

Evolution of the Group 1 late embryogenesis abundant (*Lea*) genes: analysis of the *Lea* B19 gene family in barley

Robin A.P. Stacy¹, Mari Espelund¹, Stein Sæbøe-Larssen¹, Kristin Hollung², Even Helliesen¹ and Kjetill S. Jakobsen^{1,*}

¹Division of General Genetics, University of Oslo, P.O. Box 1031, Blindern, N-0315 Oslo, Norway (*author for correspondence); ²Division of Botany, University of Oslo, P.O. Box 1045, Blindern, N-0315 Oslo, Norway

Received 26 September 1994; accepted in revised form 30 May 1995

Key words: ABA-inducible genes, coding region repeats, embryo-specific gene family, evolution, *Hordeum vulgare* L., phylogenetic analysis

Abstract

The highly conserved Group 1 late embryogenesis abundant (*Lea*) genes are present in the genome of most plants as a gene family. Family members are conserved along the entire coding region, especially within the extremely hydrophilic internal 20 amino acid motif, which may be repeated. Cloning of *Lea* Group 1 genes from barley resulted in the characterization of four family members named B19.1, B19.1b, B19.3 and B19.4 after the presence of this motif 1, 1, 3 and 4 times in each gene, respectively. We present here the results of comparative and evolutionary analyses of the barley Group 1 *Lea* gene family (B19). The most important findings resulting from this work are (1) the tandem clustering of B19.3 and B19.4, (2) the spatial conservation of putative regulatory elements between the four B19 gene promoters, (3) the determination of the relative 'age' of the gene family members and (4) the 'chimeric' nature of B19.3 and B19.4, reflecting a cross-over or gene-conversion event in their common ancestor. We also show evidence for the presence of one or two additional expressed B19 genes in the barley genome. Based on our results, we present a model for the evolution of the family in barley, including the 20 amino acid motif. Comparisons of the relatedness between the barley family and all other known Group 1 *Lea* genes using maximum parsimony (PAUP) analysis provide evidence for the time of divergence between the barley genes containing the internal motif as a single copy and as a repeat. The PAUP analyses also provide evidence for independent duplications of Group 1 genes containing the internal motif as a repeat in both monocots and dicots.

Introduction

The late embryogenesis abundant (*Lea*) genes in plants encode a diverse group of proteins induced

to high levels towards the end of embryo development and whose transcripts accumulate maximally during the post-abscission stage [24, 25]. Many groups of *Lea* genes contain sequence mo-

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X76933, X78330, X78331, X78332, X81087 and X81088.

tifs repeated several times and gene family members may differ in the number of repeats which they possess. Of functional importance, the proteins are usually very hydrophilic [2]. The Group 1 *Lea* sequences show the highest homology between species and are conserved along the entire coding region. The first analysed member of this group, the *Em* protein in wheat, see [12, 14, 31, 32, 35] represents the single most abundant protein in mature wheat embryos [6]. Cloning of *Em*-like genes from barley (three cDNA clones and one genomic clone) resulted in the characterization of four family members named B19.1, B19.1b, B19.3 and B19.4, after the observation of a strongly conserved 20 amino acid motif present 1, 1, 3 and 4 times in the B19 genes, respectively [9, 23]. The repeated nature of the very hydrophilic internal motif, first described by Espelund *et al.* [9] is a universal feature of this group since it is also present as a repeat in B19-like genes from maize, cotton, and *Arabidopsis* [5, 16, 17, 50]. The number of repeats in these 'repeat-motif' genes does vary from species to species, however.

As a part of our efforts to understand why so many variants of these genes are present in the barley genome, we have isolated and sequenced genomic clones corresponding to the three previously isolated cDNA clones (B19.1, B19.3 and B19.4). Together with the recently published B19.1b gene [23], these data have allowed us not only to examine the conservation of regions in the B19 promoters but also to address the mechanisms involved in the evolution of the entire family. On the basis of these results a model for the evolution of the 20 amino acid motif in the B19 family is proposed. Evidence for the time of divergence between the barley genes containing the internal motif as a single copy and as a repeat is presented, as well as analyses which show that duplications of Group 1 repeat-motif genes (and duplications of the repeats within these genes) have occurred relatively recently and independently in monocots and dicots.

Materials and methods

Plant material

Barley plants (*Hordeum vulgare* L. var. *disticum* cv. Bomi) were grown and staged as previously described [28, 30]. Culture of immature barley embryos was performed as described [28]. Culture conditions and incubation times are given in the figure legend.

DNA isolation and analysis

Genomic DNA was isolated from freeze-dried barley leaves as described by Blin and Stafford [3]. Gel electrophoresis conditions, blotting and Southern hybridizations have been described previously [14, 43]. cDNA hybridization probes consisted of B19.1, B19.3 and B19.4 cDNA sequences as described in Espelund *et al.* [9]. B19.4 'front' and 'back' portions of cDNA consisted of the first 230 bp and the last 670 bp, respectively, of the B19.4 cDNA described [9]. All cDNA probes were generated using biotinylated single-stranded templates bound to magnetic streptavidin-coated beads (Dynabeads M-280-Streptavidin, Dynal AS) in a standard random priming reaction [10]. Gene-specific oligonucleotides used to detect the various B19 gene family members have been described in Espelund *et al.* [9] and Hollung *et al.* [23]. Oligonucleotide end labelling, hybridization temperatures used and testing of gene specificity have also been described [9].

Screening of the genomic library

The barley (cv. Bomi) seedling genomic library used, was constructed in the lambda vector EMBL 3 (kindly provided by Dr. O.-A. Olsen, Plant Molecular Biology Laboratory NLVF, Norwegian Agricultural University, Ås, Norway). Screening of the library was performed at a hybridization and washing temperature of 68 °C ($T_m - 15$ °C).

Subcloning of genomic B19 fragments

The genomic lambda insert fragments chosen for subcloning (as shown in Fig. 1) were cloned directly into pBluescript SK⁻ (Stratagene), using the same polylinker restriction sites as those used to cut the genomic DNA. The subclones produced have been named pGB191, pGB191b-HX, pGB191b-XH, pGB193, and pGB194. Sequences presented here may be found in the EMBL database under accession numbers X62804 (cDNA), X76933 (promoter/leader) and X78330 (intron) for B19.1; X77157 (promoter/gene) for B19.1b; X62805 (cDNA), X65951 (PCR-generated promoter), X78332 (intron) and X81087 (promoter/leader) for B19.3; X62806 (cDNA), X65952 (PCR-generated promoter), X78331 (intron) and X81088 (promoter/leader) for B19.4.

DNA sequencing, sequence and phylogenetic analyses

Dideoxy-sequencing reactions [44] were carried out using the Sequenase 2.0 kit (USB) on single-stranded templates immobilized to streptavidin-coated magnetic beads (Dynabeads M-280, Dynal, Norway) [26]. The program packages PC-GENE (Intelligenetics and Genofit), UWGCG (University of Wisconsin), NTSYS-pc, version 1.70 for the UPGMA analyses [37] and PAUP version 3.1.1 [46] were used for the computer analyses. Alignments used to produce similarity matrixes were based on PC-GENE (NALIGN) using an open unit cost of 15 or 5 and a gap unit cost of 5. Adjustments were performed manually if it appeared that the program did not produce an optimal alignment. Additionally, since the NALIGN program bases its similarity values on dividing the number of similar bases between two sequences by the number of bases in the shortest sequence, gaps in the shorter sequence are counted as 0 while any gaps in the longer sequence are counted as the number of bases in the gap. Because the analyses attempted to elucidate evolution, the alignments were adjusted by

counting all gaps (in both the longer and the shorter sequences) as 1 in order to reflect the true evolutionary situation more closely. The oligonucleotides and PCR conditions were determined using the program OLIGO [42]. To search for homologous sequences, the EMBL nucleic acid databank (release 36) was screened.

mRNA isolation and primer extension

Poly(A)⁺ mRNA for primer extension was isolated directly from crude lysates (about 100 immature embryos or 0.2 g tissue per isolation) using magnetic oligo(dT) Dynabeads (Dynal) [27]. Amounts of isolated mRNA corresponding to 20 embryos were applied per gel lane. Total RNA was isolated as described by Galau *et al.* [15]. The gene-specific oligonucleotide primers used for primer extension have been previously described [9] while the B19 gene family primer (primer 2, e.g. 'non-specific' primer), which recognizes all four B19 genes, is described in Espelund and Jakobsen [8]. Primer extension was performed with AMV Reverse Transcriptase (Promega) under the conditions recommended by the manufacturer. Radioactive incorporation and sequencing reactions were performed as described above, using end-labelled primers. The extended products were run on either a 4% or 6% denaturing polyacrylamide gel [43].

