Eurosta solidaginis

Taxonomic lineage: Brachycera > Muscomorpha > Tephritoidea > Tephritidae > Eurosta

Assembly: GCA 040869045.1 ASM4086904v1 genomic

Eur_sol1 | CM081795.1:258520302-258521245 (-) | 944 nt | IncRNA:noe consensus e-value: 3.4e-16

5' motif: GCGGT 3' motif: ATCGC Internal max. Poly-T: 3nt Trailing-T: 7nt PSE: 0.96

Eur_sol2 | CM081794.1:257293876-257294122 (-) | 247 nt | IncRNA:noe consensus e-value: NA

5' motif: GCGGT 3' motif: ATCGC Internal max. Poly-T: 4nt Trailing-T: 5nt PSE: 0.96

Eur_sol3 | CM081794.1:257324205-257324451 (+) | 247 nt | IncRNA:noe consensus e-value: 4.5e-10

5' motif: GCGGT 3' motif: ATCGC Internal max. Poly-T: 4nt Trailing-T: 7nt PSE: 0.85

Eur sol4 | CM081794.1:257275967-257276196 (+) | 230 nt | IncRNA:noe consensus e-value: 4.4e-06

5' motif: GCGGT 3' motif: ATCGC Internal max. Poly-T: 4nt Trailing-T: 5nt PSE: 0.96

Eur_sol5 | CM081794.1:257294665-257294883 (-) | 219 nt | IncRNA:noe consensus e-value: 1.4e-05

5' motif: GCGGT 3' motif: ATCGC Internal max. Poly-T: 3nt Trailing-T: 6nt PSE: 0.91

GECGGTCTGTCACCTCTCAGTTGACAAGGTTGACCACATCCCTTAATTGTGGTAATCTAAATTGCTTCTAATGAAATCTAAAGTGTGACTCAAATGCTAAAGTTGTGTCTTATAGTGGGAAATGATATTTCTCCAAAAAGACGGATAAACAAAAGACGGATAAACAAAAGTCTAAATTGTTGATGATAGATGATAGATGCTCTGGGGAGTGTAGGTGATCACTTTTTT

Eur_sol6 | CM081794.1:257307612-257307829 (-) | 218 nt | IncRNA:noe consensus e-value: 1.4e-04

Eur_sol7 | CM081794.1:257278816-257278981 (+) | 166 nt | IncRNA:noe consensus e-value: 1.5e-03

5' motif: GCGGT 3' motif: ATCGC Internal max. Poly-T: 3nt Trailing-T: 6nt PSE: 0.91

GGCGGTCTGTCACCTCCCAGTTGAAAAGGTTGACCACATCCCTTAATTGTGGTAATCTAAACTGTTCTAATGAAATCTAAAGTGTGGCTCAAATGCTAAAGTTGTGTCTTATAGTGGGAAATGATATTTCTCCAAAAAGACGGATAAACA AAATGGATCGAAAGGTCTAATTGTTGATGATGATGATGATGTCTCTGGGGAGTGTAGGTEATCGCTTTTTT

5' motif: GCGGT 3' motif: ATCGT Internal max. Poly-T: 3nt Trailing-T: 5nt PSE: 0.85

GGCGGTGTGACACCTCCCAGTTGAAAAGGTTGGCCACATCCGTTAATTGTGGCAAAACAAGTCCGAAAATTACAGAAGGGTTGGAGAAACTGATCGAAATTTTGAGCAATTATAAAACAGAGATTAAAATTCTGCTGGGGAGGTACACGT CATCGTCCTTATTTTT

Eur_sol8 | CM081794.1:257285861-257286026 (-) | 166 nt | IncRNA:noe consensus e-value: 4.2e-04

5' motif: GCGGT 3' motif: ATCGT Internal max. Poly-T: 3nt Trailing-T: 5nt PSE: 0.85

ARCARTECRACACCTCCCAGTTGAAAAGGTTGGCCACATCCCTTAATTGTTGCAAAACAAGTCTAAAATTACAGAAGGGTTGGAGAAACTGATTGAAATTTGAGCAATTATATAAAACAGAGATGAAAATTCTGCTGGGGAGGTACACGTCCTTATTTTT

	Eur_sol1	Eur_sol2	Eur_sol3	Eur_sol4	Eur_sol5	Eur_sol6	Eur_sol7	Eur_sol8
Eur_sol1	-	18	21	14	14	14	11	11
Eur_sol2		-	60	86	55	57	49	47
Eur_sol3			-	63	54	53	49	46
Eur_sol4				-	61	61	52	50
Eur_sol5					-	98	54	55
Eur_sol6						-	56	55
Eur_sol7							-	96
Eur_sol8								-

Pairwise sequence identity matrix (%) showing similarity between IncRNA sequences.

