Sequence and expression of potato U2 snRNA genes

Robbie Waugh, Gillian Clark, Petra Vaux and John W.S.Brown*
Department of Cellular and Molecular Genetics, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

Received November 13, 1990; Accepted December 7, 1990 EMBL accession nos X56320-X56323 (incl.)

ABSTRACT

Plant UsnRNA multigene families show a high degree of sequence variation among individual gene members. The potato U2snRNA gene family consists of between twenty-five and forty genes. Four potato U2snRNA gene variants have been isolated. Despite the sequence variation in coding and flanking regions, all maintain the conserved U2snRNA secondary structure and all contain the plant UsnRNA promoter elements: the upstream sequence element (USE) and TATA-like box in the −70 and −30 regions respectively. In RNase A/T_1 protection analyses, one of the genes, PotU2-22, protected high levels of full length U2snRNA transcripts in potato leaf, stem, root and tuber RNA. Thus, PotU2-22 or genes with identical coding regions, are highly expressed in these potato organs and therefore represent a major subset of functional U2snRNA genes. Similar expression levels of the PotU2-22 sequence variant were also found in four genetically different potato cultivars and also in tobacco, a species closely related to potato, suggesting conservation of the coding regions of expressed U2snRNA genes. A second gene, PotU2-4, protected very low levels of full length transcripts while a third gene, PotU2-11, was not expressed in the potato organs analysed. The relative expression levels of the gene variants may reflect individual gene differences in, for example, the USE and TATA regulatory elements, or variations in gene copy number.

INTRODUCTION

The majority of eukaryotic protein-coding genes contain non-translated intron sequences which are removed from the pre-messenger RNA (pre-mRNA) by the process of splicing. Splicing occurs in a ribonucleoprotein complex, the spliceosome, comprising the pre-mRNA, U-type small nuclear ribonucleoprotein particles (UsnRNPs) and other proteins. Assembly of the spliceosome requires interactions between specific regions of the intron sequence and UsnRNPs. The 5′ splice site, branch point and probably the 3′ splice site are recognised by the U1, U2 and U5snRNPs respectively (1,2).

In the absence of an *in vitro* splicing system from plants, transient gene expression in plant cell protoplasts has provided information on pre-mRNA processing. Plasmids containing intron constructs can be routinely introduced into protoplasts by electroporation or PEG treatment and splicing analysed on isolated RNA (3−7). However, a dissection of the components of the plant spliceosome without an *in vitro* system is much more difficult. Most progress has been made in the characterisation of the U-type small nuclear RNAs (UsnRNAs). These RNAs associate with a number of protein molecules to form UsnRNPs which mediate the splicing process. U1snRNA genes have now been isolated from bean, soybean, tomato and potato (8−11); U2snRNA genes from *Arabidopsis*, pea, and maize (12−14); U3snRNA genes from tomato, *Arabidopsis* and tobacco (15,16); U5snRNA genes from *Arabidopsis*, potato and maize (17,18) and U6snRNA genes from *Arabidopsis*, tomato and maize (19−21).

Animal UsnRNA gene families are often clustered in one part of the genome but their intergenic distances vary from a few hundred base pairs to a few kilobase pairs (reviewed in 22). In animal systems, where the UsnRNA genes form multigene families, the primary sequences within a species display relatively low levels of variation (22). In the two plant UsnRNA multigene families which have been comprehensively characterised: *Arabidopsis* U2 and tomato U1, all of the isolated genes showed sequence variation (10,12). Thus, plant UsnRNAs exhibit greater sequence variation than that observed for animal UsnRNAs (12). As yet there is little information on the genomic organisation of plant UsnRNA gene families with the exception that some of the tomato U1snRNA genes were shown to be closely linked (10). Close linkage has also been found for potato U1snRNA genes (11) and for two maize U5snRNA genes (18).

Comparisons of the 5′ and 3′ flanking sequences of plant UsnRNA genes identified three conserved regions : the Upstream Sequence Element (USE) and TATA motif at approximately −70 and −30 relative to the transcription start site, and the 3′ adjacent sequence directly following the coding region (12,17). Deletion constructs transiently expressed in protoplasts have shown the USE and TATA sequences to be required for plant UsnRNA expression (12,15) and mutational analysis has begun to dissect these sequences to determine the importance of specific nucleotides, thus allowing predictions of relative expression levels for cloned UsnRNA genes to be made (23). Three of the six *Arabidopsis* U2snRNA genes and one of the eight tomato U1snRNA genes have been demonstrated to be active (10,12) and on the basis of regulatory sequences required for expression (12,23) the other *Arabidopsis* U2snRNA and tomato U1snRNA genes could be predicted to be active.

* To whom correspondence should be addressed

The functional significance of the UsnRNA sequence variation in plants is not yet understood but may be related to variation in the levels of expression of individual UsnRNA genes as observed for members of the *Arabidopsis* U2snRNA gene family (12). To examine whether different UsnRNAs (and therefore UsnRNPs) could play a role in plant development by differential splice site selection and differential pre-mRNA processing, the genomic organisation of the UsnRNA gene families in plants and the levels of expression of individual members are important aspects which remain to be established. In this paper, the genomic organisation, sequence variation and expression of U2snRNA genes in potato is reported.

MATERIALS AND METHODS

Plant materials

Five potato cultivars of different origin were selected to represent a broad genetic diversity of cultivated potatoes. The cultivars, Maris Piper, Record, BH15, Cara and Russett Burbank were developed in the UK, Netherlands, France, Ireland and the USA respectively.

Materials

Restriction enzymes were purchased from Pharmacia, Koch-Light and Boehringer-Mannheim. T4 DNA ligase, alkaline phosphatase, RNase A, RNase T_1, SP6 RNA polymerase and RNase inhibitor were obtained from Boehringer-Mannheim. Nylon transfer membrane was from Bio-Rad and Amersham. RNase-free DNase and radionucleotides were also from Amersham. Sequenase was obtained from United States Biochemicals.