Results

Two of the B19 genes containing repeats are tightly clustered

A barley EMBL3 seedling genomic library was screened with a B19.3 cDNA probe and the putative positives subsequently hybridized to gene-specific oligonucleotide (oligo)probes specific to B19.1, B19.3 and B19.4. Interestingly, a fraction of the isolated clones (8 of 29) did not hybridize to the gene-specific probes. Of the eight clones which hybridized only to the cDNA probe, four were designated likely pseudogene candidates (or

possibly undescribed expressed family members, see below) and three contained rearranged inserts, a phenomenon observed for EMBL3 libraries by others [45]. The eighth clone contained a gene of the same type as B19.1, and was therefore named B19.1b. The sequence and expression pattern of this gene has recently been described [23]. Detailed restriction mapping of several lambda clones were performed, among them a clone was found containing B19.3 and B19.4 together (Fig. 1).

Southern hybridizations of restriction digested genomic barley DNA allowed a further confirmation of B19 gene organization. Due to the high

degree of similarity between gene family members, hybridizations with several cDNA probes (B19.1, B19.3, B19.4-5', B19.4-3'; see Materials and methods) as well as with the gene-specific oligos, were necessary in order to construct the genomic map presented in Fig. 1. The hybridizations, in agreement with the library screening described above, indicates the presence of 1 or 2 extra members of the B19 gene family (not shown). Additionally, mapping found B19.3 and B19.4 to be tightly linked, with B19.3 positioned 2.5 kb upstream of B19.4 and in the same orientation, thereby confirming the integrity of the double lambda clone also described above. B19.1

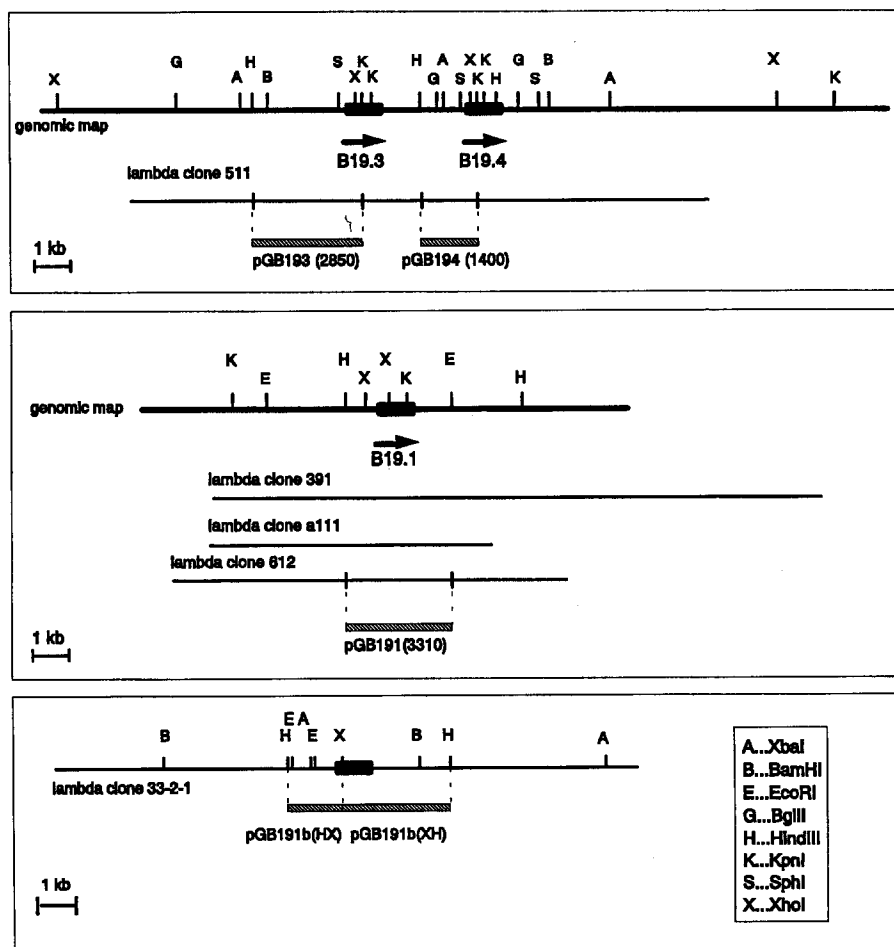


Fig. 1. Genomic map of the *Lea* B19 gene family members with corresponding inserts from isolated lambda genomic clones placed below the map. The restriction sites used to position the lambda clones in relation to the genomic map are marked on each clone. Arrows indicate gene orientation. Subclones are represented by hatched bars.

was not found within 15 kb of the B19.3/B19.4 cluster.

Putative regulatory elements are spatially conserved

Subclones from lambda clones were made for B19.1, B19.3 and B19.4 (marked in Fig. 1), and promoter and intron sequences as well as portions of the coding regions were sequenced. Coding regions from the genomic subclones were identical to the previously sequenced cDNA clones described [9], and genomic B19.3 and B19.4 promoter/leader sequences were identical

in the overlapping regions (300–350 bp) to the previously described short PCR-generated clones [8].

Figure 2 presents proximal parts of the B19 promoters with leader sequences (A) and intron sequences (B) in the form of multiple alignments. Based on the comparisons in Fig. 2A, a range of conserved regions were identified in all four B19 promoters (in bold or underlined), some of which show similarity to well characterized promoter regulatory elements (indicated in the figure). These putative elements are conserved spatially in all sequences.

A.

