Molecular Cell, ***Volume 37***

Supplemental Information

An Ancient Mechanism for Splicing Control: U11 snRNP as an Activator of Alternative Splicing

Jens Verbeeren, Elina H. Niemelä, Janne J. Turunen, Cindy L. Will, Janne J. Ravantti, Reinhard Lührmann, and Mikko J. Frilander

Supplemental Experimental Procedures

Plasmids

The human 48K USSE with flanking sequences was cloned into a pCR2.1-TOPO vector forming the plasmid p48KUSSE. For the mammalian reporter experiments, p48KUSSE was digested with EcoRI, blunt-ended, and the insert cloned into the pSmE-CFP mammalian expression vector, cut with Eco47III, to form pSmE-CFP-48KUSSE. The entire coding sequence of the human 48K gene was first amplified from a cDNA and then cloned into pCR2.1-TOPO. This was then digested with EcoRI, and the insert subcloned into pCIneo to form pCIneo-48K, which was used for mammalian overexpression studies. For the construction of the plasmid p65KJT, the region of the 3'UTR of the human *RNPC3* gene containing the USSE along with the preceding intronic sequence was amplified from HeLa genomic DNA by PCR using the primers h65K-26 and h65K-27. This fragment was then inserted into pCR2.1-TOPO. For the analysis of the 65K 3´UTR, plasmids pGL4.13F65K, pGL4.13S65K and pGL4.13L65K were constructed by amplifying the entire unspliced 3´UTR, or the short or long 3´UTR of the 65K gene, respectively, and subsequent cloning into the pGL4.13 luciferase expression vector (Promega Corporation, Madison, WI), cut with XbaI. Cloning was performed using the In-Fusion technology (Clontech Laboratories, Mountain View, CA). Primers containing the desired mutations were used to create mutants via a site-directed mutagenesis reaction.

Pattern matching and alignment of USSEs in genomes

The putative USSE [AG]TAT[CT][CT]TN(2,16)[AG]TAT[CT][CT]T was used to find all initial matches in human and mouse genomes that were downloaded from the Ensembl website (Hubbard et al. 2007). Matching was performed with both the USSE and its reverse complement to scan both 5' and 3' directions by running the locally EMBOSS-program 'fuzznuc' (Rice et al. 2000). After initial screening confirmed the existence of the putative USSEs, the pattern matching was expanded to encompass a multiple alignment of 29 vertebrate genomes with the mouse genome (Jul. 2007 (mm9) / multiz30way) from the UCSC Genome Browser website (Kent et al 2002). Positive hits displayed complete agreement in the alignment (except for the ambiguous N(2,16)-part) and were present in at least four genomes in the multiple alignment. Custom-made Python programs were used for alignment matching, scoring, sorting and finally linking the patterns that were found to the UCSC Genome Browser for manual verification and for further processing. Manual verification was used to confirm that the pattern was located within a transcribed region and in the correct orientation with respect to the transcription direction. Phylogenetic conservation plots were generated with MULAN (mulan.decode.org; Ovcharenko et al., 2005) and multiple sequence alignments were performed using the MUSCLE algorithm (www.ebi.ac.uk/Tools/muscle/; Edgar, 2004).

Supplemental Figures

Figure S1. The USSE sequences are conserved in animals as well as plants (related to Figure 1).
(A) Sequence alignment of the conserved intronic region in 48K genes in mammalian, fish and insect species. The polypyrimidine tract, 3'ss, exon 4i and the USSE are indicated. Sequence alignments were performed using the MUSCLE algorithm (www.ebi.ac.uk/Tools/muscle/; Edgar, 2004). The exonic sequence was not identified in insect species.
(B) Sequence alignment of the conserved 3'UTR region of 65K genes in mammals, birds, lizards and fishes. USSE elements, the upstream U2-type 3'ss and the PPT are indicated. The alignment was performed as in (A).
(C) Sequence alignment of the conserved 3'UTR region in plant 48K genes. A comparison of those species for which complete genomic data was available is shown in the upper panel (Plant genomic). USSE elements, an upstream U2-type 3'ss and the polypyrimidine tract are indicated. Alignment was performed as in (A). The sequence of *Physcomitrella patens* was aligned by hand, based on the experimentally determined splice site information available at www.phytozome.net. A comparison of plant 48K gene-specific ESTs containing the USSE element is shown in the lower panel. The sequences were identified from Plant Transcript Assemblies web site (plantta.jcvi.org/) by blast searches using *A. thaliana* and other known full length plant 48K sequences.
(D) Identification of 48K 3'UTR isoforms from *A. thaliana* and *P. trichocarpa*. Total RNA from each plant was used for cDNA synthesis, followed by RT-PCR amplification with 3'UTR-specific primers. Arrows indicate the locations of the primers used for RT-PCR. In both cases the RT-PCR generated three bands as shown in the gel images. The identities of the PCR bands were confirmed by DNA sequencing, which revealed that they represent different splicing isoforms, generated using consensus 5' and 3' splice sites. The gene models representing different 48K splicing isoforms are shown as schematic pictures, and the actual sequences, as determined by sequencing of the RT-PCR products are shown below. In the *Arabidopsis thaliana* panel, the genomic sequence of *Arabidopsis lyrata* is included for comparison.