Southern analysis

DNA was isolated from freeze dried leaf material from different potato cultivars using the CTAB extraction method (24) with additional phenol and chloroform extraction steps. Approx. 7 μg of DNA was digested with the relevant restriction enzyme (2U/μg), electrophoresed on 1% agarose gels and transferred to Zetaprobe nylon membrane. The membranes were hybridised with the random primed 560 bp Hind III fragment of the *Arabidopsis* U2snRNA gene clone, pGEM U2.3 (12) (a kind gift from Drs P. Vankan and W. Filipowicz, Friedrich Miescher Institute, Basel). Prehybridisation and hybridisation in the presence of 10% dextran sulphate were by standard procedures. Following a final wash at 65°C in 0.1×SSC, 0.1% SDS, hybridising fragments were detected by autoradiography.

Isolation of potato U2snRNA genes

An Mbo I-partial potato genomic library in lambda EMBL4 was screened using standard procedures with the 560 base pair (bp) Hind III fragment of the *Arabidopsis* U2 clone, pGEM U2.3. Three U2-containing clones, lambda PotU2-4, lambda PotU2-11 and lambda PotU2-22 were isolated (Fig. 1). The 4.8 kilobase pair (kbp) Eco RI fragment from lambda PotU2-4 was cloned into pUC13 to produce pPotU2-4.O14. Lambda PotU2-11 contained two U2snRNA genes which were cloned on a 6.0 kbp Bam HI and a 3.3 kbp Eco RI fragment into pUC13 to give the plasmids pPotU2-11.O5 and pPotU2-11.O11 respectively. The U2snRNA gene from lambda PotU2-22 was cloned into pUC13 as a 3.4 kpb Hind III-Eco RI fragment to give plasmid pPotU2-22.O1. The complete sequences of the coding regions and extensive 5′ and 3′ flanking regions of the four U2snRNA genes were obtained by the chain termination method with Sequenase using primers complementary to the *Arabidopsis* U2snRNA sequence in positions 4 to 21 and 68 to 51 (12).

Subclones for *in vitro* transcription were constructed by cloning the 525 bp Ssp I fragment from pPotU2-4.O14 into the Sma I site of pSP64 to produce plasmid pPotU2-4.O43. The 600 bp Eco RI-Bgl II fragment from pPotU2-11.O5 was cloned into the Eco RI and Bam HI sites of pSP64 and pSP65 to give plasmids pPotU2-11.O24 and pPotU2-11.O25 respectively. Finally, the 500 bp Eco RV-Cla I fragment from pPotU2-22.O1 was cloned into the Sma I and Acc I sites of pSP64 and pSP65 to give pPotU2-22.O7 and pPotU2-22.O6 respectively.

Preparation of RNA transcripts

pPotU2-4.O43, pPotU2-11.O25 and pPotU2-22.O7 were linearised at either the Eco RI or Hind III sites for the production

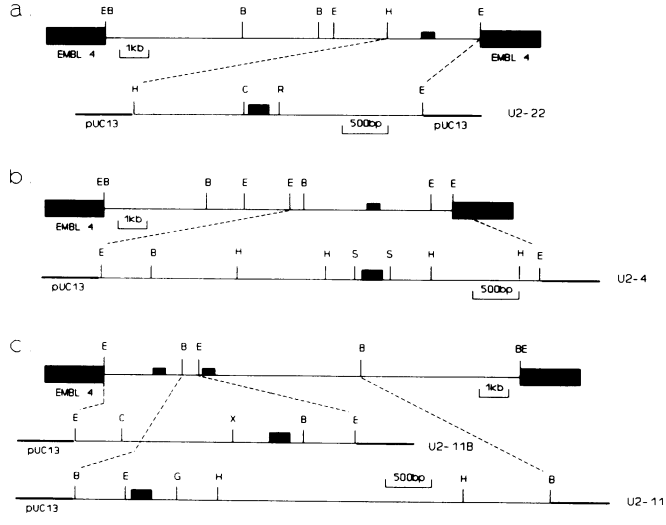

Figure 1. Restriction maps of potato U2snRNA clones. Lambda and plasmid clones of (a) PotU2-22, (b) PotU2-4 and (c) PotU2-11 and PotU2-11B. B-Bam HI, C-Cla I, E-Eco RI, G-Bgl II, H-Hind III, R-Eco RV, S-Ssp I and X-Xba I. Small boxes represent the U2 coding regions.

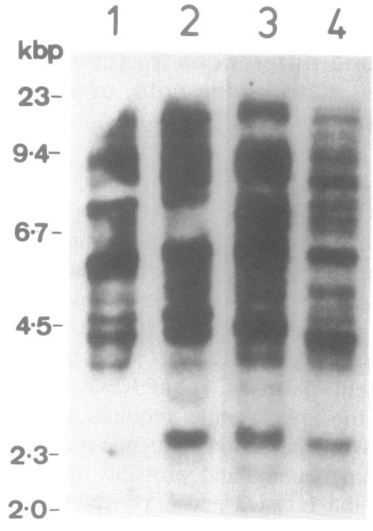

Figure 2. Southern analysis of U2snRNA genes in four potato cultivars. Total genomic DNA, digested with Eco RI, was hybridised with a [^{32}P]-labelled *Arabidopsis* U2snRNA probe. Lane 1—Cara; 2—Record; 3—Maris Piper and 4—Russett Burbank.

of [^{32}P]-labelled RNAs complementary to the PotU2-4, PotU2-11 and PotU2-22 coding sequences. Labelled control RNAs representing the coding sense of PotU2-11 and PotU2-22 were generated from linearised plasmids pPotU2-11.O24 and pPotU2-22.O6. *In vitro* transcription was carried out with 2μg of linearised plasmid DNA in a 40 μl reaction volume containing 40 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM DTT, 100 μg/ml BSA, 500 μM each of ATP, CTP and GTP and 20 μM UTP, 20 μCi [^{32}P]-UTP, 10 units of SP6 polymerase and 50 units of RNase inhibitor at 40°C for 75 min. (25). RNA transcripts were gel purified (26) prior to RNase A/T_1 protection analyses.