1	gatttgagcacatgacaactgcacacccggtAgtttagGactaaTct-AatttctaactCtaaacAattttgCaAtggtttGgacGtagatGtTtaAaaaATgcyg	-270
1b	agaagaaaaagcagtagacaatgctcgacCCATgcCtgGtttccC-gACCG----GCCGACGAccAa-CGcaC---GtacG-Gg-aG-T-GA-ttA----	-292
3	gagtaacagctctcccagagccttgcatgCCATgcCccGcagctCT---CCGagcaGCCGcCGAccAccCGAgCacgGtgaGtGctAaGTAaAccAT---	-270
4	gtggtgcccgcggtacagcatgtagagTCCATtgCtcGtagtCTCTgACCG----GCCGcCGAacAccCGAgCatgGcga-Gc--GcgaGAaCCAT---	-245
1	G--CCTG-----gGt--gACGTGg-CGA-----CagacGACaa---tG-CAGgACAACACAggtGtCACagcTtGtgtgta-----	-208
1b	G-TgCTGtagttcaGtagTACGTaTaCGAtcccacgcccACagGACgacagtGcCAagACAacgCaggtG-----TtGtacgtagggtcctatagc	-200
3	GaTCCTG-----Gc--TACGTGT-CcA-----CtctgT-----GaCAGcACAAGcAaC-G-CACgagTcG-----	-219
4	GaTCCTa-----Gc--TACGTGT-C-A-----cCag--GAC-----GaCAGcACAggACgac--tCACgagTcG-----	-194
	ABRE[19]	
1	---CTAGC--T-----Cg-TTCCggcGG----CtTcgtCagTC-cTCGGaTaaCCATGcc--cTACgCgCGTG-Cg----caggcAtGt	-142
1b	tctcTAGC--G-----Ca-TTCCggtGG----C-T-g-----C-cTCGGaTaaGcAcGC---T-CcgACGTGTcaTCCgtcCggt--GC	-137
3	---CTAGCatTgGcagctcaatctCgcTTCCt-cGGacaacCgTgcaCACtCatTCGG-TgcCCATGCAagtTAC-CACGTGTGTCca-aCcagAgGC	-126
4	---CTAGCatTgGcagctca-----TCC-----aaaacCct--aCA-Tg--TCGG-TgcccCATGCAagcTAC-CACaTGTcaTCCc-aCcagAgGC	-119
	RY[7, 22] G-Box[18, 51] (B19.3) RY (B19.1)	
1	Ttc-CATGCACGc---tcgCatGTgTcg-----cctCca-AgAgcctcG---CATggggCGCGtgcagacctcAGAGGcACACGCGCCGGCa-CT	-61
1b	cTcGcctG-ACGcggtgtca-AtGcaTcaggacagcaacagCc-G--AggcttG--CATc-----CGGtctgctc-----GgACACGCGCCGTCTCCT	-57
3	TT-GCATGCACGa-----CacGtCttaga-----ctCttGatAcagcaGga-----gCGCG-----gAGAGGTACACGCGCCGGCTCCT	-58
4	<u>TT-GCATGCACGa</u> -----CgcGtCtT-----GATAcacaGgaaCATg-----gAGAGGTACACGCGCCGGCTCCT	-59
	Sph/RY[22] GC-rich[rev. by 36]	
1	cgGGCCGtgaCGCAccCTCCCTCC--G-CcGGTATAaAAgGCGcGgccaTCGCTgcCTCCCTCAtCacCAGCATCAcGCAATCA---CAaAcGcgtt	+33
1b	-gGC-CCGcgaCGc-cCTCCTC--GcCGGTATAaAAgGCGcGAG---TCGCTgcCTCCCTCAtCacCAGCATCAaAATCAactaCAaacaCgta	+35
3	a-GCGC-G---C-CAagCTCCTCcttGtCGGTATAaAAcGCGAGC---TCGCTagCTCCCTCCgCAtCAGCATCAgGCAATCA---CagA-Gcacc-	+28
4	a---GCCGtc-CG-AagCTCCCTCctGcCGGTATAaAAcCTcGAGC---TCGCTagCCTCCCTCCgCAtCacCATCAgGCAATCA---CagA-Gcacc-	+28
	Py-rich[see 34] TATA-box[29] Py-rich	
1	C-----AAGCAGC-----cACAcAAcG---CtTCAAGTtGAgATTT-CG--TTCaag--TAGC-A-----AGGCAGaaAGC-A---ATG	+93
1b	CcaagctgAAGCAGCA-cAG--cACAcA-cgtttcCtTCcAGctGAG-----G--TTaaaG--TAGCcAGGccagCCCAGGCAgaaAGagAggtgATG	+118
3	-----AAG-AGCAGTcGCATAcCAAAac---Cct-AAGTcGgc-TTAcGgTTTCC-G--TAGCtAGGt---CCAGGCAGcgtGC-A---ATG	+100
4	-----AcG-AGCAGTAGCAtAC-aaaac---Cct-AcaTccAc-TTAcTgTTTCC-GttTAGCtAGGt---CCAGGCAGcgcGC-A---ATG	+101
	---->	

B.

1	GTAcGTacgtGcAcGcgcga-GgAaTGT-----CctTgTact-a-TGcTtGAGCGATCT-GTGACTTAcGATgCTgATTgCTaGCgTTGttgtactactg	91
1b	GTAcGT-----GcatGcgcga-GaAaTGT-----CGaTgTactGaTgGcTtGAGCGATCT-GTGACcTccATACTgATTgCTaGCaTTGctc	81
3	GTATAaT--gaG--tGcgtttGtAcTgTacttccGcT-T-C-GtTgAT-GA-CatTgTtGTGgCTTCTGATACTcATTtCTG-----	75
4	GTAtGT-----GtgcTgTacttgcGTT-T-C-GcPT-----GAGCGATCT-GTGACTTCTGATACTcAgTcCTgCATTG-----	66
1	tatttgTaTGCTGCATTtgtgTGTGCAG	119
1b	-----TaTGCTGCATT---TGTGCAG	99
3	-----GCgcCATT---TGTGCAG	90
4	--tttgTgTa-TGCgTT---TGTGCAG	87

Fig. 2. Comparisons of *Lea* B19 genomic sequences. Gaps are marked by -, and the consensus is written in upper-case letters. A (top). Alignment of the B19.1, B19.1b, B19.3 and B19.4 promoters. Putative *cis* elements, with references, are in bold or underlined. Position (bp) in relation to transcription start (double-underlined) is given on the right-hand side and translation start is marked by a horizontal arrow. Alignments were not performed upstream of the vertical arrow. Only proximal promoter regions are shown. Additional upstream sequences are available in the EMBL database (see Materials and methods). B (bottom). Alignment of the B19.1, B19.1b, B19.3 and B19.4 intron sequences. Intron length in bp is given on the right hand side.

Although none of the B19 promoters possess regions showing striking similarity (0 or 1 mismatch) to the dehydration-responsive element (DRE) of *Arabidopsis* [52], the closest similarity (2 mismatches) was found at -384 in B19.1. Also, B19.1b was found to possess the most regions showing similarity (0 or 1 mismatch) to the RRC-CCR ABA-responsive element of a *Cra-terostigma plantagineum* gene [38]. The only identical sequence, however, was found in the B19.1b leader region at +95 bp.

Transitions from GC- to AT-rich regions are observed for the four B19 genes between ca. 275 and 350 bp upstream of transcription start. Proper alignments of all four sequences against each other were difficult upstream of ca. -330 bp and were, therefore, not performed. Additional distal sequences for all four genes are available in the EMBL database (see Materials and methods).

The B19 intron sequences, presented in Fig. 2B, were found inserted at the same position relative to the coding sequences (117 bp from ATG start). All intron sequences possess the standard GT/AG borders [21] and vary in length from between 87 and 119 bp.

Primer extension indicates the presence of more expressed B19 gene family members

In order to determine the number of B19 transcripts in the embryo and to identify their 5' ends, primer extension experiments were performed. Since the genes were so similar, it was not feasible to design gene-specific primers complementary to leader regions. The only region of the B19 genes different enough to allow design of gene-specific primers was in the trailer region (see [9]). This creates problems when using a genomic sequence ladder since there is an intron in all genes (see above). Therefore, in order to determine transcription start specifically for each gene it was necessary to perform extensions using both a 'non-specific' primer (complementary to leader regions of each gene), run next to each B19 genomic sequence, and a gene-specific primer (spe-

cific for each gene's trailer region), run next to each B19 cDNA sequence. Extensions utilizing specific and non-specific primers were always performed on mRNA from the same preparation and each set of extensions was repeated several times on independently isolated mRNA preparations.

Using the B19 non-specific primer (Fig. 3A), several 'series' of banding patterns are observed, consisting of one or two stronger bands above several weaker ones. Similar patterns are observed regardless of whether the mRNA was isolated from ABA or mannitol-treated embryos, confirming that no alternative splicing or initiation events occur on different treatments. The number of banding patterns observed in Fig. 3A suggest that at least four, more likely five or six, transcripts exist in the barley embryo.

Combining results from the non-specific primer with the gene-specific primers (Fig. 3B and results not shown) allowed us to assign identity to, and determine the sites of initiation for each gene. This was performed by counting the number of bases over the end of the cDNA ladder sequence (as seen in Fig. 3B), and then reading the actual sequence (over the cDNA terminus) from the genomic ladders in Fig. 3A. The specific primers (B19.1 shown) identified one of the lower banding series from Fig. 3A to correspond to B19.1 (filled arrow), while the weaker, overlapping middle patterns were found to correspond to B19.4 and B19.3 (filled arrows). The signal strength of the patterns in Fig. 3A are in agreement with the mRNA levels seen on northern blots [9]. The uppermost series of bands has been shown to correspond to the B19.1b transcript [23]. The observation that several series of bands remain which do not correspond to any of the four characterized B19 genes is in agreement with the Southern hybridization and library screening data.

