Doc S4

XLOC_006026 genomic fragment

Ch1: 13222626 – AACGCTCTAATAGCTCACCACATATACATATGCAAAATTTACCACCCCAAATTTGCTCAGCTCTAAATGATCCACTATAAATAACCATCTCATATCGAACAATTACCACAACCACAATTTGCCTCAGTCTAAATGATCTCTCAGAATTTT

In black – exons, Red – introns. Orange – part of the coding sequence of the Q9LQH8 protein that is included in the extended first intron (Orange and red sequence) spliced to give the sequence of TCONS_00009618.

GT and ag – the alternative donor and acceptor splice sites of the extended intron removed in TCONS_00009618.

GT and AG – indicate the correct donor and acceptor sites of intron 1 and 2.

Pairwise alignment of XLOC_006026 and TCONS_00009618:
Position 107 to 508 with stop at 422-424

**F15O4.6/Q9LQH8 protein sequence from Uniprot:**

>tr|Q9LQH8|Q9LQH8_ARATH F15O4.6 OS=Arabidopsis thaliana OX=3702 PE=4 SV=1
MISQNFYIAR YFLCIALEMFCYTELYEQAIEYEKGSLITSNGALTMLSGAKTDRE ISVLEMRLLRMSFGRIDYFIFILYFGLLIFVFDFFDFGBKLSNIEIDEHTFMVNREAVD YLNSLEKILSHSHGCGMDGDPALFFFELLGIATF

**In red** – the polypeptide part of the protein coded by 3’ half of exon 1 (see above in orange) that is excluded from TCONS_00009618