RNA extraction and RNase A/T_1 protection assay

RNA was prepared from leaf, stem, root and tuber material from the potato cultivar, Maris Piper, and from leaf tissue from the cultivars, Record, Cara, BH15 and Russett Burbank using the guanidinium isothiocyanate method (27). RNA preparations were treated with RNase-free DNase prior to the protection analyses. 10 μg of total RNA was precipitated with 1×10^5 dpm of [^{32}P]-labelled RNA transcripts. The RNAs were dissolved in 10 μl of 80% formamide, 400 mM NaCl and 10 mM Pipes, pH 6.4, denatured at 90°C for 2 min. and hybridised at 45°C for 16 h.. The RNase A and T_1 digestion was carried out by the addition of 100 μl of RNase buffer (10 mM Tris-HCl, pH 7.5, 0.3 M NaCl, 5 mM EDTA) containing 8 μg/ml of RNase A and 0.4 μg/ml of RNase T_1 at 25°C for 40 minutes (7). The reaction was stopped by the addition of 2 μl 10% SDS and 3 μl Proteinase K (10 mg/ml) and incubated at 37°C for 15 min.. Following phenol/chloroform and chloroform/isoamyl alcohol extractions and the addition of 5 μg carrier tRNA, the RNA duplexes were ethanol precipitated. The RNA was separated on a 6% polyacrylamide-urea sequencing gel alongside a DNA sequence marker and autoradiographed.

RESULTS

Organisation of the potato U2snRNA gene family

Southern analysis of genomic DNA from different potato cultivars, digested with either Eco RI (Fig. 2), Hind III, Eco RV or Bam HI (results not shown) showed great variation in the number and size of bands hybridising with the *Arabidopsis* U2snRNA probe. From the number of hybridising fragments it can be estimated that the potato genome potentially contains twenty-five to forty U2snRNA genes although it is not possible to distinguish between actively transcribed genes and inactive or pseudogenes. The random size distribution of the hybridising DNA fragments with all four enzymes and the varying intensities of the hybridisation signals suggest that potato U2snRNA genes are not arranged in tandem repeats and may show some dispersal through the potato genome. The four potato cultivars exhibited a considerable degree of polymorphism in the lengths of hybridising fragments.

Isolation of potato U2snRNA genes and sequence analysis

Of fifteen potato U2snRNA lambda clones isolated using the *Arabidopsis* U2snRNA gene probe, three were subcloned and sequenced. Two clones contained a single U2snRNA gene while

```
                                                        -150                                                 -100

U2-11     TATTACTCATTAACCTTCATAATTAATCTTTTCACTGCTCTTTGATTAATAATATATTAATGAGATTGAATTAGTTTAATATAAAGTTAAGAGATTAATT
U2-11B    ATCA*TAA*AAG*GAC*AGAT*G*CTCA*CA*ATAAATATA*A*TACT**C**TA*A**G*A*A*CCAG*GA**AGATTAC**GT*AATCTT*TC*CT*G
U2-4      G*AATG*G*GC**ATA*T*AC*AAG*GTAC*CTTTGAAGT**CA*ACTT****G*T**C*A*T**G**CG*CTCA*CTTGT*TT****C*AGTT*CT***
U2-22     *TA*G*AAG**GTTTA*ATA*C*G*GAAGA*GG*TCATCAGG*AT*C*T*TCC**G*G*GGAG*CAGC*GC*GA*A*CTATG**GCAA*GATA**GGGGA

                -90                                     -50

U2-11     TGTGAATCTTCAAGCTCTTCTAGTCCCATACAGACTGCAAAGAGAAATTTCATGTAACAAAGTGGCATATAAAACAGCTCGGCATAACTCGTGGAAAATC
U2-11B    GAAA**AG*AT*GAAA**CT*CTAGT*CC***CTGACTGG*A***C*GA*TT*A*GTA*C*AAATG*C******T**TG*********A*A*******G*
U2-4      *T*CT*GTG**TGTA*G**AGTT*****C*TC**AG*****AGC*GGAC*TGAACGCA*GT**TTTG**AT***AGC*AGTCAC*T**A*TA*AC***G*
U2-22     *A*T*T*T*AT*GAGCT**A**TGTA*CACATCGGAAT**GATA*TG*A*AGACAGC*G**AGCTGC**C*T**GCAAATATTCGTTGAGT**TT***G*

                                                         50                                                  100

U2-11     ATACCTTTCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTAATATCTGATATGTGAGTCATCGACTCACACGATATTAACTCTA
U2-11B    ****************************************************************************************************
U2-4      *******************************************************************G*C***T*GTC****T***********A*
U2-22     *******************************************************************G*C***T*GTC****************A*

                                                         150                                                  200

U2-11     TTTTTTGAGGGG-AGGGTCTATCACGG-TAGCTTGCTATCGGGACCTTTAAGCATCGCCTTGGTGCTGCACTATAACCTTGGCCTGGTGCACCGTACCCCAC
U2-11B    ************G*A****C*C*****-*************A******C***TC******A****T*******CTG***********CA****-----****
U2-4      ******A*****A**A****G*T*A**T**********C*T**GTTC**GT**G*****CAT*C*T*******CTG*ACG*******C*****-----****
U2-22     ******A*****A**A***C**T*A**-*********GC*T**G*TC**GC**G*****CAT*C*T*******CTG*ACG*******C*****-----****

                                                        250                                                  300

U2-11     CAATTCTATTCCAAAATTTAGATTATTTCAGTTGATATGTAATAATGTTCTTTTGTAAAGAATTCTATGAATCTACAAAATATTTTTGTTGTTATGAAAG
U2-11B    ********G*TTT***AACCTT********TAAC*GTACAC*ATT*A*C*CG*CTA*T*TTGGG*AG*TTGATCCAT***AGC*C*CTC*C*ATGTTCCA
U2-4      ******A*G*T****T**CTT*A*TGC***TAAAT*C*T*TCGTG*TACA*****G**T*TTAGAAC*TTCCT*CGCC*CA*A**AGC**TGA**T*CG*
U2-22     ***A****G*A***CT***TTCA*T**GAT**ATCAT*TGTT*TG*T*CT*CA*AA*TTTT**CGATGT**CAA*TCTGTATACAG*C**A*C*AC**GC
```