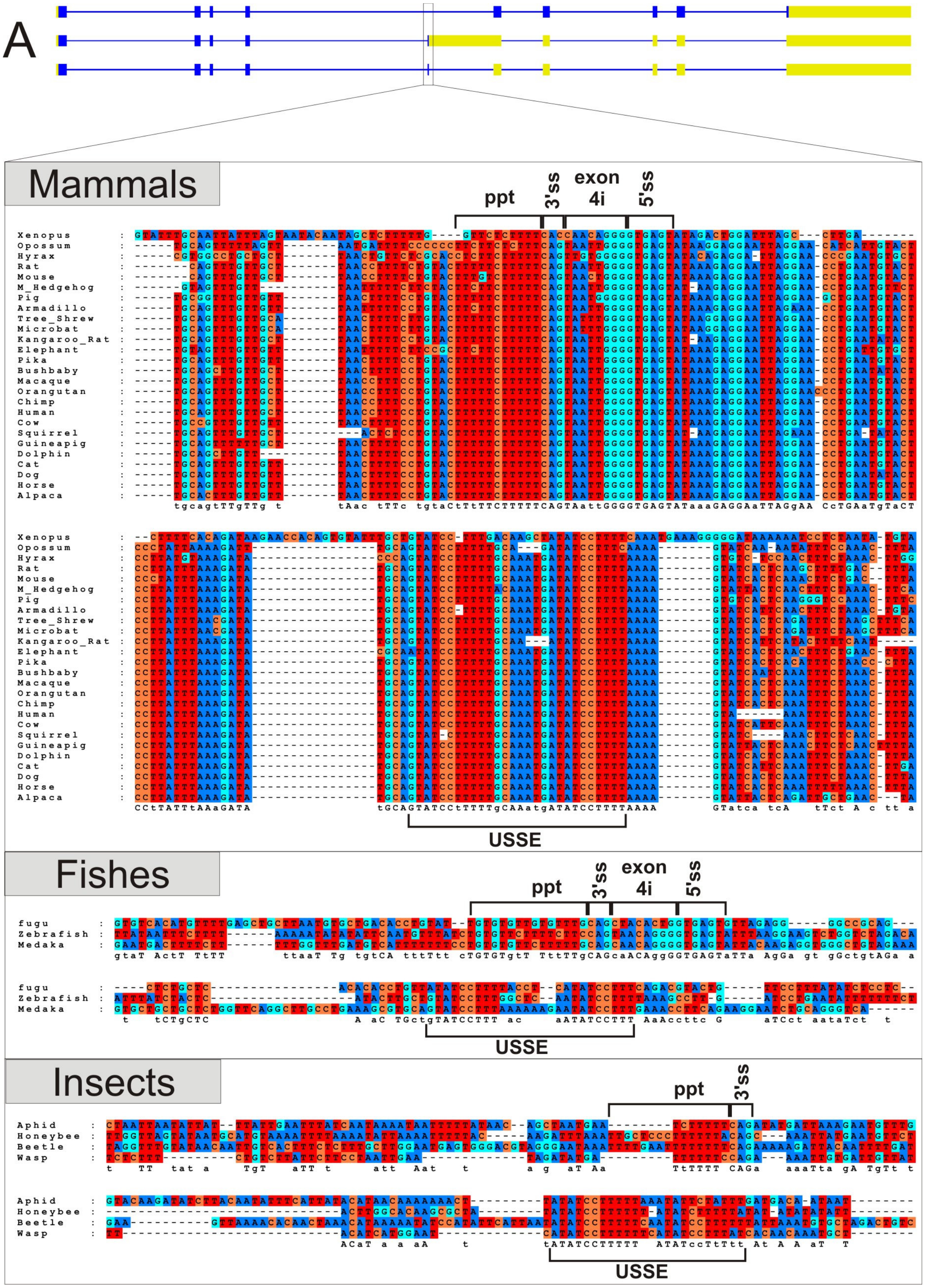
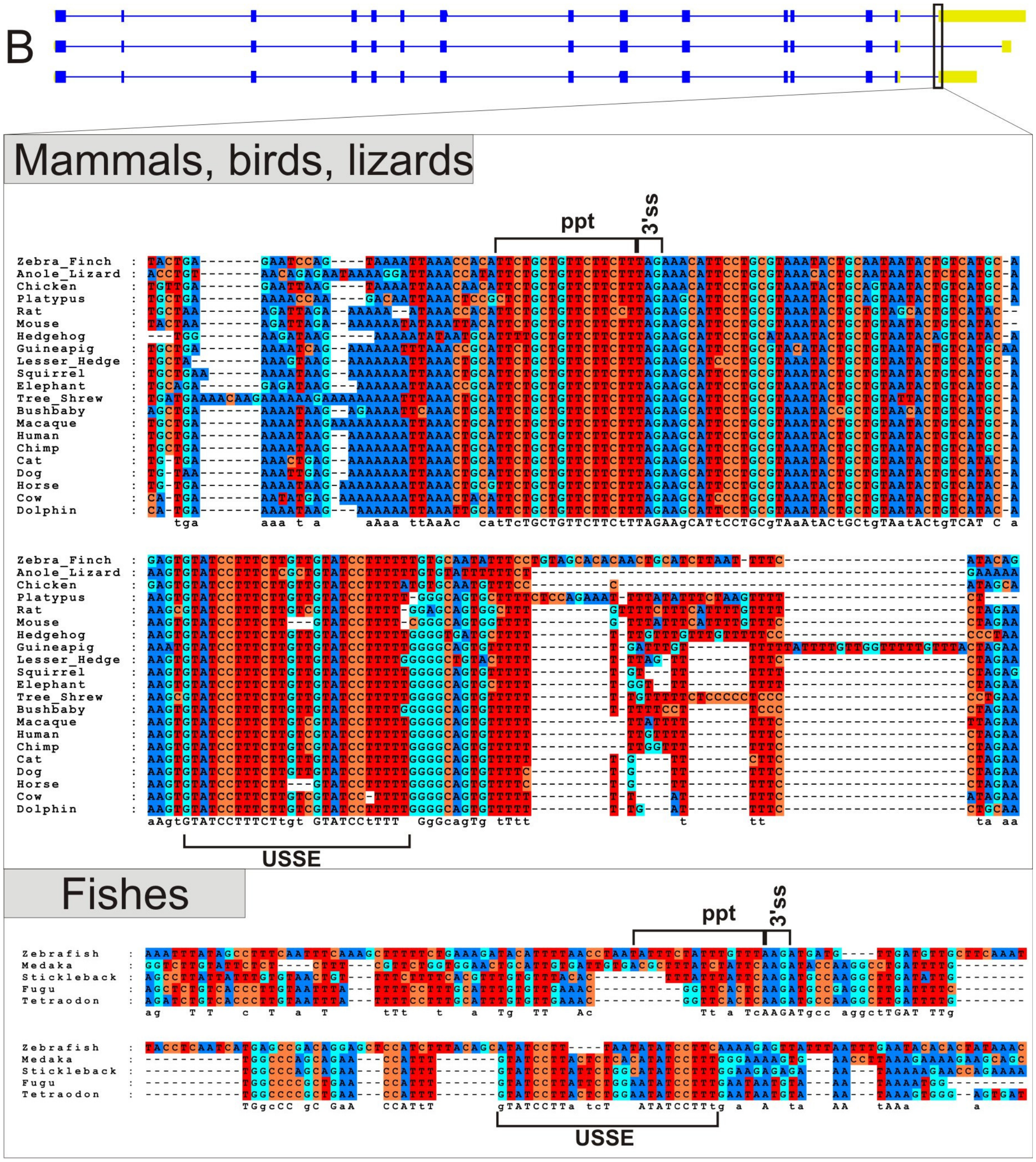
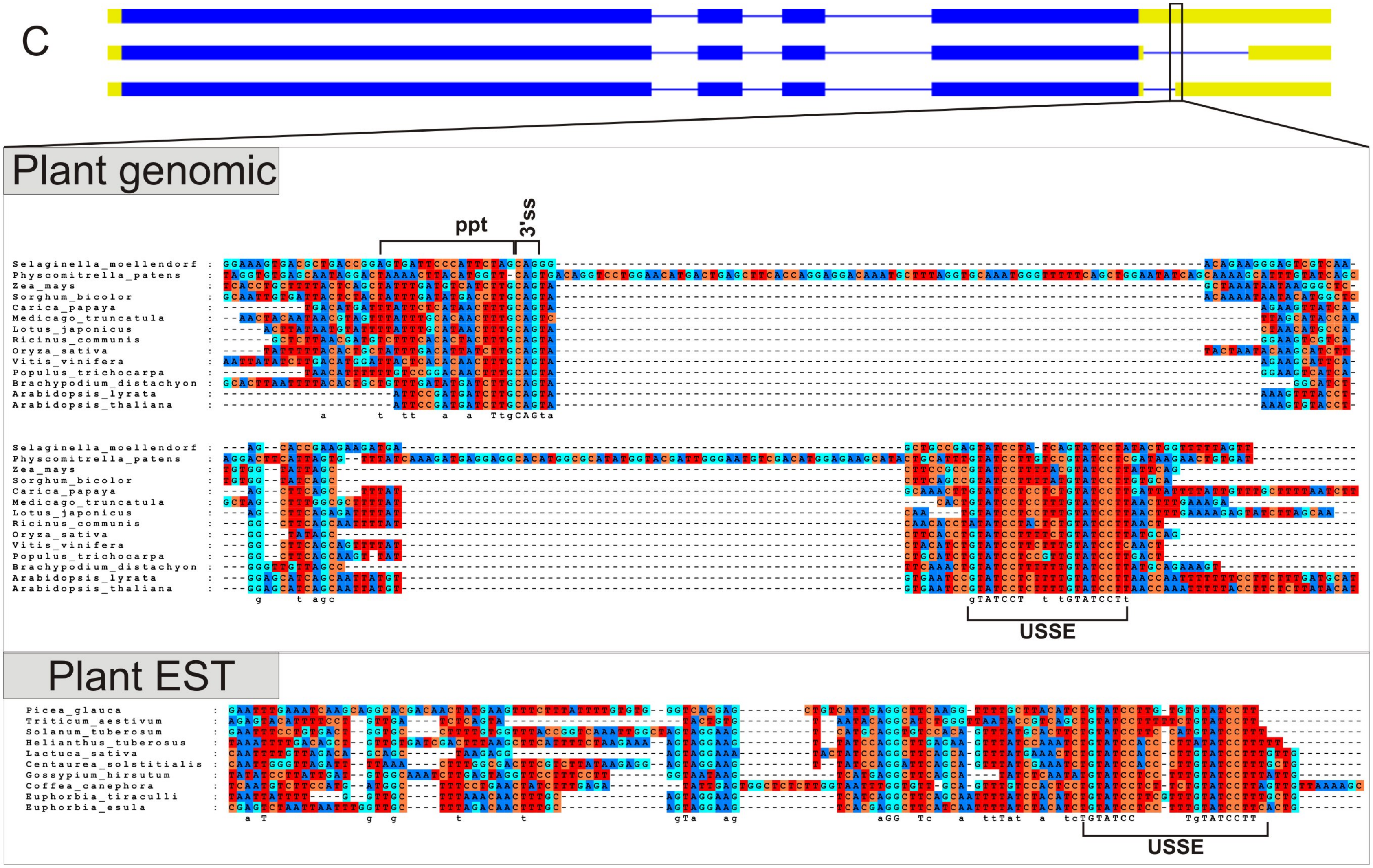
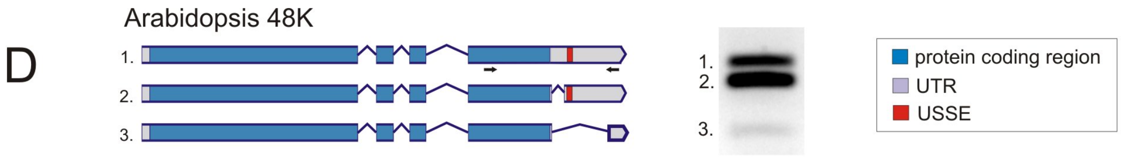

