phytohormones, in cell cycle control, as well as in photomorphogenesis and plant circadian clock functioning (Sopers et al., 2004; Schwechheimer and Calderón Villalobos, 2004). Within the plant F-box family, a small subgroup comprising FKF1, ZTL, and LKP2 is characterized by several C-terminal Kelch repeats. The translated Lu05142 and Lu05146 paralogous ESTs present high similarities with this specific C-terminus Kelch repeat. FKF1, shown to modulate the Arabidopsis circadian clock, contains an F-box domain characteristic of proteins that direct ubiquitin-mediated degradation and a PAS domain similar to the flavin-binding region of some photoreceptors (Nelson et al., 2000). It must be noticed that Lu05142 and Lu05146 are detected at the late maturation phase but also in the mature desiccated seed. Moreover, it was shown recently that maintenance of an mRNA stock within the mature seed is more important for germination success than mRNA neosynthesis during this process (Rajjou et al., 2004). Thus, these two putative flax F-box proteins could be involved, if not in the seed development process, perhaps in seed germination. In this case, the hypothesis of a possible involvement of targeted protein degradation during the early seed germination process could be supposed. Considering the role played by FKF1, which shows the strongest similarities with Lu05142 and Lu05146, in modulating the Arabidopsis circadian clock by means of light (Nelson et al., 2000), such a hypothesis could be related to the control by light of the end of dormancy and early seedling initiation.

Transcript localization of previously identified HAP3 and MYB factor by in situ hybridization on flax embryo

With the aim of going further in discovering new mechanisms involved in seed maturation, focus was on the three identified flax putative transcription factors expressed during the storage phase (20–30 DAF): Lu10015, similar to a MYB factor, and Lu11142 and Lu11146, both similar to a HAP3 subunit of the CCAAT-box binding transcription factor. For this purpose, in situ hybridization experiments were done to locate their expression specificities at the tissue level. These experiments were performed using a labelled antisense RNA probe on 30 DAF flax embryo sections, on which hybridization was detected by a purple coloration of the cells (Fig. 6). Flax EST Lu11142 and Lu11146 tally with the same part of the HAP3 genes and, while their sequences show sufficient differences to show that they are derived from two different genes, they present strong nucleic acid similarities (~92% similar nucleic acids, not shown). For this reason, only one Lu11146
probe was used, assuming it is likely to hybridize to both \textit{Lu11146} and \textit{Lu11142} mRNAs. In each experiment, a hybridization negative control was performed using a labelled \textsc{gus} antisense RNA probe.

\textit{Lu10015} transcripts are localized in cotyledons (Fig. 6A), with a strongest hybridization signal at their top end. A hybridization signal also appears in the cortex but is confined in some cell lines, some others presenting no signal (Fig. 6B). As far as is known, such \textsc{myb} factor transcript localization is original and has never been described in a plant embryo during seed maturation.

\textit{Lu11146} transcripts are localized in the procambium of the embryo axis level, in the root cap (Fig. 6C), and in the three upper layers of the shoot apical meristem (Fig. 6D). This tissue-specific expression is radically different from the one of the \textsc{lec1}-type \textit{AHAP3} (\textit{LEC1} and \textit{L1L}), shown to be mainly expressed in the outer cell layers of the whole embryo (Lotan \textit{et al}., 1998; Kwong \textit{et al}., 2003). The accuracy of their proven tissue specificity confirms that these \textit{Lu11146 \textsc{hap3}} transcripts are not of the \textsc{lec1}-type. Thus, it appears that some non-\textsc{lec1}-type \textit{HAP3s} are involved in a previously unknown gene expression regulation mechanism during seed maturation.

\section*{Gene expression of \textit{HAP3} and \textsc{myb} factor \textit{Arabidopsis} counterparts during silique development}

Results concerning \textsc{myb} factor and \textit{HAP3} genes assess the existence of previously unknown mechanisms acting during seed maturation. Are these mechanisms conserved within plant species or are they restricted to some species like \textit{flax}? The use of \textit{flax} as a model is of interest to answer such questions. Indeed, by comparison of the present results with those obtained on another plant model phylogenetically far removed, like \textit{Arabidopsis} (see Introduction), it might be possible to distinguish ubiquitous from non-ubiquitous seed maturation mechanisms.

Kinetic expression during silique development was measured by real-time RT-PCR for \textit{Arabidopsis} genes \textit{At2g37060}, a non-\textsc{lec1}-type \textit{AHAP3} showing the most sequence similarity with \textit{Lu11142} and \textit{Lu11146} (\textit{E}=9e-22 and \textit{E}=6e-23, respectively), for the nine other \textit{AHAP3} including \textit{LEC1} and \textit{L1L}, and for \textit{At5g02840} and \textit{At3g09600}, two \textsc{myb} factors showing the most sequence similarity with \textit{Lu10015} (\textit{E}=1e-21 each).