Figure 3. Nucleotide sequences of potato U2snRNA genes. The sequences of PotU2-11B, PotU2-4 and PotU2-22 are directly compared to that of PotU2-11. An asterisk indicates indentity with the PotU2-11 sequence in the top line. Gaps have been introduced to maximise alignment of the coding regions of the genes (boxed). Regulatory sequences conserved in plant UsnRNA genes are underlined.

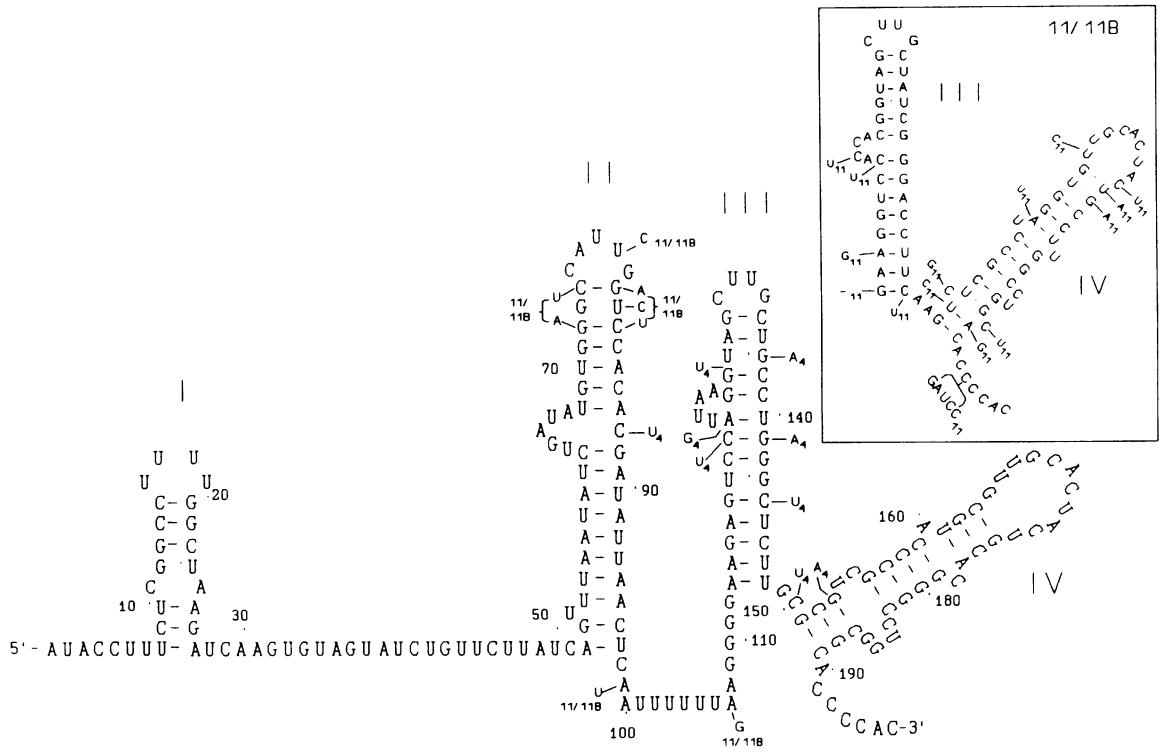

Figure 4. Secondary structure of the potato U2-22 snRNA deduced from the DNA sequence. The structure is based on the model of Keller and Noon (29). Sequence variations in the PotU2-4 sequence, and the PotU2-11 and 11B sequences over the first 110 nt are indicated on the main structure. The secondary structure of PotU2-11B over stem-loops III and IV is shown in the inset where sequence differences in PotU2-11 are indicated.

	USE	TATA	Coding	3′ adjacent sequence
Plant UsnRNA	5′...RTCCCACATCG.....33-34.....TATAaata.....23-27.....C	UsnRNA	CAN_{4-8}	AGTN$\frac{A}{T}$AA.....3′
Potato U2-11	5′...GTCCCATACAG.....34.....TATAaaac.....25.....C	UsnRNA	CAN_5	ATTCCAA.....3′
Potato U2-11B	5′...GTCCCACACTG.....34.....TAAAatag.....23.....C	UsnRNA	CAN_5	AGTTTTA.....3′
Potato U2-4	5′...TTCCCACATCG.....35.....TAATaaaa.....25.....C	UsnRNA	CAN_6	AGTTCAA.....3′
Potato U2-22	5′...GTACCACATCG.....33.....TACAtaag.....25.....C	UsnRNA	CAN_5	AGTACAA.....3′

Figure 5. Comparison of potato U2snRNA gene 5′ and 3′ regulatory sequences with the plant UsnRNA concensus sequences (17). Asterisks indicate positions of single nucleotide differences.

the third contained two, closely linked genes (Fig. 1). A direct comparison of the coding regions and flanking sequences of the four genes is shown in Fig. 3. The 5′ and 3′ ends of the coding region have been determined by comparison to the sequence of *Arabidopsis* U2snRNA genes (12) taking into consideration the high degree of sequence conservation in U2snRNA genes from different species (28). The PotU2-11B and PotU2-22 genes potentially code for RNAs of 196 nt in length. PotU2-4 potentially codes for a 197 nt RNA while PotU2-11 would code for a 200 nt RNA. The increased size of the latter RNA is due to a 5 nt insertion at the 3′ end of the RNA at position +192 (Fig. 3). The sequence of this insertion preserves the sequence, CCCCAC at the extreme 3′ end of PotU2-11 as is found in the other potato U2snRNA sequences.