A
Mammals
ppt
3'ss
exon 4i
5'ss
Xenopus
Opossum
Hyrax
Rat
Mouse
M_Hedgehog
Pig
Armadillo
Tree_Shrew
Microbat
Kangaroo_Rat
Elephant
Pika
Bushbaby
Macaque
Orangutan
Chimp
Human
Cow
Squirrel
Guineapig
Dolphin
Cat
Dog
Horse
Alpaca
tgcagtTTgTTg t tAac tTTc tgtaCtTtTTCTTTTTCAGTAattGGGGTGAGTATaaaGAGGAaTTAGgAA CcTGAaTgTaCT
CCtTATTtAAaGATA tGCAGTATCCtTTTTgCAAatgATATCCTTTTAAAA GTatca tcA tTct A c tt a
USSE
Fishes
ppt
3'ss
exon 4i
5'ss
fugu
Zebrafish
Medaka
gtaT ActT TtTT ttaaT Tg tgtCA tttTtt cTGTGTgtT TttTTgCAGcaACAggGGTGAGTaTTa AgGa gt gGctgtAGa a
t tCTgCTC A aC TGctgTATCCTTT ac aATATCCTTT AaAccttc G aTCct aataTct t
USSE
Insects
ppt
3'ss
Aphid
Honeybee
Beetle
Wasp
t TT tat a TgT aTT t att Aat t a g aT Aa TtTTTT CAGa aaaTta gA TgTt t
ACaT a a aA t tATATCCTTTTT ATATccTtTtt At A A aT T
USSE

Figure S1 (continued on the next page)


B
Mammals, birds, lizards
ppt
3'ss
Zebra_Finch
Anole_Lizard
Chicken
Platypus
Rat
Mouse
Hedgehog
Guineapig
Lesser_Hedge
Squirrel
Elephant
Tree_Shrew
Bushbaby
Macaque
Human
Chimp
Cat
Dog
Horse
Cow
Dolphin
tga aaa t a aAaa ttAaAc catTcTGCTGTTCTTCtTTAGAAgCATtCCTGCgTAaAtACtGCtgTAatACtgTCAT C a
aAgtGTATCCTTTCTtgt GTATCCtTTT GgGcagTg tTtt t tt ta aa
USSE
Fishes
ppt
3'ss
Zebrafish
Medaka
Stickleback
Fugu
Tetraodon
ag T T c T a T tTt t a Tg TT Ac Tt a TcAAGATgcc aggctTGAT TTg
TGgcCC gC GaA CCATtT gTATCCTTa tcT ATATCCTTtg a A ta AA tAAa a
USSE


C
Plant genomic
ppt
3'ss
Selaginella_moellendorf
Physcomitrella_patens
Zea_mays
Sorghum_bicolor
Carica_papaya
Medicago_truncatula
Lotus_japonicus
Ricinus_communis
Oryza_sativa
Vitis_vinifera
Populus_trichocarpa
Brachypodium_distachyon
Arabidopsis_lyrata
Arabidopsis_thaliana
a t tt a a TtgCAGta
g t agc
gTATCCT t tGTATCCTt
USSE
Plant EST
Picea_glauca
Triticum_aestivum
Solanum_tuberosum
Helianthus_tuberosus
Lactuca_sativa
Centaurea_solstitialis
Gossypium_hirsutum
Coffea_canephora
Euphorbia_tiraculli
Euphorbia_esula
aT g g t t gTa ag aGG Tc a ttTat a tcTGTATCC TgTATCCTT
USSE