The primer extension analyses described above found transcription to be initiated at several different positions in a conserved tandem repeated motif CA(C/G/T) for all B19 genes. The transcripts initiate mainly at positions 93, 100 and 101 bp upstream of the ATG start codons for

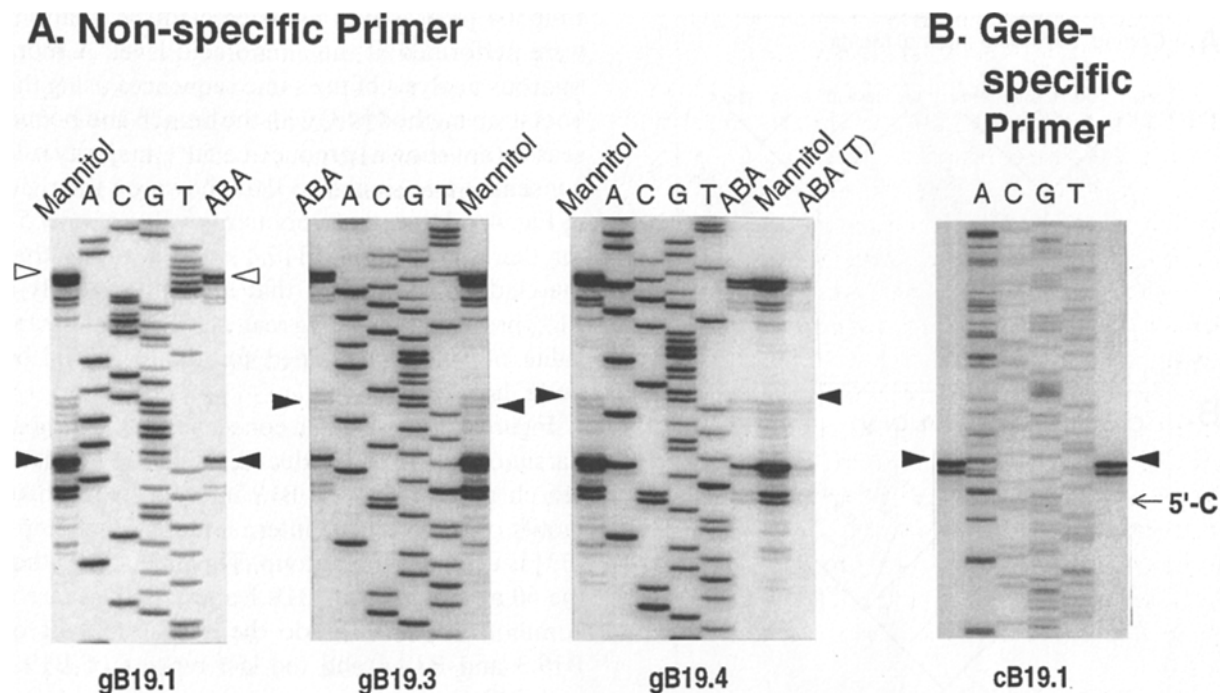


Fig. 3. Determination of transcription start for *Lea* B19 family members using both (A) an 'unspecific' primer complementary to a conserved region in the B19 genes and (B) a gene-specific primer. In (A) primer extension was performed on poly(A)⁺ or total (lane T) RNA from embryos 20 days after anthesis incubated 24 h on 1/2 MS plus either 20% mannitol or 50 μ M ABA, as indicated. Closed arrows show the position of the strongest expressed products from experiments with the gene-specific primers (shown in panel B for B19.1, not shown for B19.3 and B19.4). The sequence ladders shown next to the extension products represent the minus strand of a genomic B19.1 sequence, and PCR/promoter-cDNA constructions of B19.3 and B19.4 which are equivalent to their genomic sequences (for details see text). The open arrow indicates the primer extension product corresponding to B19.1b. B. Primer extension was performed on poly A⁺ RNA from embryos 20 days after anthesis with the B19.1 gene-specific primer next to the B19.1 cDNA sequence (same primer, minus strand shown). 5'C marks the 5' end of the B19.1 cDNA. The solid arrowhead marks the strongest expressed primer-extension product.

B19.1, B19.3 and B19.4, respectively. These initiation points are indicated in Fig. 2A along with that of B19.1b, which initiates 116 bp upstream of the ATG start codon, within a CAT motif [23].

The B19 single-motif type is the ancestral form of the gene

Since most nucleotide changes within the coding regions of the B19 genes were found to occur either at the wobble position or to code for synonymous amino acids (observed for 84% of the changes within repeats, 67% of the changes outside of repeats) selection may be supposed to have played a minor role at these sites. A cladistic

analysis may, therefore, be considered a reliable measure of the genes' evolutionary relationships. Figure 4 presents comparisons of B19 coding regions using the program package PAUP, version 3.1.1 [46] and shows the most parsimonious trees produced based on sequence data. For the analyses, the single motif rice *Lea* Group 1 gene Emp1 [33] has been used as an outgroup to root the tree since this species is the closest diploid relative to barley for which sequence data is available.

Figure 4A presents the two most parsimonious trees produced, using the heuristic search method, for B19 coding regions, excluding the internal hydrophilic motifs. The main difference between these two trees is the position of the B19.1b branch, which directly shares either a common

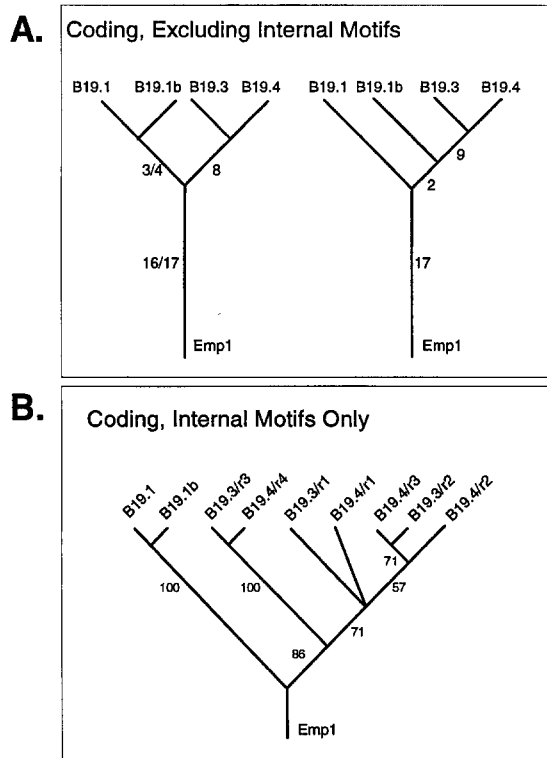


Fig. 4. The proposed evolutionary relationships between (A) the *Lea* B19 coding regions and (B) the internal motifs within each *Lea* B19 family member, based on the PAUP phylogenetic analysis, performed at the nucleotide level. All trees have been rooted by the rice gene Emp1 (see text). A. Most parsimonious trees based on B19 sequence similarity of coding regions minus internal motifs. Numerals beside each branch show the numbers of informative characters for each branch. Both trees have a tree length of 57, a consistency index (CI) of 0.947 and a CI excluding uninformative characters of 0.833. B. Majority rule consensus of the 7 most parsimonious trees based on B19 coding sequence similarity within the internal motifs only. Motifs are designated by which gene member they belong to, r for repeat (motif), and then the number of the repeat (motif) within that gene. Numerals beside each branch show the percentage of the most parsimonious trees which match consensus.

ancestor with B19.1 or with B19.3 and B19.4. Both trees suggest that a gene containing a single copy of the internal motif was the ancestral type. Also, in both trees B19.3 and B19.4 are shown to be most closely related as well as being the most recent descendants of the family. The same two trees were produced when rooted with the Group 1 cotton gene D19 [2] or with the maize gene

Emb564 [49] as well as when cladistic analyses were performed at the amino acid level. A more rigorous analysis of the same sequences using the bootstrap method [46] with the branch and bound search (not shown) produced a 50% majority rule consensus tree similar to that shown on the right in Fig. 4A. However, a bootstrap value of only 51 for the B19.1b/B19.3/B19.4 node suggests that the clade supported by that node shows only a 51% probability of being real. A highly significant value of 99 was obtained for the B19.3/B19.4 node, however.

Figure 4B presents a consensus of the most parsimonious trees produced, using the heuristic search pattern, for the B19 internal hydrophilic motifs only, where the internal motif from Emp1 [33] is used as an outgroup. This tree shows that the 60 nt motif within B19.1 and B19.1b share a common ancestor, as do the middle repeats of B19.3 and B19.4 and the last repeats of B19.3 and B19.4.

A 'chimera' of the single-motif genes most likely arose during the evolution of the B19 family

Due to the lower degree of similarity between B19 non-coding sequences (i.e. due to small insertions/deletions) a phenetic instead of cladistic study was performed using the program UPGMA, from the NTSYS-pc package version 1.70 [37] (see Materials and methods). Figure 5 presents the phenetic trees produced for each gene region compared. When analysing the non-coding regions one may again observe the grouping together of B19.1 and B19.1b sequences as well as the grouping of B19.3 and B19.4 sequences. However, some divergence from this trend is seen for intron and trailer comparisons, where B19.4 is more similar to the single-motif genes within the intron and B19.1b is more similar to the repeated-motif genes within the trailer.