The first 71 nt of all four genes are identical while the rest of the coding regions are more variable (Fig. 3). From the sequence comparison, the genes can be divided into two pairs: PotU2-11 and PotU2-11B, and PotU2-4 and PotU2-22. Pot U2-11 differs from PotU2-11B by 16 base substitutions, 5 base insertions and 1 deletion giving 83% homology from position 72 to the end of the coding region. Similarly, PotU2-4 differs from PotU2-22 by 6 substitutions, 1 insertion and 2 deletions giving 93% homology over the 3′ part of the coding region from position 72. Comparisons between individuals of the two groups give

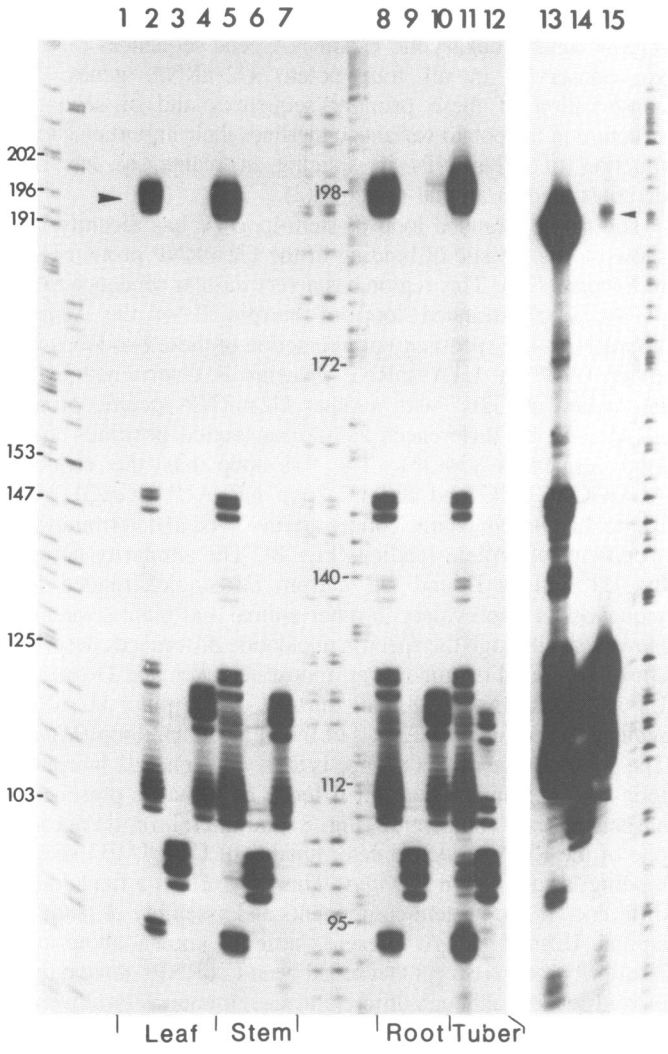

Figure 6. RNase A/T_1 protection analyses of potato leaf, stem, root and tuber RNA with different U2snRNA complementary probes. Leaf RNA (lanes 1–4 and 13–15), stem RNA (lanes 5–7), root RNA (lanes 8–10) and tuber RNA (lanes 11 and 12) were hybridised with a PotU2-22 coding sense probe (lane 1), a PotU2-22 complementary probe (lane 2,5,8,11 and 13), a PotU2-11 complementary probe (lanes 3,6,9,12 and 14) or a PotU2-4 complementary probe (lanes 4,7,10 and 15) and protected RNAs separated on a sequencing gel. Lanes 13 to 15 have been exposed for a longer period than lanes 1 to 12. Full length protected transcripts are arrowed.

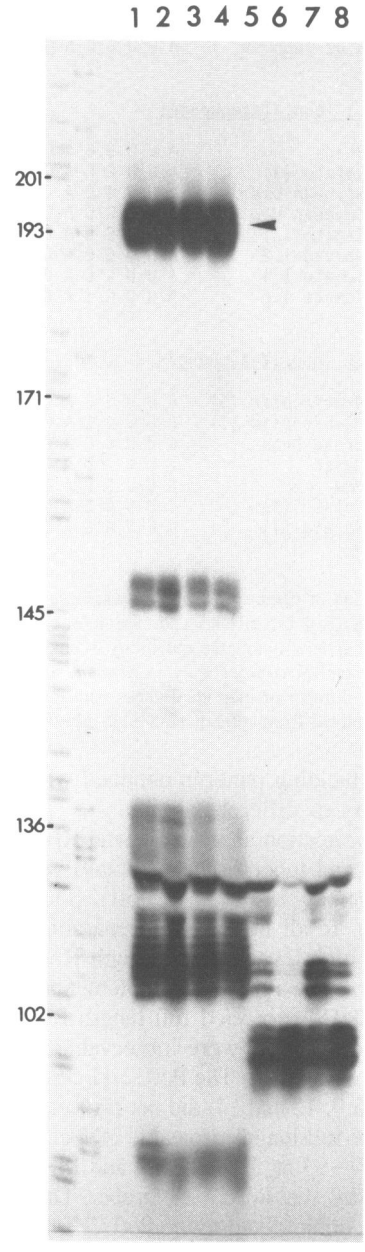

Figure 7. RNase A/T_1 protection analyses of leaf RNA from different potato cultivars. RNA fragments protected from RNase A/T_1 digestion following hybridisation of total potato leaf RNA with [^{32}P]-labelled potato U2–11 and U2–22 RNA probes. a) Lanes 1 and 5–Record; lanes 2 and 6—Cara; lanes 3 and 7—BH15 and lanes 4 and 8–Russett Burbank. The PotU2-22 complementary probe was used for lanes 1–4 and the PotU2-11 complementary probe for lanes 5–8. Full length protected transcripts are arrowed.

reduced% homologies of between 67% and 74% over this same region. Many of the sequence changes are compensatory allowing all four RNAs to be folded into the secondary structure proposed for the human U2snRNA gene (29) (Fig. 4). Alternatively, all four genes contain the Box A (positions 48-53 and 62-67) and Box B (positions 54-61 and 89-95) palindromic sequences allowing the RNAs to be folded into a pseudoknot structure predicted from sequences which are conserved in all known U2snRNA genes (30).