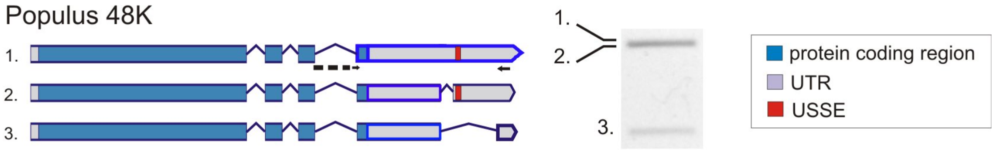
```
                                                                       5'ss
A_lyrata-genomic            : AATCTTATCATCGAAACCACAGATCATCTCGTGAGAAGAGTTCTTCAGATTGTAAGACCAAAAGGGATGATCCTTATGACCGCTGCAGTCGGGAACCTAGAATCAAAATTCTTTTGAAG
A_thaliana-genomic          : AATCTTATCA---AAACTACAGATCATCTCGTGAGAAGAGTTCTTCAGATTATAAGACCAAAAGGGATGATCCTTATGACCGCCGCAGTCAGCAACCTAGGAATCAAAATTTGTTTGAAG
A_thaliana_splice_variant_2 : AATCTTATCA---AAACTACAGATCATCTCGTGAGAAGAGTTCTTCAGATTATAAGACCAAAAGGGATGATCCTTATGACCGCCGCAGTCAGCAACCTAGGAATCAAAATTTGTTTGAAG
A_thaliana_splice_variant_3 : AATCTTATCA---AAACTACAGATCATCTCGTGAGAAGAGTTCTTCAGATTATAAGACCAAAAGGGATGATCCTTATGACCGCCGCAGTCAGCAACCTAGGAATCAAAATTTGTTTGAAG

                                                                       5'ss
A_lyrata-genomic            : ATAGATACATACCAACAGAAAGGGAGTGAACTTAAAAGAAGGTAAGTCATTTAAAAAAACATCTGTCTATCTCCTTACCATCGTGTGGGCTAGTCTGACTTTGAAATGGGGTGGTTTATT
A_thaliana-genomic          : ATAGATACATACCAACAGAAAAAGAGTGAACTTAAAAGAAGGTAAGTCATTTTAAAA---------CTCTCTCTCTATCACCGTGTGGGCTGGTCTGACTTTGAAATGGGGTGGTTTATT
A_thaliana_splice_variant_2 : ATAGATACATACCAACAGAAAAAGAGTGAACTTAAAAGAAG-------------------------------------------------------------------------------
A_thaliana_splice_variant_3 : ATAGATACATACCAACAGAAAAAGAGTGAACTTAAAAGAAG-------------------------------------------------------------------------------

                                   3'ss-2                                           USSE
A_lyrata-genomic            : CCGATGATCTTGCAGTAAAAGTTTACCTGGAGCATCAGCAATTATGTGTGAATCCGTATCCTCTTTTGTATCCTTAACCAATTTTTTTCCTTCTTTGATGCATCTAAGTGTTAACCAAAT
A_thaliana-genomic          : CCGATGATCTTGCAGTAAAAGTGTACCTGGAGCATCAGCAATTATGTGTGAATCCGTATCCTCTTTTGTATCCTTAACCAAATTTTTACCTTCTCTTATACATCTCA-------------
A_thaliana_splice_variant_2 : ---------------TAAAAGTGTACCTGGAGCATCAGCAATTATGTGTGAATCCGTATCCTCTTTTGTATCCTTAACCAAATTTTTACCTTCTCTTATACATCTCA-------------
A_thaliana_splice_variant_3 : ------------------------------------------------------------------------------------------------------------------------

                                                                                                                      3'ss-3
A_lyrata-genomic            : GTTGGTGATATTCTCAAATTTCCAGCTATCTCTTGCGTTTATAGATTCACTGTAGTTGCCATTCCTGTCTGTCTATATATTTTCAGTTTCTTTTTTCTTCTTTGGTTGTGACAGATTGAA
A_thaliana-genomic          : CTTGGTGATACTCTCAAATTTCCAGCTGTC-----------TAGATTCACTGTAGTTACCATTCCCGTCTA--GATATGTTTTCAGTTTC-TTTTTCGTCTATGGTTGTGACAGATTGAA
A_thaliana_splice_variant_2 : CTTGGTGATACTCTCAAATTTCCAGCTGTC-----------TAGATTCACTGTAGTTACCATTCCCGTCTA--GATATGTTTTCAGTTTC-TTTTTCGTCTATGGTTGTGACAGATTGAA
A_thaliana_splice_variant_3 : ------------------------------------------------------------------------------------------------------------------ATTGAA


A_lyrata-genomic            : GTTAATTCAAGTATGTTCTTGCTATGAACAAGCCATGAGCCATGAGATTGCTGGGAAGAAGTGCCAAACTCCA----------TAATTGTGAGAGACTGCAATGTTTTAAGTCTACTATT
A_thaliana-genomic          : GTTAATTGAAGTATGTTTTTGCTGTGAACAAGCCATGAGCCGTGAGATTGCTGGGAAGAAATGGTAAACTCCACAAGTGTTTGTAATTGTGAGAGACTGCAATGTTTTAACTTTTAAGTC
A_thaliana_splice_variant_2 : GTTAATTGAAGTATGTTTTTGCTGTGAACAAGCCATGAGCCGTGAGATTGCTGGGAAGAAATGGTAAACTCCACAAGTGTTTGTAATTGTGAGAGACTGCAATGTTTTAACTTTTAAGTC
A_thaliana_splice_variant_3 : GTTAATTGAAGTATGTTTTTGCTGTGAACAAGCCATGAGCCGTGAGATTGCTGGGAAGAAATGGTAAACTCCACAAGTGTTTGTAATTGTGAGAGACTGCAATGTTTTAACTTTTAAGTC
```



```
Genomic          : TATGACATGCATGAGGATGATGTTTACACTAGCATTACATATGCCAGGGAAGATGTTCATGATTGATCTAGCAGGTGCAGTGCAAAAAGGCAATTAAGCAGATTACCTTGATAAATTCTC
Splice_variant_2 : TATGACATGCATGAGGATGATGTTTACACTAGCATTACATATGCCAGGGAAGATGTTCATGATTGATCTAGCAGGTGCAGTGCAAAAAGGCAATTAAGCAGATTACCTTGATAAATTCTC
Splice_variant_3 : TATGACATGCATGAGGATGATGTTTACACTAGCATTACATATGCCAGGGAAGATGTTCATGATTGATCTAGCAGGTGCAGTGCAAAAAGGCAATTAAGCAGATTACCTTGATAAATTCTC

                                  5'ss                                                                                     3'ss-2
Genomic          : ACTTCAGAATGTGAGGTAAGCTATTGTGGTTATTTTTATTGTTTTGAAAAATGTTTATCTGCACTAGATAAAAGTTCTTGCTTTTAACATTTTTTGTCCGGACAACTTTGCAGTAGGAA
Splice_variant_2 : ACTTCAGAATGTGAG--------------------------------------------------------------------------------------------------TAGGAA
Splice_variant_3 : ACTTCAGAATGTGAG--------------------------------------------------------------------------------------------------------

                                                 USSE
Genomic          : GTCATCAGGCTTCAGCAAGTTATCTGCATCTGTATCCTCCGTTGTATCCTTGACTGTTTTATTTTCTATACTGTTAAGGTCCTTACTGTCTAGGCAGTGATTTCATTATCCTCCTACTGC
Splice_variant_2 : GTCATCAGGCTTCAGCAAGTTATCTGCATCTGTATCCTCCGTTGTATCCTTGACTGTTTTATTTTCTATACTGTTAAGGTCCTTACTGTCTAGGCAGTGATTTCATTATCCTCCTACTGC
Splice_variant_3 : ------------------------------------------------------------------------------------------------------------------------

Genomic          : CCATACAATGCCAGTACTTTTTTCCAATATATGATGGAATAAGTAGTGGTCACATTTGAGTGAAAAAATAATGTCTCGGTTATTTACATAGTGAAAGAGGAAATAAGATTGTACAGGAAC
Splice_variant_2 : CCATACAATGCCAGTACTTTTTTCCAATATATGATGGAATAAGTAGTGGTCACATTTGAGTGAAAAAATAATGTCTCGGTTATTTACATAGTGAAAGAGGAAATAAGATTGTACAGGAAC
Splice_variant_3 : ------------------------------------------------------------------------------------------------------------------------

Genomic          : ATTCTAGCAGAACATGCCTTATTGCAGCTTGTAACTGTTCTGTTATTTATACTGTATATAAAATCTGAAGTATATCATATGTGATTTTGTTGTGTTGACATCAGTACCTTTGTGGATTAA
Splice_variant_2 : ATTCTAGCAGAACATGCCTTATTGCAGCTTGTAACTGTTCTGTTATTTATACTGTATATAAAATCTGAAGTATATCATATGTGATTTTGTTGTGTTGACATCAGTACCTTTGTGGATTAA
Splice_variant_3 : ------------------------------------------------------------------------------------------------------------------------

Genomic          : TAAAATATGCAAATTCTTGGCAATCCATGCTTTGATGGTAAGGTCATTGGTTACCATGGAAGAACTGGTAAATAACCCAAATCGGCAGGGTTTACAGGTTTGCTTAGAATGGTTACCTGT
Splice_variant_2 : TAAAATATGCAAATTCTTGGCAATCCATGCTTTGATGGTAAGGTCATTGGTTACCATGGAAGAACTGGTAAATAACCCAAATCGGCAGGGTTTACAGGTTTGCTTAGAATGGTTACCTGT
Splice_variant_3 : ------------------------------------------------------------------------------------------------------------------------

Genomic          : TTATGATTTTGTTTTTCTAGGTAAAAGCTGATACTTATTGTTTATGGGTCTTTTTTATTGCATACAATATGTATGGATAATTTATTAATTTGAATGTTCTTATAGCTCACGTTGCTTCTT
Splice_variant_2 : TTATGATTTTGTTTTTCTAGGTAAAAGCTGATACTTATTGTTTATGGGTCTTTTTTATTGCATACAATATGTATGGATAATTTATTAATTTGAATGTTCTTATAGCTCACGTTGCTTCTT
Splice_variant_3 : ------------------------------------------------------------------------------------------------------------------------

                        3'ss-3
Genomic          : GTAGATTGCAAAAGAAGACCAGCTTTTTCCAGAGAATATTTGAAGTTTGTTTGGCTCCTTTTCTCATGGTGGTCTTTCATGGCAAGAAGGTTGATGATGAAGCAAAGTGATTTCTTCAGC
Splice_variant_2 : GTAGATTGCAAAAGAAGACCAGCTTTTTCCAGAGAATATTTGAAGTTTGTTTGGCTCCTTTTCTCATGGTGGTCTTTCATGGCAAGAAGGTTGATGATGAAGCAAAGTGATTTCTTCAGC
Splice_variant_3 : ----ATTGCAAAAGAAGACCAGCTTTTTCCAGAGAATATTTGAAGTTTGTTTGGCTCCTTTTCTCATGGTGGTCTTTCATGGCAAGAAGGTTGATGATGAAGCAAAGTGATTTCTTCAGC
```