As expected regarding previous studies (Lotan \textit{et al}., 1998; Kwong \textit{et al}., 2003), \textit{LEC1} and \textit{L1L} show a strong expression during silique development (Fig. 7). Among all other non-\textsc{lec1}-type \textit{AHAP3s}, only \textit{At2g37060} is specifically expressed in siliques. Its expression kinetic is wave-shaped, showing a first over-expression peak around 5–6 DAF and a second highest peak around 15–17 DAF (Fig. 7). This result indicates that, contrary to \textit{LEC1} and \textit{L1L}, \textit{At2g37060} is specifically expressed in siliques not only during the morphogenesis stage but also during the late maturation stage of the seed. This expression specificity, in correlation with previous \textit{flax} results, sustains the hypothesis that non-\textsc{lec1}-type \textit{HAP3s} could be involved in a seed maturation mechanism of gene expression regulation that should be ubiquitous and that is different from those related to \textit{LEC1} or \textit{L1L}.

On the contrary, \textsc{myb} factor \textit{Lu10015} appears to be related to a non-ubiquitous gene expression regulation mechanism. Indeed, neither \textit{At5g02840} (Fig. 7) nor \textit{At3g09600} (not shown) show specific expression during silique development.

\section*{Concluding remarks}

With the intention of improving the knowledge of seed maturation mechanisms, several studies were carried out,
using a genome-wide expression analysis approach on seed (Girke et al., 2000; White et al., 2000; Lee et al., 2002; Ruuska et al., 2002; Suh et al., 2003; Dong et al., 2004; Sreenivasulu et al., 2004). These studies were performed on a small group of monocot and dicot plant models, using a macro- or microarray analysis approach, which considerably improved the amount of data obtained because they provide a wide vision of transcriptome variations during this essential step in a plant’s lifespan. However, despite its ability to characterize, for example, functional metabolic pathways, such approaches are not always the ones most adopted to identify new genes on the basis of their expression specificities. This observation is especially relevant when considering gene expression regulators which are often expressed at low levels and belong to multigenic families in which specific expression of one gene is difficult to distinguish from the others. As an example, these approaches did not allow the identification of new seed-specific regulators. WRINKLED1, the only transcription factor characterized so far (in Arabidopsis) to be specifically involved in regulating the synthesis of some storage products in the embryo (Focks and Benning, 1998; Cernac and Benning, 2004) was identified through a classic mutant approach. WRINKLED1 is clearly a prominent factor regulating central lipid and carbohydrate metabolism during seed maturation (Ruuska et al., 2002), but is possibly not the only one.

The aim of this present study was to identify new genes specifically expressed during the seed maturation phase. In order to make up for specific limitations of previous studies, use was made of (i) a new plant model that offers several advantages for studying seed maturation; (ii) kinetic points covering the whole seed maturation stage, from embryo growth phase to mature seed; and (iii) the cDNA-AFLP technique as a tool for gene expression analysis. This approach led to development of a database containing ESTs from 256 genes, among which most had not previously been characterized for such expression.

With the purpose of validating information contained in the present database and to highlight its interest, this study focused on gene expression regulators. The first interest in this database was to identify new genes expressed during seed maturation and, then, to underline the existence of seed maturation mechanisms still unknown. While such regulators are usually characterized with difficulty by transcript profiling tools because of their multigenic family membership (e.g. in Arabidopsis ~250 MYB factors genes and ~750 F-box genes) or their low expression levels, 13 regulators were identified that had never been characterized before for their expression either in seed maturation or in early seedling development. Moreover, some of them are related to mechanisms never described during seed maturation: circadian clock regulation and post-transcriptional and post-translational regulations. These results show the originality of the information found on seed maturation. These new regulators might be involved in still unknown seed maturation mechanisms as was assessed for HAP3 and MYB factor genes, regarding the original localization of their transcripts. The second interest in this database was the development of a flax gene database, a plant relatively phylogenetically far removed from Arabidopsis. Indeed, for each gene of interest, a transversal analysis performed between flax and Arabidopsis allows genes involved in ubiquitous seed maturation mechanisms to be distinguished, as the new non-LEC1-type HAP3 identified in
this work, and genes involved in more specific seed maturation mechanisms, as the flax Lu10015 putative MYB factor identified in this work. New genes involved in ubiquitous seed maturation mechanisms, conserved within these species, were identified thanks to the strong sensitivity and specificity of the cDNA-AFLP analysis. New genes involved in more specific seed maturation mechanisms, restricted to a limited number of species, were identified due to the unusual flax model used, allowing insight to be gained into the diversity of biological mechanisms involved during seed maturation. Therefore, this database could constitute a very informative tool for studying new seed maturation mechanisms.

Supplementary data

The flax embryo EST database described in this paper is available online. In this database, each of the 256 flax ESTs identified was associated with its closest similar sequence, a description of the encoded protein function, the corresponding BLASTX (E) value and its cDNA-AFLP pattern under a clustered form. Flax ESTs were named Lu (for Linum usitatissimum) followed by a code number corresponding to the AFLP fragment. Similarities shared were preferentially searched within Arabidopsis proteins and eventually with other plant proteins. Clusters are shown in three levels of grey intensity to illustrate differences in EST accumulation levels.

Acknowledgements

This work was supported by Alternatech and the Conseil Régional de Picardie, Amiens. Many thanks to Sonia Rippa, Carol Darcq, Morgan David, Fanny Henrion, and Romain Hecquet for EST cloning assistance, to Stéphanie Troufland for supplying flax plants, to François Guérineau and Jérôme Pelloux for helpful discussions and for reviewing the manuscript, and to Philippe Grappin for providing the protocol of the cDNA-AFLP method.

References