The USE, TATA and 3′ adjacent sequences of the potato U2snRNA genes are compared to the plant UsnRNA consensus sequences in Fig. 5. The USE regions of PotU2-4 and PotU2-22 differ from the consensus by a single nucleotide while PotU2-11B and PotU2-11 differ in 2 and 3 positions respectively. The TATA element is conserved in PotU2-11 while PotU2-4, PotU2-11B and PotU2-22 have TAAT, TAAA and TACA respectively. The spacing between the USE and TATA element (33–34 nt) and the TATA element and the transcription start site (23–27 nt) (23,31) is conserved in all four potato U2snRNA genes.

Expression of potato U2snRNAs

Expression of U2snRNAs in potato has been examined by Northern analysis and by RNase A/T_1 protection analysis. A high resolution Northern analysis with total RNAs from Maris Piper leaf, stem, root and tuber tissues and from leaf tissues of the other four cultivars showed the U2snRNA population to range in size from 195–200 nt (results not shown). No variation was

Xenopus U1 loop II	A U U G C A C U C C
Xenopus U2 loop IV	A U U G C A G U A C C

Plant U1 loop II sequences

Bean	A U U G C A C U C A
Potato (3)	A U U G C A C U U U
Soybean U1a	A U U G C A C U G A
Soybean U1b	A U U G C A C U U A
Tomato (5)	A U U G C A C U U U
Tomato 1.2	A U U G C A C U U C
Tomato 1.4	A U U G C A C U U A
Tomato 1.6	A U U G C A C U A A

Plant U2 loop IV sequences

Arabidopsis (5)	G U U G C A C U A C U
Arabidopsis (1)	G U U G C A C U A U U
Broad bean	G U U G C A C U A U A
Maize	G U U G C A C U A U U
Pea	G U U G C A C U A U A
Potato (3)	G U U G C A C U A C U
Potato (1)	G C U G C A C U A U A

Figure 8. Comparison of plant and *Xenopus* sequences of the single-stranded regions of stem-loop II of U1snRNAs and stem-loop IV of U2snRNAs. The nucleotides involved in determining the specificity of binding of U1A and U2B″ to U1snRNA and U2snRNA respectively in *Xenopus* are boxed. Numbers in brackets refer to the number of individual genes with that particular sequence. Data have been extracted from references 8–11, 12–14 and 34.

observed in the banding patterns obtained with RNA from the different cultivars or different organs.

RNase A/T_1 protection analysis of total RNA extracted from leaf, stem, root and tuber tissue of potato with labelled RNA probes complementary to PotU2-4, PotU2-11 and PotU2-22 are shown in Fig. 6. In all four RNA preparations, the PotU2-22 probe protected high levels of full length RNAs (Fig. 6, lanes 2,5,8,11 and 13). The PotU2-4 complementary probe (Fig. 6, lanes 4,7,10 and 15) protected full length RNAs at extremely low levels (faint signals were observable after longer gel exposures—Fig. 6, lane 15). The PotU2-11 complementary probe (Fig. 6, lanes 3,6,9, 12 and 14) did not protect full length RNAs. In addition to the full length protected RNAs, shorter RNAs of approximately 90–93 nt, 105–125 nt and 148–151 nt in length were protected by the PotU2-22 probe. The PotU2-4 probe (which is >96% homologous to the PotU2-22 probe) was similar to the PotU2-22 probe in showing protected fragments of 105–123 nt in size but otherwise it and and the PotU2-11 probe showed different patterns of protected fragments. No protected RNA fragments were obtained with the [^{32}P]-labelled RNA probes corresponding to the coding sequence of the PotU2-11 and PotU2-22 (Fig. 6, lane 1) genes or when potato or tobacco RNA was substituted with *E. coli* tRNA (results not shown).

The high levels of full length protection signals obtained with the PotU2-22 probe relative to those obtained with PotU2-11 probe were also observed in RNase A/T_1 protection analyses of total leaf RNA from Record, Cara, BH15 and Russett Burbank (Fig. 7). This relationship was also seen with total leaf protoplast RNA extracted from tobacco (*Nicotiana tabacum* L.) (results not shown).

DISCUSSION

Structure of potato U2snRNA genes

The first 70 nt, the Sm antigen binding site (Fig. 3, positions 100–107) and the single stranded loop sequences of stem-loop III, CUUG (positions 130–133) and stem-loop IV, (G)UUGCACUAC(U) (positions 163–172) are the most highly conserved regions among eukaryotic U2snRNA gene sequences (28) and are conserved in all four potato U2snRNA genes. The conservation of these primary sequences and of secondary structure in the potato variants underlines their importance to the function of U2snRNPs in splicing in plants, as has been demonstrated in animal systems (2).