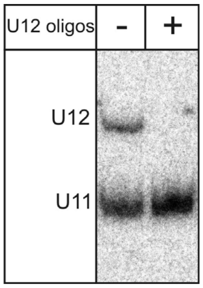


Figure S2. Cleavage of U12 snRNA by RNase H does not affect U11 binding to the USSE (related to Figure 2). Splicing reactions were incubated with or without oligonucleotides against U12 (U12-9C and U12$_{3\text{-}20}$), as indicated, prior to adding biotinylated short 48K RNA. The latter was pulled down with streptavidin beads and co-purifying RNA was identified by Northern blot analysis with probes against U11 and U12.

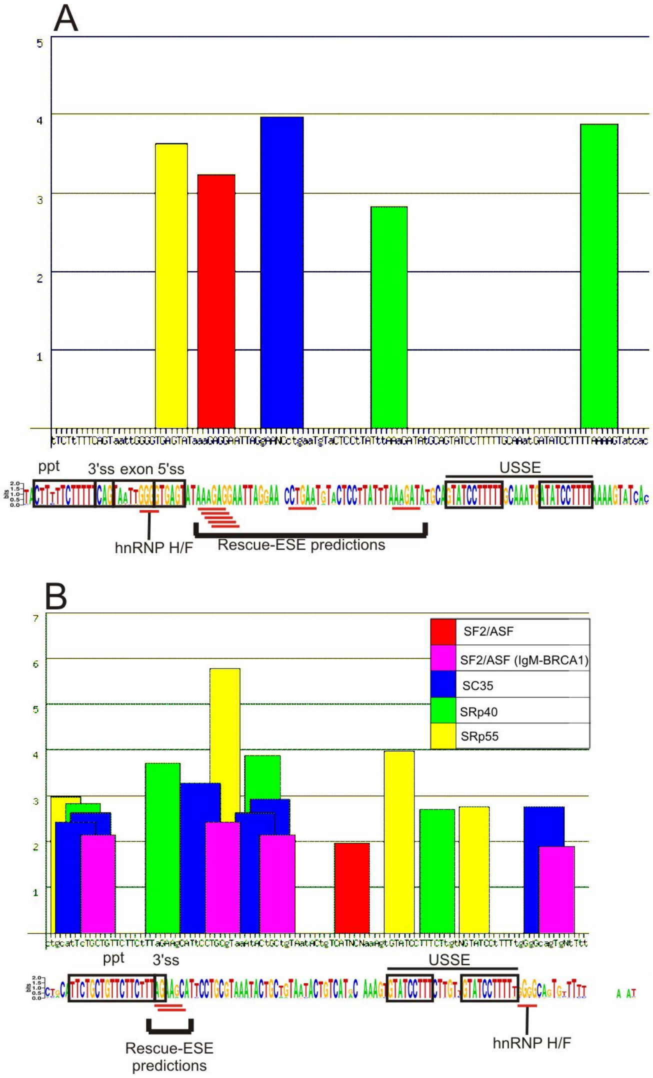


Figure S3. Identification of putative splicing enhancers and SR-protein binding sites in the conserved regions containing the USSE element (related to Figure 6). Mammalian consensus sequences were scanned with ESE-Finder 3.0 (Cartegni et al. 2003; Smith et al. 2006) and RESCUE-ESE (Fairbrother et al., 2002) using default parameters. The ESE-Finder output is shown as a bar chart above the sequence. The Y-axis depicts the ESE-finder score. RESCUE-ESE hits indicating putative ESE elements are underlined with red lines. Additionally, G-rich regions resembling hnRNP H/F binding sites (Schaub et al. 2007; Buratti et al. 2004) are indicated. (A) Putative SR and hnRNP H/F protein binding sites within the 48K RNA sequence. (B) Putative SR and hnRNP H/F protein binding sites within the 65K RNA sequence.