Since the analyses produced different UPGMA trees for some of the non-coding regions, the original identity matrixes, for non-coding as well as coding regions, were examined in more detail. These data, presented in Table 1, were obtained

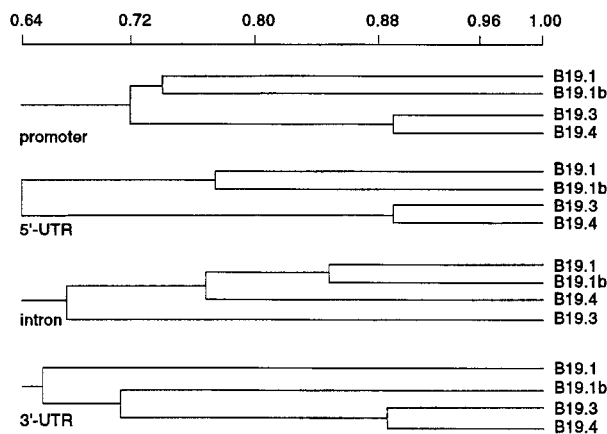


Fig. 5. UPGMA trees for *Lea* B19 non-coding regions. The scale shows percentage of identity.

from individual alignments of pairs of B19 sequences (see Materials and methods), and show the same trends as the cladistic analyses shown in Figure 4A, B19.3 and B19.4 were found to be,

on average, 6% more like each other than B19.1 to B19.1b, consistent with the idea that the divergence between B19.3 and B19.4 occurred later than the divergence between B19.1 and B19.1b. An interesting observation, not apparent from the analyses presented in Fig. 4 or Fig. 5, is made when the percentage of identity values are compared for single-motif genes to repeated-motif genes. For 5' regions (promoter, 5'-UTR sequences, coding regions 5' of the repeats and intron sequences) B19.3 and B19.4 are slightly more like B19.1 (1.6% more similar on the average, than like B19.1b) while for 3' regions (repeats, coding region 3' of the repeats and 3'-UTR) B19.3 and B19.4 are more like B19.1b (8%, on average, more than like B19.1). This suggests that during the evolution of the B19 family, a 'chimera' of B19.1 and B19.1b was formed which later gave rise to B19.3 and B19.4. Based on the results presented in both Fig. 5 and Table 1, a transition or 'break point' between B19.1-like and

Table 1. Summary of identity values for B19¹ subsequence comparisons.

Comparison	Promoter ²	5'-UTR	Coding 5' of repeats	Intron	Repeats ³	Coding 3' of repeats	3'-UTR
B19.1 to B19.1b	0.75	0.77	0.96	0.84	0.93	0.91	0.66
B19.1 to B19.3	0.71	0.65	0.92	0.68	1 0.78 0.85 0.83	0.85	0.61
B19.1 to B19.4	0.74	0.67	0.94	0.76	1 0.85 0.85 0.83 0.83	0.86	0.66
B19.1b to B19.3	0.71	0.62	0.92	0.63	1b 0.85 0.90 0.90	0.89	0.73
B19.1b to B19.4	0.73	0.62	0.94	0.77	1b 0.92 0.92 0.88 0.88	0.93	0.67
B19.3 to B19.4	0.89	0.89	0.98	0.75	3r1 0.90 0.90 0.87 0.80 3r2 0.95 0.98 0.98 0.92 3r3 0.88 0.88 0.88 0.95	0.93	0.88
most like							
B19.3 to B19.4 ⁴	(1)	1	1&1b	(1)	1b	1b	1b

¹ Identity is calculated from pairwise comparisons of nucleotide sequences as described in Materials and methods.

² Sequences from transcription start to the most distal RY element (see Fig. 2A) were used for the promoter comparison.

³ Repeat matixes compare the internal motif(s) of the B19 genes to each other, where the motifs are designated by which gene member they belong to and r, for repeat, followed by the number of that repeat within the gene.

⁴ Summary of comparisons between the repeated-motif genes to the single-motif genes, where a question was posed, 'which single-motif gene, B19.1 or B19.1b, is most like the repeated-motif genes, B19.3 and B19.4?'

B19.1b-like for the proposed 'chimera' is suggested at or near the intron.

Discussion

The proximal part of the B19 promoters is highly conserved

Comparative studies have uncovered a strikingly high degree of similarity between the B19 promoter regions. Additionally, putative regulatory regions such as the ABRE, G-box, RY and Sph-RY-like sequences (rev. in [47]) have been identified in all genes, within the first 280 bp of transcription initiation. Interestingly, comparisons of promoter regions from all *Lea* Group 1 upstream sequences characterized to date identified perfect RY or RY-like (0 or 1 mismatch) sequences in every monocot *Lea* Group 1 promoter within

350 bp of putative TATA boxes. However, no RY-like (0 or 1 mismatch) sequences were found within the same region for any of the dicot Group 1 promoters. This could prove interesting in light of the observed differences in regulation, during embryogenesis, between monocots and dicots [4, 9]. Functional studies are underway to analyse the putative regulatory regions of the four B19 genes.

A model for the evolution of the B19 family

Based on our observations, a model which includes sequential gene duplications is proposed for the evolution of the B19 gene family, including the internal motifs (Fig. 6). According to the model, and based on results presented in Figs. 4 and 5 and Table 1, the B19 family has arisen

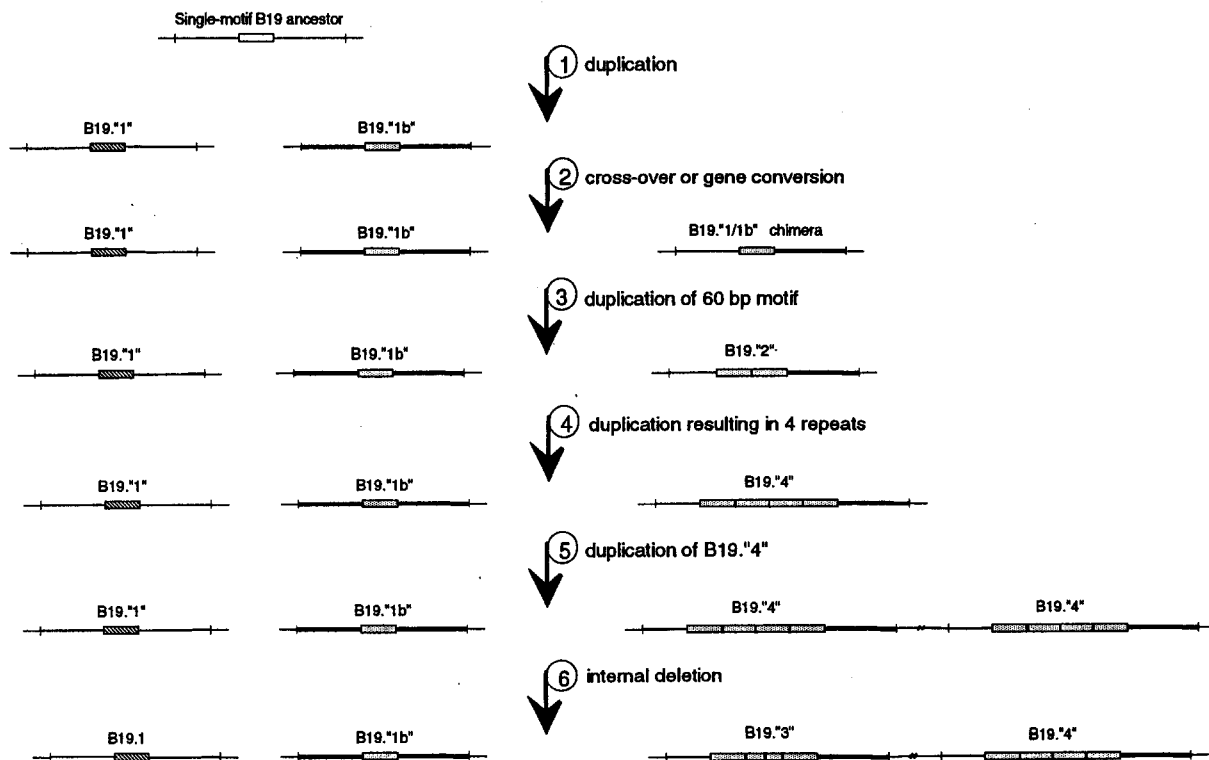


Fig. 6. Suggested model for evolution of the *Lea* B19 gene family. Boxes represent internal motifs within each gene. Sequence divergence is represented by shading of boxes and thickness of lines (no further divergence is shown for B19.1 and B19.1b after step 2). Steps 4–6 show the scenario in which B19.4 is created directly from a double-motif ancestor, the possible scenario in which B19.3 is created first is not shown (see text).

through duplications of a single-motif ancestor. This ancestor must have been duplicated, giving rise to B19.1 and B19.1b, which group together in most analyses performed here (Fig. 6, step 1).