The single stranded loop of stem-loop IV has recently been shown to be the site of binding of the U2snRNP protein U2B″ in *Xenopus* (34). This region has a very similar sequence to that of the single-stranded loop of hairpin II of the *Xenopus* U1snRNA. The specificity of interaction of these two loops with either U2B″ or U1A snRNP proteins is determined by the interaction of U2B″ with another U2snRNP specific protein, U2A′, and by differences at two nucleotide positions in the loop sequences (34,35). The U1 loop has the sequence AUUGCACUCC and the U2 loop has AUUGCAGUACC, where the underlined nucleotides are involved in determining the specificity of protein binding (Fig. 8). The similarity between the U1 hairpin II and U2 hairpin IV single-stranded loop sequences is also evident in other animal and plant sequences. However, although the specific nucleotide differences described above are found in human, rat, mouse, chicken and *Drosophila* U1 and U2snRNAs (28), all thirteen of the plant U2snRNA sequences contain a C instead of the G in the U2 loop (Fig. 8). The consequence of this change is that, in plants, U2 hairpin IV loops differ from U1 hairpin II loops only by the presence or absence of the A nucleotide (Fig. 8) and, therefore, do not have one of the RNA sequence determinants of U1A/U2B″ specific binding, as defined in *Xenopus*. This may reflect a fundamental difference in the protein components and assembly of plant and animal U2snRNPs. At present, little is known about plant UsnRNP proteins except that broad bean UsnRNPs contain many proteins which, although different in size, immunologically cross-react with antibodies raised against animal UsnRNP proteins (36). An investigation of the consequences of the similarity between the single-stranded regions in plant U1 and U2snRNAs, in terms of protein complement and specificity, will require the cloning and characterisation of genes coding for the specific plant UsnRNP proteins.

From mutation analysis of the USE and TATA sequence elements of plant UsnRNA genes, some prediction of the effects of sequence changes in the USE can be made (23). PotU2-4 and 22 differ from the concensus sequence, RTCCCACATCG by having a T in position 1 and an A in position 3 respectively. Neither of these changes affected expression levels of an *Arabidopsis* U2snRNA dramatically (23). PotU2-11B differs by having a C and a T at positions 9 and 10 respectively while PotU2-11 has a T at position 7 and a C and an A at position 9 and 10 respectively. Of these five changes, four are changes of a pyrimidine for a pyrimidine and were not directly tested in the investigation of effects on expression (23). However, the C to A change at position 7, found in PotU2-11, led to much lower expression levels of the *Arabidopsis* U2 gene in protoplasts (23) and, therefore, PotU2-11 may not be expressed. Although, single base mutations in the TATA element were not tested in the mutational analysis, the presence of a C in PotU2-22 in this otherwise strongly AT-rich element could potentially affect both transcription levels and the site of initiation.

Expression of potato U2snRNA genes

Northern and RNase A/T_1 protection analyses showed variation in both transcript length and sequence as has been seen for

Arabidopsis and maize U2snRNAs (12,14). Transcript sequence variation can be deduced from the RNase A/T_1 protection analysis because the first 71 nt of the 195−200 nt coding region are identical in all plant U2snRNA genes. Thus, a protected fragment of length greater than 72 nt is highly likely to represent a transcript which is identical to the labelled probe from position +1. For example, the fragments of 148−151 nt protected by the PotU2-22 probe (Fig. 6) must represent transcripts which are identical (within the limits of detection of the RNase A/T_1 protection assay) to the PotU2-22 sequence from position +1 to positions +148 −151 at which point sequence divergence has permitted digestion by RNase A/T_1. Thus, the patterns of protected fragments point to greater sequence variation among potato U2snRNA genes than is represented by the four cloned genes presented here.

In the four potato organs analysed, full length transcripts homologous to the PotU2-22 gene were detected at high levels while transcripts corresponding to PotU2-4 were only visible after longer gel exposures. The relative intensities of the fragments protected by the PotU2-22 probe indicate that a major subset of potato U2snRNA transcripts are complementary to this probe. From the estimated U2snRNA gene number it is likely that these transcripts derive from a number of genes with coding sequences which are virtually identical to that of PotU2-22 and, therefore, that these genes represent a subset of functional potato U2snRNA genes. The shorter protected fragments represent transcripts from other genes which are identical to PotU2-22 over the first 150 bp or to PotU2-4 and PotU2-22 over the first 100−125 bp. The low level of expression of PotU2-4 could not be predicted on the basis of its USE or TATA-like sequence and may reflect a low copy number of gene(s) with sequences identical to PotU2-4. The lack of expression of PotU2-11 and the nucleotide changes in its USE suggest that it is a non-functional pseudogene. However, the coding sequences of PotU2-11 and PotU2-11B differ substantially from PotU2-22 and PotU2-4. It is unlikely that the PotU2-11/11B sequences have diverged from the PotU2-22/4 sequences by virtue of their being pseudogenes because they have not accumulated mutations in, for example, the first 71 nt of the coding region. Therefore, they may still represent a distinct class of sequence variants whose expression pattern is yet to be determined.

The four potato cultivars used in the expression analysis were selected on the basis that they reflected a broad genetic diversity in cultivated potatoes. The variation observed in band size and intensity in the Southern analysis was consistent with the expected genetic variation among the cultivars. However, the RNase protection patterns were virtually identical for all four cultivars with the PotU2-22 and PotU2-11 probes suggesting that the population of U2snRNA transcripts in the leaf tissue of these cultivars are very similar. The contrasting variation observed at the DNA and RNA transcript levels suggests that the coding sequence of expressed classes of potato U2snRNA gene variants is highly conserved while there is extensive sequence divergence in their flanking regions (as also evident in the four gene sequences presented in Fig. 3).

From this initial expression analysis of U2snRNA transcripts in five potato cultivars and four different organs of potato using RNA probes made from three of the four cloned genes, there is no evidence for cultivar or organ specific expression of particular U2snRNA variants. In some animal systems, UsnRNA gene variants are expressed at different stages of development and in different tissues. For example, in *Xenopus* different sets of U1snRNA and U4snRNA genes are expressed during early embryogenesis, in tadpoles and in the adult frog (37,38; see also 21). Furthermore, in mouse, two U1snRNA gene variants are expressed at different stages of embryo development (39) and two chicken U4snRNA genes are expressed to different levels in different tissues and developmental stages (40). These observations have led to the proposal that the differential expression of UsnRNA genes during development may alter patterns of pre-mRNA processing by, for example, affecting splice site selection, particularly in alternative splicing pathways (21,37−40). Although only one alternative splicing event has as yet been described in plants (41), the extensive sequence variation in plant UsnRNAs may provide the basis for regulation of gene expression at the pre-mRNA splicing level. Further analysis of potato U2snRNA gene variant expression will require the cloning of other sequence variants and a more detailed analysis of different potato tissues and developmental stages. In order to relate the relative intensities of expression of different subsets of sequence variants to differences in either their regulatory sequences or their gene copy number, the detailed molecular organisation of the potato U2snRNA multigene family is currently being investigated.