Table S1. Oligonucleotides used in this study

Name	Purpose	Target	See Fig.	Sequence (5'-3')[1]
a48K-11	RT-PCR	*Arabidopsis* 48K cDNA	S1	AGCGACATGATGAGTATGATTC
a48K-9	RT-PCR	*Arabidopsis* 48K cDNA	S1	CTGAACTGGTGTTCTGTTGTAA
h48K-10	RT-PCR	48K cDNA	4C, 5A,B,H	CAGGCACCAGGTGTGATCTA
h48K-11	RT-PCR	48K cDNA	4C, 5A,B,H	TGGCAGTTGACTTGCCAATA
h48K-113	Northern Probing	48K pre-mRNA	2E	CTGCATATCTTTAAATAAGGAGTACATTCAGGTTCCTAATTC
h48K-22	RT-PCR	48K cDNA	4C, 5A,B,F,H	CGGAGGATGAAGTTGTGA
h48K-70	RT-PCR	48K cDNA	4C, 5A,B,F,H	CCCAATTATTTGATTATAACAATCA
h48K-block	USSE block in cell culture	48K mRNA	4C	**aaggauaucauuugcaaaaaggaua**
h48K-mock	mock block in cell culture	48K mRNA	4C	**auucuaauuauaguaggcauuucau**
h65K-9	RT-PCR	65K cDNA	4B,D,E 5D,E,F,G,H	CTTAACCTGGATCAACAGGTG
h65K-11	RT-PCR	65K cDNA	4D, 5D,E,F,H	CCATGGTGGTTCAGTTTGCT
h65K-14	RT-PCR	65K cDNA	5H	ACGGGATTCAAAAAAAC/CTTTA
h65K-15	RT-PCR	65K cDNA	5G	TCCATAGTCCAGGTGGTCAATT
h65K-26	Cloning	65K gen. DNA		GAACACGAAGTAATTGGTCACAA
h65K-27	Cloning	65K gen. DNA		CCTGATTCACTAACTTTCCCTCA
h48K-57	RNase H cleavage	48K pre-mRNA	3A, 3B	TAAGGAGTACATTC
h65K-69	Northern Probing	65K pre-mRNA	2F	TTTGCATGACAGTATTACAGCAGTATTTACGCAGGAA
h65K-79	RT-PCR	pGL4.13-65KUTR cDNA	4F	GGGGGAAGGACAAACATTT
h65K-80	RT-PCR	pGL4.13-65KUTR cDNA	4F	GGCAAGATCGCCGTGTAATA
h65K-82	RT-PCR	pGL4.13-65KUTR cDNA	4F	CTTTCAAAGCTGTTACGCACA
h65K-M1	USSE block in cell culture	65K mRNA	4D	AAAGGATACGACAAGAAAGGATACA
h65K-M2	mock block in cell culture	65K mRNA	4D	TTCTCCCATCTCGGACTTGCAAAGT
hGAPDH-1	RT-PCR	GAPDH cDNA	4A,C, 5A, B	CACCAGGGCTGCTTTTAACT

[1] DNA uppercase, LNA lowercase, Morpholino uppercase underlined, 2'-*O*-methyl RNA lowercase bolded

Oligonucleotides used in this study (continued)

Name	Purpose	Target	See Fig.	Sequence (5'-3')[1]
hGAPDH-2	RT-PCR	GAPDH cDNA	4A,C, 5A, B	TGGAAGATGGTGATGGGATT
Luc-9	RT-PCR	pGL4.13-65KUTR cDNA	4B,E, 5G	CAGTCGTCGTGCTGGAAC
P120-13	RNase H cleavage	P120 splicing substrate (5'ss)	2C	GCAAGGATATCCTG[2]
p48K-5	RT-PCR	populus 48K cDNA	S1	ATCTAAGGTGAATATCCAAGTGCTA
p48K-7	RT-PCR	populus 48K cDNA	S1	GACATGCATGAGGATGATGT
$U1_{1-14}$	*in vitro* splicing block	U1 snRNA	3A, 3B	**ugccagguaaguau**[3]
U11-10	RNase H cleavage	U11 snRNA	3A, 3B	GTGTGCCACTCACGACAGAAG
U11-6L	Northern Probing	U11 snRNA	2D, 2F, 3	TCtCTtGAtGTcGAtTCcGCaC
U1-1L	Northern Probing	U1 snRNA	2D, 2F, 3	GcAGtCCcCCaCTaCCaCAaATtAT
$U1_{64-75}$	RNase H cleavage	U1 snRNA	3A, 3B	GTAACGTGAGGC
$U12_{3-20}$	RNase H cleavage	U12 snRNA	S2	TACTCATAAGTTTAAGGC
U12-9C	RNase H cleavage	U12 snRNA	S2	TCTCACATTAGCAGTGAGG
U12-9L	Northern Probing	U12 snRNA	2D, 2F, 3	AGaTCgCAaCTcCCaGGcATcCCgC
U2-3L	Northern Probing	U2 snRNA	2D, 2F, 3	TtTAaTAtATtGTcCTcGGaTAgAG
U2b	*in vitro* splicing block	U2 snRNA	2C	**auaaiaacaiauacuacacuuia**[4]

[1] DNA uppercase, LNA lowercase, Morpholino uppercase underlined, 2'-*O*-methyl RNA lowercase bolded, [2] see also Turunen et al. (2008), [3] see also Tarn and Steitz (1994), [4] i denotes an inosine residue, see also Lamond et al. (1989)

Table S2. PCR primers used for generating transcription templates

Name	**Type[a]**	**Template[a]**	**Sequence (5'-3')**
h48K-46	Fwd	Long 48K	GCGAAGCTTAATACGACTCACTATAGGGAATGCAGTTTGTTGCTTAACC
h48K-48	Rev	48K WT	AATTTTACTTTTAAAAGGATATC
h48K-49	Rev	48K 5' A3G	AATTTTACTTTTAAAAGGATATCATTTGCAAAAAGGACACTG
h48K-50	Rev	48K 3' A3G	AATTTTACTTTTAAAAGGACATCATTTG
h48K-51	Rev	48K 2×A3G	AATTTTACTTTTAAAAGGACATCATTTGCAAAAAGGACACTG
h48K-96	Fwd	Short 48K	GCGAAGCTTAATACGACTCACTATAGGAATTAGGAACCTGAATGTACTCC
h48K-101	Rev	48K 5' CC5/6GG	AATTTTACTTTTAAAAGGATATCATTTGCAAAAACCATACTG
h48K-102	Rev	48K 3' CC5/6GG	AATTTTACTTTTAAAACCATATCATTTG
h48K-103	Rev	48K 2×CC5/6GG	AATTTTACTTTTAAAACCATATCATTTGCAAAAACCATACTG
h65K-28	Fwd	65K	GCGAAGCTTAATACGACTCACTATAGGGTAGTCACAAGTCTTATAAACAC
h65K-30	Rev	65K WT	AAACAAAAAAACACTGCCCC
h65K-31	Rev	65K 2×A3G	AAACAAAAAAACACTGCCCCAAAAAGGACACGACAAGAAAGGACACAC
h65K-32	Rev	65K 2×CC5/6GG	AAACAAAAAAACACTGCCCCAAAAACCATACGACAAGAAACCATACAC