The most likely mechanism responsible for this initial duplication is non-allelic homologous recombination between repeated elements, a mechanism which may lead to the amplification of relatively small genomic regions, and has been proposed by Sutliff *et al.* [45] for the amplification of the α -amylase family in rice. Here, analyses of flanking sequences from individual gene family members revealed the presence of repetitive elements both upstream and downstream of the genes. Similar observations have been made concerning sequences upstream of B19.1 and B19.1b (R. Stacy, K. Hollung and K.S. Jakobsen, unpublished data). Stabilizations of meiotic products could then have occurred as suggested by Pichersky [39].

The next step in the model is the creation of a third single-motif gene of chimeric (B19.1/B19.1b) nature (step 2), which later gives rise to B19.3 and B19.4. What could have caused the chimeric nature of this third single-motif gene? Three evolutionary scenarios are suggested to explain this. A cross-over event could have occurred between B19.1 and B19.1b, creating the repeated genes' single-motif ancestor. Evidence of cross-over events have been observed for other plant gene families [45]. Another possible explanation is a gene-conversion of the back portion of the repeated genes' single-motif ancestor to B19.1b-like. Gene conversion has been implicated in the evolution of large gene families but may also have an effect on smaller groups of dispersed repeated sequences [20]. A third possibility is that, if B19.1 and B19.1b were found close to each other in the genome, a duplication of the putative cluster and subsequent deletion (duplication/deletion mechanism) within one of the two clusters could produce a chimeric single-motif ancestor to B19.3 and B19.4. The characterization of a genomic clone from wheat containing two clustered single-motif genes [13] may support this mechanism. As an added note, we believe the chimeric nature of B19.3 and B19.4 could explain the production of

the two (rather than only one) most parsimonious trees shown in Fig. 4A.

The third step in the model involves the duplication of the internal motif of the B19.1/B19.1b chimera to two copies. We feel that the creation of a chimeric gene preceded the duplication of the internal motif to two copies (as opposed to a two-repeat 'chimera' being created directly from B19.1 and B19.1b, with one repeat being of B19.1 nature and the other of B19.1b nature) since all repeats of B19.3 and B19.4 are more like B19.1b than like B19.1 (Table 1). How have the genes with repeats (B19.3 and B19.4) arisen from a chimeric single-motif ancestor? An initial amplification of the 60 bp internal motif must have occurred, resulting in a gene containing two repeats (Fig. 6, step 3). Although no 'B19.2'-like genes have been identified in barley, such examples have been isolated from maize [50], cotton [16] and *Arabidopsis* [5]. Based on the cladistic analyses presented in Fig. 4B, further motif duplications within the double-motif gene could then have given rise to B19.3 and B19.4. Data from *Arabidopsis* *Lea* Group 1 genes support the idea of a double motif intermediate, since genes containing two and four copies of the internal motif (Ale [5] from the Landsberg ecotype and GEA1 [17] from the Columbia ecotype, respectively) show 99.5% similarity at the nucleotide level within the coding region (analysis performed excluding repeats).

How did the duplication from two to three and four repeats occur? High sequence similarity between the middle two repeats of B19.4 (Fig. 4, Table 1) indicate that a sequential amplification of the internal motif from two to three to four repeats could have occurred, as proposed for the *Arabidopsis* gene GEA1 by Gaubier *et al.* [17]. However, the average similarity amongst the B19.3 repeats is nearly 10% lower than amongst the B19.4 repeats (comparing first, middle, and last repeats in each gene, see Table 1), a phenomenon difficult to fully explain should the two genes be of the same age. The following alternative explanation of the reduced internal similarity of B19.3, involving the creation of B19.4 directly from a 'two-repeat' ancestor (Fig. 6, step 4), may provide a solution to the problem. If, after a du-

plication of the B19.4-like gene (Fig. 6, step 5), a deletion was to occur within one of the B19.4-like genes (the last half of repeat 2 being deleted along with the first half of repeat 3; Fig. 6, step 6), the resulting middle motif would be identical to the middle repeat of B19.3. Additionally, the internal similarity would be reduced by 2%. Our data does not make it possible to prove that the ancestral gene to B19.3 and B19.4 contained four repeats. However, the presence of genes from *Arabidopsis* which contain two and four repeats (Ale and GEA1, respectively) and are suggested to be allelic (see above and [5]), makes this scenario plausible.

The evolution of the Group 1 Lea gene family and the 20 amino acid motif in dicots and monocots

In most dicot and monocot species investigated small, multigene Group 1 *Lea* families exist. In both dicots and monocots the internal motif is present as a repeat. A notable exception is in rice where Ranjhan *et al.* [41], as a part of their chromosomal mapping study, detect only one Group 1 *Lea* gene (Emp1). If the initial duplication of the internal motif occurred prior to the divergence of monocots and dicots, the situation in rice would imply that either the double motif gene has diverged enough to reduce homology to below detectable levels at the Southern hybridization conditions used by Ranjhan *et al.* [41] or that rice has lost this double-motif gene. In contrast, the initial duplication of the internal repeat may be an independent event in monocots and dicots. In this case, the occurrence of a single rice gene with a single internal motif is possible. To address the evolutionary relationship of the different Group 1 genes and the duplications of the internal motif we analysed the coding sequences of the 21 reported *Lea* group 1 genes using the maximum parsimony method (PAUP). The analysis was performed on the coding region (excluding the internal motifs) at the nucleotide and amino acid level, as well as on the 'repeats' themselves at the nucleotide level (Fig. 7A, B, C).

The resulting trees match relatively well with

known systematic relationships. The consensus trees produced for the coding sequence (excluding repeats) at the nucleotide (a) and amino-acid (b) level are in good agreement with some exceptions. These include the grouping of the maize genes EMB564 and Emb5, the rice Emp1 with wheat TAEMG, and the *Arabidopsis* GEA1 and Ale with the radish p8B6 at the amino acid level but not at the nucleotide level. The cotton gene D132, at both levels, does not form a clade with any dicot gene. The consensus trees (Fig. 7A and B) show that the barley genes may be divided into two sub-sets; one sub-set (B19.3 and B19.4) which groups together with the wheat gene EmH5 and another subset (B19.1 and B19.1b) which groups together with the wheat genes EmH2, pMTK1a and TAEMR. This suggests that the divergence of the ancestor to B19.3 and B19.4 from the ancestor to B19.1 and B19.1b occurred prior to the divergence of barley and wheat. Divergence of B19.3 and B19.4 must then have followed (as shown in Fig. 6). Additional examples grouping of genes across species rather than within species are EMB1 and D19, and the repeat-containing *Arabidopsis* genes, which are divergent from the single-motif genes. These data also support the idea of ancient duplication(s) of Group 1 genes. However, this does not directly address the evolution of the repeats.

While repeat-motif and single-motif genes from the same species do not group together in the examples above, it is worth noting that the coding regions of these repeat-motif genes do group together according to species (B19.3 with B19.4; GEA1 with Ale) rather than according to the number of repeats (i.e. the four-repeat barley gene, B19.4, does not group together with the four repeat *Arabidopsis* gene, GEA1). These groupings suggest that the duplications of these repeat-motif genes from a common ancestor are relatively young events which have occurred independently of each other. This is also supported by the conclusions reached by Gaubier *et al.* [17]. However, our results do not support that GEA1 (four repeats in *Arabidopsis*) is most like B19.3 and B19.4, possibly due to the fact that more sequences were available for our analyses.