ACKNOWLEDGEMENTS

We would like to thank Dr. W. Powell for his support and encouragement and Dr. F. Guerineau for critically reading the manuscript. We are grateful to Drs P. Vankan and W. Filipowicz for the kind gifts of the *Arabidopsis* U2.3 gene and U2 sequencing primers. This research was supported by grants from the Agricultural and Food Research Council and the Nuffield Foundation.

REFERENCES

1. Krainer, A.R. and Maniatis, T. (1988) In Transcription and Splicing, Hames, B.D. and Glover, D.M. eds., IRL Press, Oxford, pp. 131−206.
2. Steitz, J.A., Black, D.L., Gerke, V., Parker, K.A., Krämer, A., Frendewey, D. and Keller, W. (1988) In Birnstiel, M. (ed.), Small Nuclear Ribonucleoprotein Particles. Springer Verlag, Berlin, pp 115−154.
3. Wiebauer, K. Herrero, J.-J. and Filipowicz, W. (1988) Mol. Cell Biol. *8*, 2042−2051.
4. Goodall, G.J. and Filipowicz, W. (1989) Cell *58*, 473−483.
5. Van Santen, V.L. and Spritz, R.A. (1987) Gene *56*, 253−265.
6. Goodall, G.J. and Filipowicz, W. (1990) Plant Mol. Biol. *14*, 727−733.
7. Goodall, G.J., Wiebauer, K. and Filipowicz, W. (1989) Meths. Enzymol. *181*, 148−161.
8. Van Santen, V.L. and Spritz, R.A. (1987) Proc. Natl. Acad. Sci. USA. *84*, 9094−9098.
9. Van Santen, V.L., Swain, W.L. and Spritz, R.A. (1988) Nucl. Acids Res. *16*, 4176.
10. Abel, S., Kiss, T. and Solymosy, F. (1989) Nucl. Acids Res. *17*, 6319−6337.
11. Vaux, P., Waugh, R. and Brown, J.W.S. Manuscript in preparation.
12. Vankan, P. and Filipowicz, W. (1988) EMBO J. *7*, 791−799.
13. Hanley, B. A. and Schuler, M. A. (1989) Nucl. Acids Res. *17*, 10106.
14. Brown, J.W.S. and Waugh, R. (1989) Nucl. Acids Res. *17*, 8991−9001.
15. Kiss, T. and Solymosy, F. (1990) Nucl. Acids Res. *18*, 1941−1949.
16. Marshallsay, C., Kiss, T. and Filipowicz, W. (1990) Nucl. Acids Res. *18*, 3459−3466.
17. Vankan, P., Edoh, D. and Filipowicz, W. (1989) Nucl. Acids Res. *16*, 10425−10440.
18. Leader, D., Waugh, R. and Brown, J.W.S. manuscript in prepartion.
19. Waibel, F. and Filipowicz, W. (1990) Nucl. Acids Res. *18*,
20. Szkukalek, A., Kiss, T. and Solymosy, F. (1990) Nucl. Acids Res. *18*, 1295.
21. Veretnik, S. and Rubenstein, I. (1990) Nucl. Acids Res. *18*, 8661.
22. Dahlberg, J.E. and Lund,E. (1988) In Birnstiel, M. (ed.), Small Nuclear Ribonucleoprotein Particles. Springer Verlag, Berlin, pp 38−70.
23. Vankan, P. and Filipowicz, W. (1989) EMBO J. *8*, 3875−3882.
24. Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A. and Allard, R. W. (1984) Proc. Natl. Acad. Sci. USA *81*, 8014−8018.

25. Melton, D.A., Kreig, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M. (1984) Nucl. Acids, Res. *12*, 7035–7056.
26. Frendewey, D. and Keller, W. (1985) Cell *42*, 355–367.
27. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
28. Reddy, R. and Busch, H. (1988) In Birnstiel, M. (ed.), Small Nuclear Ribonucleoprotein Particles. Springer Verlag, Berlin, pp 1–37.
29. Keller, E.B. and Noon, W.A. (1985) Nucl. Acids Res. *13*, 4971–4981.
30. Igel, A.H. and Ares, M. (1988) Nature *334*, 450–453.
31. Waibel, F. and Filipowicz, W. (1990) Nature *346*, 199–202.
32. Kiss, T., Abel, S. and Solymosy, F. (1989) Plant Mol. Biol. *12*, 709–711.
33. Kiss, T., Antal, M. and Solymosy, F. (1987) Nucl. Acids Res. *15*, 1332.
34. Scherly, D., Boelens, W., Dathan, N.A., van Venrooij, W.J. and Mattaj, I.W. (1990) Nature *345*, 502–506.
35. Scherly, D., Dathan, N.A., Boelens, W., van Venrooij, W.J. and Mattaj, I.W. (1990) EMBO J. *9*, 3675–3681.
36. Palfi, Z., Bach, M., Solymosy, F. and Lührmann, R. (1989) Nucl. Acids Res. *17*, 1445–1458.
37. Forbes, D.J., Kirschner, M.W., Caput, D., Dahlberg, J.E. and Lund, E. (1984) Cell *38*, 681–689.
38. Lund, E. and Dahlberg, J.E. (1987) Genes and Dev. *1*, 39–46.
39. Lund, E., Kahan, B. and Dahlberg, J.E. (1985) Science *229*, 1271–1274.
40. Korf, G.M., Botros, I.W. and Stumph, W.E. (1988) Mol. Cell Biol. *8*, 5566–5569.
41. Werneke, J.M., Chatfield, J.M. and Ogren, W.L. (1989) Plant Cell *1*, 815–822.