[a] Type is forward (Fwd) or reverse (Rev). Reverse primers contain the USSE mutations indicated in Template field

Table S3. Animal 48K and 65K genomic sequences

Latin name	Common name	Database	Genome release	48K gene	65K gene
Acyrthosiphon pisum	Pea aphid	www.aphidbase.com	Aphibase 1.0	GENSCAN_mRNA_SCAFFOLD3504_0006	maker-SCAFFOLD13652-gene-1-mRNA-1
Aedes aegypti	Aedes	www.ensembl.org	Ensembl release 54	AAEL002105	AAEL005083
Anolis carolinensis	Anole lizard	www.ensembl.org	Ensembl release 54	ENSACAG00000002811	ENSACAG00000002029
Apis mellifera	Honey bee	www.ncbi.nlm.nih.gov/mapview/	Amel_4.0	LOC551392	GB16643-PA
Bombyx morii	Silk moth	silkworm.genomics.org.cn/silkdb	SilkDB V2.0	BGIBMGA002859-TA	n/a
Bos taurus	Cow	www.ensembl.org	Ensembl release 54	ENSBTAG00000001123	ENSBTAG00000019091
Branchiostoma floridae	Florida lancelet	www.metazome.net	JGI V1.0	67055 [2]	estExt_fgenesh2_pm.C_1530006
Canis familiaris	Dog	www.ensembl.org	Ensembl release 54	ENSCAFG00000009574	ENSCAFG00000019972
Cavia porcellus	Guinea pig	www.ensembl.org	Ensembl release 54	ENSCPOG00000007910	ENSCPOG00000004437
Ciona intestinalis	Vase tunicate	crfb.univ-mrs.fr/aniseed/	V3.0	KH.S2264.1.v1.A.ND2-1	KH.C9.303.v1.A.SL1-1
Danio rerio	Zebrafish	www.ensembl.org	Ensembl release 54	ENSDARG00000039989	ENSDARG00000011247
Dasypus novemcinctus	Armadillo	www.ensembl.org	Ensembl release 54	ENSDNOG00000014978	n/a
Dipodomys ordii	Kangaroo rat	www.ensembl.org	Ensembl release 54	ENSDORG00000015140	ENSDORG00000000599 [1]
Drosophila melanogaster	Fruitfly	www.ensembl.org	Ensembl release 54	n/a	FBgn0050327
Echinops telfairi	Lesser hedgehog tenrec	www.ensembl.org	Ensembl release 54	ENSETEG00000011102	ENSETEG00000015567
Equus caballus	Horse	www.ensembl.org	Ensembl release 54	ENSECAG00000026937	ENSECAG00000021148
Erinaceus europaeus	Hedgehog	www.ensembl.org	Ensembl release 54	ENSEEUG00000015437	ENSEEUG00000013148 [1]
Felis catus	Cat	www.ensembl.org	Ensembl release 54	ENSFCAG00000018853	n/a
Gallus gallus	Chicken	www.ensembl.org	Ensembl release 54	ENSGALG00000013005	ENSGALG00000005162
Gasterosteus aculeatus	Stickleback	www.ensembl.org	Ensembl release 54	ENSGACG00000017756	ENSGACG00000002048
Homo sapiens	Human	www.ensembl.org, genome.ucsc.edu	Ensembl release 54, GRCh37	ENSG00000168566	Rnpc3
Lottia gigantea	Owl limpet	www.metazome.net	JGI V1.0	LgGsHFWreduced.11881	LgGsHFWreduced.8681
Loxodonta africana	Elephant	www.ensembl.org	Ensembl release 54	ENSLAFG00000011756	Chr1:103,893,839-103,897,533
Macaca mulatta	Macaque	www.ensembl.org	Ensembl release 54	ENSMMUG00000014785	ENSMMUG00000018039
Monodelphis domestica	Opossum	www.ensembl.org	Ensembl release 54	ENSMODG00000009996	ENSMODG00000002946

Animal 48K and 65K genomic sequences (continued)

Latin name	Common name	Database	Genome release	48K gene	65K gene
Mus musculus	Mouse	www.ensembl.org	Ensembl release 54	ENSMUSG00000021431	ENSMUSG00000027981
Nasonia vitripennis	Jewel wasp	www.ncbi.nlm.nih.gov/mapview/	Nvit_1.1	Scaffold17:1058726-1063520	LOC100122396
Ochotona princeps	Pika	www.ensembl.org	Ensembl release 54	ENSOPRG00000015471	ENSOPRG00000013972[1]
Ornithorhynchus anatinus	Platypus	www.ensembl.org	Ensembl release 54	ENSOANG00000019888[1]	ENSOANG00000008220
Oryzias latipes	Medaka	www.ensembl.org	Ensembl release 54	ENSORLG00000015654	ENSORLG00000005694
Otolemur garnettii	Bushbaby	www.ensembl.org	Ensembl release 54	ENSOGAG00000005275	GeneScaffold_5339:72,946-111,275
Pan troglodytes	Chimpanzee	www.ensembl.org	Ensembl release 54	ENSPTRG00000017703	ENSPTRG00000033902
Pongo pygmaeus	Orangutan	www.ensembl.org	Ensembl release 54	ENSPPYG00000016209	n/a
Procavia capensis	Hyrax	www.ensembl.org	Ensembl release 54	ENSPCAG00000011991	ENSPCAG00000000224[1]
Pteropus vampyrus	Megabat	www.ensembl.org	Ensembl release 54	ENSPVAG00000002854	ENSPVAG00000000245[1]
Pyretophorus gambiae	Anopheles	www.ensembl.org	Ensembl release 54	n/a	AGAP005296
Rattus norvegicus	Rat	www.ensembl.org	Ensembl release 54	ENSRNOG00000013756	ENSRNOG00000017310
Spermophilus tridecemlineatus	Squirrel	www.ensembl.org	Ensembl release 54	ENSSTOG00000014635	n/a
Strongylocentrotus purpuratus	Sea urchin	www.metazome.net	JGI Build 2.1	XM_001184482.1	XM_778095.2
Taeniopygia guttata	Zebra finch	www.ensembl.org	Ensembl release 54	ENSTGUG00000007575	ENSTGUG00000004795
Takifugu rubripes	Fugu	www.ensembl.org	Ensembl release 54	ENSTRUG00000007031	ENSTRUG00000001130
Tarsius syrichta	Tarsier	www.ensembl.org	Ensembl release 54	ENSTSYG00000000698	n/a
Tetraodon nigroviridis	Tetraodon	www.ensembl.org	Ensembl release 54	ENSTNIG00000012977	ENSTNIG00000018963
Tribolium castaneum	Red flour beetle	www.beetlebase.org	Beetlebase 3.0	TC001134	TC010372, TC010373[3]
Tupaia belangeri	Tree Shrew	www.ensembl.org	Ensembl release 54	ENSTBEG00000015620	ENSTBEG00000016291[1]
Tursiops truncatus	Dolphin	www.ensembl.org	Ensembl release 54	ENSTTRG00000016715	scaffold_99960: 5,827-34,058
Vicugna pacos	Alpaca	www.ensembl.org	Ensembl release 54	ENSVPAG00000003557	ENSVPAG00000011414[1]
Xenopus tropicalis	Pipid frog	www.ensembl.org	Ensembl release 54	ENSXETG00000020447	ENSXETG00000025027