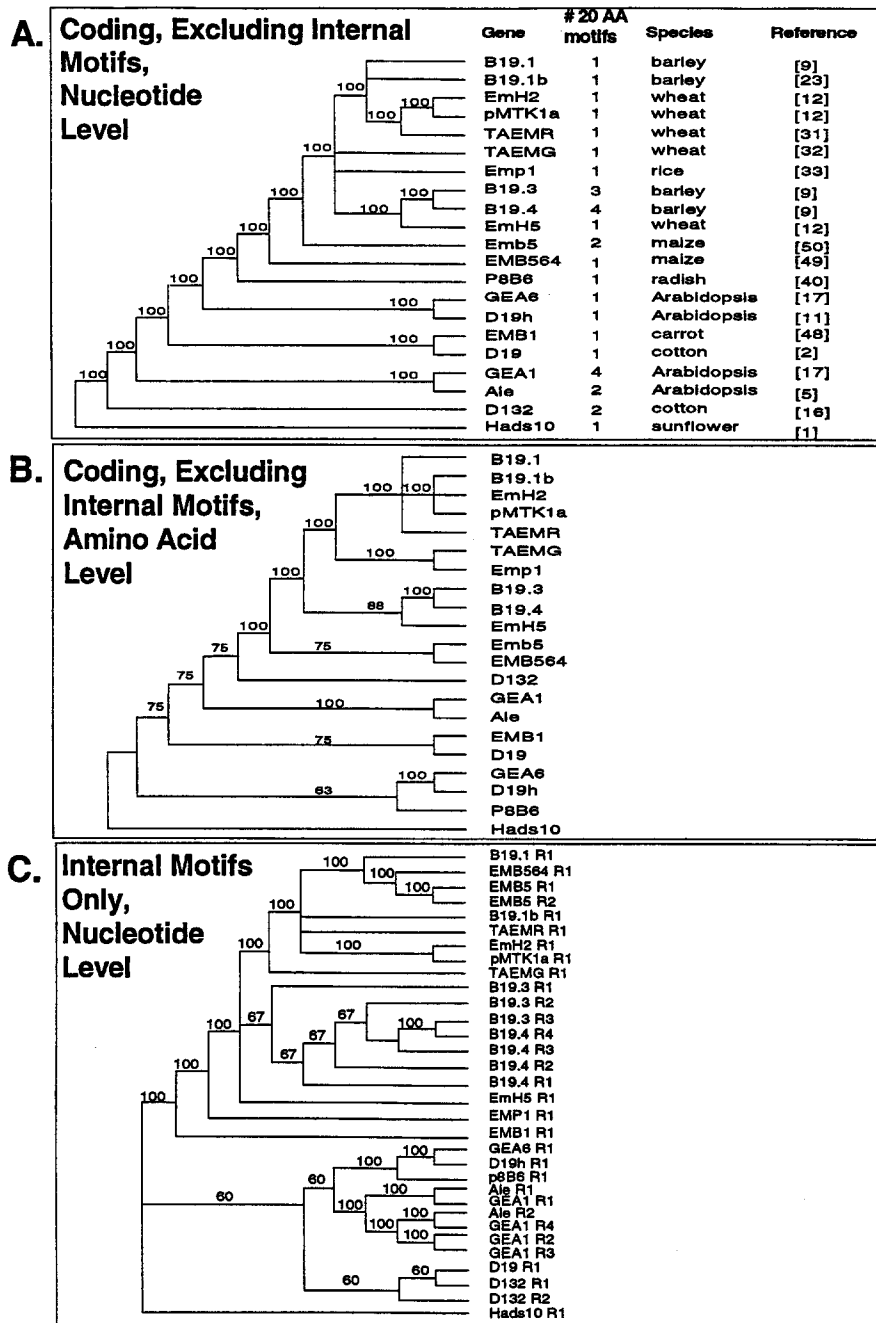


Fig. 7. Phylogenetic consensus trees for 21 *Lea* Group 1 genes. Analyses shown in A and B were performed on coding region sequences, excluding the internal motifs, using the program PAUP at both (A) the nucleotide level and (B) the amino acid level. Numerals beside each branch show the percentage of the 8 (for nucleotide level) and 11 (for amino acid level) most parsimonious trees which match consensus. (C) shows the phylogenetic consensus tree for *Lea* Group 1 internal motifs at the nucleotide level. Numerals beside each branch show the percentage of the fifteen most parsimonious trees which match consensus. All trees have been rooted with the Hads10, since sunflower is the most distantly related species for which *Lea* Group 1 sequence data are available.

The analysis of the internal motifs alone (Fig. 7C) show, in agreement with Fig. 7A and B, that motifs present as a repeat group within a species and not across species. Analyses of the internal motifs at the amino acid level produce very low resolution, nevertheless, the same trend was observed (data not shown). Notably, a 'tighter' and more evident grouping by species is observed for the dicot internal motifs alone than for the rest of the coding region of the dicot genes (compare Fig. 7C to A and B). For example, all cotton internal motifs group together while the rest of the coding regions are separated by several branches at both the nucleotide and amino acid level. In agreement with the proposed functional importance of the internal motif [9], this indicates that there is a stronger selective pressure on the internal motifs, and therefore the motif has a different evolutionary rate than the rest of the coding sequence.

Taken together, the PAUP analyses of the Group 1 genes strongly suggest ancient duplication(s) of the entire gene, and further duplications once repeats have arisen. Our data does not allow determination of when the gene duplication(s) occurred. Furthermore, the trees are not conclusive about an initial duplication of the internal motif. However, based on the widespread occurrence of the repeats and since the probability of additional repeats arising from an already repeated sequence is higher than of arising several times *de novo* from a single motif, we suggest that the initial duplication of this motif is a single event. If so, this must have occurred prior to the divergence of dicots and monocots. The characterization of more *Lea* Group 1 genes, both from barley (since our results indicate the presence of more transcribed members) and from pseudogenes and active members from primitive species, will hopefully aid in forming an even better understanding of the evolution of this group.

Acknowledgements

Erik Hornes and Dynal A.S. are gratefully acknowledged for the gifts of oligo (dT) and strepta-

vidin magnetic beads and oligonucleotides, as is Odd-Arne Olsen for providing the barley genomic library. We are indebted to Odd E. Stabbetorp, Brita Stedje, John E. Stacy and Bjørn Magne Fangen for fruitful discussions and help in performing the cladistic and phenetic analyses. Thanks are due to Solveig Hauge Engebretsen for skillful technical assistance and to Reidunn B. Aalen and John E. Stacy for carefully reading the manuscript. This work was supported in part by a grant from the Norwegian Research Council (NFR) to K.S.J.

References

1. Almoquera C, Jordano J: Developmental and environmental concurrent expression of sunflower dry-seed-stored low-molecular-weight heat-shock protein and *Lea* mRNAs. *Plant Mol Biol* 19: 781–792 (1992).
2. Baker J, Steele C, Dure III L: Sequence and characterization of 6 *Lea* proteins and their genes from cotton. *Plant Mol Biol* 11: 277–291 (1988).
3. Blin N, Stafford DF: A general method for isolation of high molecular weight DNA from eucaryotes. *Nucl Acids Res* 3: 2303–2308 (1976).
4. Crouch M: Regulation of gene expression during seed development in flowering plants. In: Browder LW (ed), *Developmental Biology: A Comprehensive Synthesis*, vol. 5, pp. 367–404. Plenum Press, New York (1988).
5. Calvo ES, Rodermel SR, Shoemaker RC: A third highly conserved group 1 *Lea* gene from *Arabidopsis thaliana* L. *Plant Physiol* 106: 787–788 (1994).
6. Cuming AC, Lane BG: Protein synthesis in imbibing wheat embryos. *Eur J Biochem* 99: 217–224 (1979).
7. Dickenson CD, Evans RP, Nielsen NC: RY repeats are conserved in the 5'-flanking regions of legume seed-protein genes. *Nucl Acids Res* 16: 371 (1988).
8. Espelund M, Jakobsen KS: Cloning and direct sequencing of plant promoters using primer-adaptor mediated PCR on DNA coupled to a magnetic solid phase. *Biotechniques* 13: 74–81 (1992).
9. Espelund M, Sæbøe-Larssen S, Hughes DW, Galau GA, Larsen F, Jakobsen KS: Late embryogenesis-abundant genes encoding proteins with different numbers of hydrophilic repeats are regulated differentially by abscisic acid and osmotic stress. *Plant J* 2: 241–252 (1992).
10. Espelund M, Prentice Stacy RA, Jakobsen KS: A simple method for generating single-stranded DNA probes labeled to high activities. *Nucl Acids Res* 18: 6157–6158 (1990).
11. Finkelstein RR: Abscisic acid-insensitive mutations provide evidence for stage-specific signal pathways regu-