n/a - Ortholog was not identified; [1] Partial sequence that did not contain the USSE region; [2] Weak homology to the 48K gene in other species; [3] 65K gene annotated to two consecutive transcription units

Table S4. Plant 48K and 65K genomic sequences

Latin name	Common name	Database	Genome release	48K		65K	
				Gene	Coordinates	Gene	Coordinates
Arabidopsis thaliana	Thale cress	www.gramene.org	TAIR release 8	At3g04160	Chr3: 1,091,154-1,094,353	AT1G09230.1	Chr1: 2,979,522-2,982,626
Arabidopsis lyrata	Lyre-leaved rock-cress	www.gramene.org	JGI Araly1	fgenesh2_kg.3__379__AT3G04160.1	scaffold_3: 1,383,202-1,386,296	fgenesh2_kg.1__966__AT1G09230.1	Scaffold_1: 3,528,455-3,532,039
Branchypodium distachyon	Purple false brome	www.brachybase.org	JGI 4X draft genome release	Super_3.3309	super_3:21035761-21039686	super_2.2460	super_2:15866108-15876108
Zea mays	Corn	www.maizesequence.org	Release 3b.50	GRMZM2G022107	Clone AC198229.4: 46,662-50,955	AC208339.3_FG027	Clone AC208339.3:78,274-82,606
Medicago trunculata	Barrel medic	www.medicago.org	IMGAG Version 2.0	n/a	chr1:17605724-17610091	2598.m00004	chr07: 17574850-17582132
Carica papaya	Papaya	www.ncbi.nlm.nih.gov	ENTREZ ID 20267	n/a	gi|187572611|gb|DS981571.1|, Supercontig_51: 589649-594000	n/a	gi|187570087|gb|DS982095.1| ,supercontig_594: 11505-22503
Physcomitrella patens	Moss	www.phytozome.net	Phypa1_1	160817	scaffold_26:1497775-1492558	1907212	Scaffold_116:422145-426259
Populus trichocarpa	Black cottonwood	www.gramene.org	JGI v1.1	gw1.180.17.1	scaffold:jgi2004:scaffold_180:213639:218344:1	estExt_Genewise1_v1.C_LG_XIII1925	Scaffold LG_XIII: 3,975,073-3,979,944
Oryza sativa ssp. japonica	Rice	www.gramene.org	MSU/TIGR pseudomolecule assembly release 5	LOC_Os04g28170.1, Q7XKS8_ORYSA	Chr4: 16,452,087-16,455,192	LOC_Os03g21020, Os03g0326600	Chr3: 11,903,301-11,908,550
Ricinus communis	Castor bean	castorbean.jcvi.org	Release_0.1	29950.m001185	29950:410000-416000	29950.m001170	29950: 293,859-298,357
Selaginella moellendorffi	Spikemoss	www.phytozome.net	JGI v1.0	406818	scaffold_6:436102-434218	1866650	Scaffold_80:623300-625097
Shorgum bicolor	Shorgum	www.gramene.org	JGI v1.0	Sb01g003350	chromosome:Sbi1:1:2683440:2689804:-1	Sb01g036600	Chr1: 60,231,230-60,236,025
Vitis vinifera	Grape	www.gramene.org	Genoscope-8x-2007	GSVIVG00026753001	chromosome:8X:4:10481891:10491985:1	n/a	n/a

Table S5. Plant 48K ESTs

Latin name	Common name	Accession number	Type	Database
Helianthus tuberosus	Artichoke	EL450721	EST	plantta.jcvi.org
Centaurea solstitialis	Yellow starthistle	EH789393	EST	plantta.jcvi.org
Coffea canephora	Coffee robusta	DV679512	EST	plantta.jcvi.org
Gossypium hirsutum	Cotton	DV849442	EST	plantta.jcvi.org
Euphorbia esula	Leafy spurge	TA13745_3993	EST	plantta.jcvi.org
Euphorbia tirucalli	Indiantree spurge	BP954788	EST	plantta.jcvi.org
Lactuca saligna	Willow leaf lettuce	DW073610	EST	plantta.jcvi.org
Solanum tuberosum	Potato	CK257796	EST	plantta.jcvi.org
Triticum aestivum	Wheat	TA108318_4565	EST	plantta.jcvi.org
Lotus japonicus	Lotus	LjT26B16	Partial genomic clone	www.kazusa.or.jp/lotus
Picea glauca	White spruce	PUT-162b-Picea_allspecies-60833	EST	www.plantgdb.org

Supplemental References

Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala, Y.M., and Baralle, F.E. (2004). hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHβ genes. Nucleic Acids Res *32*, 4224-4236.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. (2003). ESEfinder: a web resource to identify exonic splicing enhancers. Nucleic Acids Res *31*, 3568-3571.

Edgar, R. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Fairbrother, W.G., Yeh, R.-F., Sharp, P.A., and Burge, C.B. (2002). Predictive identification of exonic splicing enhancers in human genes. Science *297*, 1007-1013.

Hubbard et al. (2007). Ensembl 2007. Nucleic Acids Res *35*: Database issue:D610-D617.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM and Haussler D. (2002). The human genome browser at UCSC. Genome Res *12,* 996-1006.

Lamond, A.I., Sproat, B.S., Ryder, U., and Hamm, J. (1989). Probing the structure and function of U2 snRNP with antisense oligonucleotides made of 2'-OMe RNA. Cell *58*, 383-390.

Ovcharenko, I., Loots, G.G., Giardine, B.M., Hou, M., Ma, J., Hardison, R.C., Stubbs, L., and Miller, W. (2005). Mulan: multiple-sequence local alignment and visualization for studying function and evolution. Genome Res *15*, 184-194.

Rice, P. Longden, I. and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet *16*, 276-277.

Schaub, M.C., Lopez, S.R., and Caputi, M. (2007). Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. J Biol Chem *282*, 13617-13626.

Smith, P.J., Zhang, C., Wang, J., Chew, S.L., Zhang, M.Q., and Krainer, A.R. (2006). An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Hum Mol Genet *15*, 2490-2508.