- lating expression of an *Arabidopsis* late embryogenesis-abundant (*lea*) gene. *Mol Gen Genet* 238: 401–408 (1993).
12. Futers TS, Onde S, Turet M, Cuming AC: Sequence analysis of two tandemly linked *Em* genes from wheat. *Plant Mol Biol* 23: 1067–1072 (1993).
 13. Futers TS, Vaughan TJ, Sharp PJ, Cuming AC: Molecular cloning and chromosomal location of genes encoding the 'Early-methionine-labelled' (*Em*) polypeptide of *Triticum aestivum* L var. Chinese Spring. *Theor Appl Genet* 80: 43–48 (1990).
 14. Galau GA, Hughes DW, Dure III L: Abscisic acid induction of cloned cotton late embryogenesis-abundant (*Lea*) mRNAs. *Plant Mol Biol* 7: 155–170 (1986).
 15. Galau GA, Legocki AB, Greenway SC, Dure III L: Cotton messenger RNA sequences exist in both polyadenylated and nonpolyadenylated forms. *J Biol Chem* 256: 2551–2560 (1981).
 16. Galau GA, Wang HYC, Hughes DW: Cotton *Lea4* (D19) and *LeaA2* (D132) group 1 *Lea* genes encoding water stress-related proteins containing a 20-amino acid motif. *Plant Physiol* 99: 783–788 (1992).
 17. Gaubier P, Raynal M, Hull G, Huestis GM, Grellet F, Arenas C, Pages M, Delseny M: Two different *Em*-like genes are expressed in *Arabidopsis thaliana* seeds during maturation. *Mol Gen Genet* 238: 409–418 (1993).
 18. Giuliano G, Pichersky E, Malik VS, Timko MP, Sednik PA, Cashmore AR: An evolutionary conserved protein binding sequence upstream of a plant light-regulated gene. *Proc Natl Acad Sci USA* 85: 7089–7093 (1988).
 19. Guiltinan MJ, Marcotte Jr WR, Quatrano RS: A plant leucine zipper protein that recognizes an abscisic acid response element. *Science* 250: 267–271 (1990).
 20. Guthrie C, Abelson J: Organization and expression of tRNA genes in *Saccharomyces cerevisiae*. In: Strathern JN, Jones EW, Broach JR (eds) *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, pp. 487–528. Cold Spring Harbor Press, Cold Spring Harbor, NY (1982).
 21. Hanley BA, Schuler MA: Plant intron sequences: evidence for distinct groups of introns. *Nucl Acids Res* 16: 7159–7176 (1988).
 22. Hattori T, Vasil V, Rosenkrans L, Hannah LC, McCarty DR, Vasil IK: The *Viviparous-1* gene and abscisic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes Devel* 6: 609–618 (1992).
 23. Hollung K, Espelund M, Jakobsen KS: Another *Lea* B19 gene (Group 1 *Lea*) from barley containing a single 20 amino acid hydrophilic motif. *Plant Mol Biol* 25: 559–564 (1994).
 24. Hughes DW, Galau GA: Temporally modular gene expression during cotyledon development. *Genes Devel* 3: 358–369 (1989).
 25. Hughes DW, Galau GA: Developmental and environmental induction of *Lea* and *LeaA* mRNAs and the post-abscission program during embryo culture. *Plant Cell* 3: 605–618 (1991).
 26. Hultman T, Ståhl S, Hornes E, Uhlén M: Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. *Nucl Acids Res* 17: 4937–4946 (1989).
 27. Jakobsen KS, Breivold E, Hornes E: Purification of mRNA directly from crude plant tissues in 15 minutes using magnetic oligo dT microspheres. *Nucl Acids Res* 18: 3669 (1990).
 28. Jakobsen K, Klemsdal SS, Aalen RB, Bosnes M, Alexander D, Olsen O-A: Barley aleurone cell development: molecular cloning of aleurone-specific cDNAs from immature grains. *Plant Mol Biol* 12: 285–293 (1989).
 29. Joshi CP: An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucl Acids Res* 15: 6643–6653 (1987).
 30. Klemsdal SS, Kvaale A, Olsen O-A: Effects of the barley mutants Risø 1508 and 527 high lysine genes on the cellular development of the endosperm. *Physiol Plant* 67: 453–459 (1986).
 31. Litts JC, Colwell GW, Chakerian RL, Quatrano RS: The nucleotide sequence of a cDNA clone encoding the wheat *Em* protein. *Nucl Acids Res* 15: 3607–3618 (1987).
 32. Litts JC, Colwell GW, Chakerian RL, Quatrano RS: Sequence analysis of a functional member of the *Em* gene family from wheat. *J DNA Sequencing Mapping* 1: 263–274 (1991).
 33. Litts JC, Erdman MB, Huang N, Karrer EE, Noueiry A, Quatrano RS, Rodrigues RL: Nucleotide sequence of the rice (*Oryza sativa*) *Em* protein gene (*Empl*). *Plant Mol Biol* 19: 335–337 (1992).
 34. Maas C, Schaal S, Wolfgan W: A feedback control element near the transcription start site of the maize *Shrunken* gene determines promoter activity. *EMBO J* 9: 3447–3452 (1990).
 35. McCarty DR, Hattori T, Carson CB, Vasil V, Lazar M, Vasil IK: The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. *Cell* 66: 395–905 (1991).
 36. Mitchell PJ, Tjian R: Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245: 371–378 (1989).
 37. Nei M: *Molecular Evolutionary Genetics*, pp. 293–298. Columbia University Press, NY (1987).
 38. Nelson D, Salamini F, Bartels D: Abscisic acid promotes novel DNA-binding activity to a desiccation-related promoter of *Craterostigma plantagineum*. *Plant J* 5: 451–458 (1994).
 39. Pichersky E: Nomad DNA: a model for movement and duplication of DNA sequences in plant genomes. *Plant Mol Biol* 15: 437–448 (1990).
 40. Raynal M, Depigny D, Cooke R, Delseny M: Characterization of a radish nuclear gene expressed during late seed maturation. *Plant Physiol* 91: 829–836 (1989).
 41. Ranjhan S, Litts JC, Foolad MR, Rodriguez RL: Chro-

- mosomal localization and genomic organization of α -amylase genes in rice (*Oryza sativa* L.). *Theor Appl Genet* 82: 481–488 (1991).
42. Rychlik W, Rhoads RE: A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and *in vitro* amplification of cDNA. *Nucl Acids Res* 17: 8543–8551 (1989).
 43. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
 44. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463–5467 (1977).
 45. Sutliff TD, Huang N, Litts JC, Rodriguez RL: Characterization of an α -amylase multigene cluster in rice. *Plant Mol Biol* 16: 579–591 (1991).
 46. Swofford DL: *Phylogenetic analysis using parsimony: program and documentation*. Illinois Natural History Survey, Champaign, Urbana (1990).
 47. Thomas TL: Gene expression during plant embryogenesis and germination: an overview. *Plant Cell* 5: 1401–1410 (1993).
 48. Ulrich TU, Wurtele ES, Nikolau BJ: Sequence of EMB-1, an mRNA accumulating specifically in embryos of carrot. *Nucl Acids Res* 18: 2826 (1990).
 49. Williams B, Tsang A: A maize gene expressed during embryogenesis is abscisic acid-inducible and highly conserved. *Plant Mol Biol* 16: 919–923 (1991).
 50. Williams BA, Tsang A: Nucleotide sequence of an abscisic acid-responsive, embryo-specific maize gene. *Plant Physiol* 100: 1067–1068 (1992).
 51. Williams ME, Foster R, Chua NH: Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. *Plant Cell* 4: 485–496 (1992).
 52. Yamaguchi-Shinozaki K, Shinosaki K: A novel *cis*-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low temperature, or high-salt stress. *Plant Cell* 6: 251–264 (1994).