Eur_sol1	taactaacggaaaattttattgctattgcagacgatgcAATTCGCAGATGCTTTTTCggtaaaatgcagaataaataggacttccaagttaactgcatgc
Eur sol2	tatatgaaatttcgctgatgatataaatttggctAcTtcGcAcATGcTTTTTCGtcaaataatttgtataaataccgttttcaatgcattaatatagcat
Eur_sol3	agctcaaccaatttcattgatgttagaatattcttttA <mark>GTTCACGGGTGCTTTTTC</mark> gttaaatgaattgcataaataccgttttcgatacccactagcaa
Eur_sol4	catacqaaatttcqatqatqatataaatttqtqtAATTCGCAGATGCTTTTTCqttaaataatttqtataaataccqttttcaatacqtcqaqccatcat
Eur_sol5	ttt <mark>o</mark> ctt <mark>o</mark> ctccttatta <mark>u</mark> caataucaattutututa <mark>GCAGATTCCCTTTTTC</mark> uttuaacattttutatataaatautaattuoaatacacaaattatt
Eur_sol6	tttgcttgactccttgttagcattagcaaatttgtgtAGTTCGCAGATGCTTTTTCgttgaacattttgtataaatagtaattaggaatatgaaatatt
Eur_sol7	tt <mark>gacaac</mark> atttttcgtagatgttataaaattttttc <mark>AATTCGCTGGTGCTTTTTC</mark> attgaaggttatgtataacaaattgtaaaactgtaaattgc
Eur_sol8	tt <mark>gacaacatttttcu</mark> ta <mark>uatauttataaaatttttcucAATTCGCTGGTGCTTTTTC</mark> attaaa <mark>u</mark> attataattaaataacaaatt <mark>uaaaatca</mark> tattttut
RNase_MRP	attcaccct <mark>gcaaacagaacaaAATTCACAGATGCTTTTTC</mark> gttacatacaaagaataaatactgtttagctgaccatttatataaggtcggcgctaatt
U6atac	<mark>cag</mark> atgcagttatttcattcgattcgattcgattggtgtat ACTTCTCAGATGCTTTCTC ggtacaccttacttataaatacatatggaaaagagttttaattt
U1	tttt <mark>cattocaatttaaaaaatoctacatttoco</mark> tcotat AATTCTTAACTTGCTAATG aaaaaaottatacaotaoatttootactoaaaaaaoc
U1	tttcgttcgactttaaactcacaagcatcctacgaataca AATTCTTAACAACCTAATA acaaattccacttaagtaatttcatttgcgctaagaaaatc
U1	tgttc <mark>gttgcttatg</mark> aaaaagtgct <mark>gcaattgagtcctctAATTCTTAACTTCTTATG</mark> aaaaaagttatacagtagatcttttataagctgatgaaa <mark>g</mark> c
U1 U1	ttto <mark>octcatcauou</mark> natacatattuatecttoaactaacAATTOTTAAACACTITATEctttecatutetoaatacttuatttuattuatutaacaacttoattuatetoataetuatetetetoataetuatetetetoaataetaetuatetetoaatuataetaetaetaetaetaetaetaetaetaetaetaetae
U1	
U1	tttettaateateatusetusetusetusetusetusetusetusetusetuse
U1	
U1	atttcattcatcacacagattgaaatgaactgcctatAATTCTCATTTATTTTATGcgtgagtccgttgaaatagttagttggtaggtagaaagc
U1	ctotograattititcactogacattatticcaccatttATACTCAACGACTTTACTaaaactocatotatattattottataocoggagaaagc
U1	tgtacaattaaatttcactcgttactattaaccgctgttaAATTCTTAGATACTTAAATaatattggtattgaaatagttcattgtgagctgactaaagc
Ü1	atttcatttaagcatttaaagtttgcatttgcgtactcgtTATTCTCAACTATTTTATAgaaattgttattgaagtgattgcttttttactaaataaagc
U1	ctotogtaattitttcacttgacattatttccaaccotttAATACTCAACGACTTTACTTtaaactocatotcaaatauttttattotaocoggaagaagac
Ú2	<mark>cacc</mark> ataaaatttcttttaaatatttgcttttgtgtgtatgc <mark>AATTCTCAGTTAAAC</mark> aacttttgtgttgaaatatttaaaaattttctggcttaattt
U2	ttcgttgaagtcggattttgaagagtgatgtctcgtttgAATTCTCAAACATTTTATAactttttggaattaaaatagaccattttattttcgtttaacttt
U2	ttc <mark>octoaqutcoo</mark> atttt <mark>oaquacto</mark> atotctcattt <mark>oACTTCTCAAACATTTTATA</mark> actttt <mark>oo</mark> aatttaaaata <mark>oacc</mark> ottttattctattoaaottt
U2	ttttcgttgaagtcggatttggatgtgaggtgctatttcAATTCTTAAACACTTTATAactttttggaatttaaatacactgttttgtgccgtttaaagtc
U2	ttcat <mark>ccacctctcattttcttgcagtgttccatcg</mark> taa AATTCTCATTTATTTAATA cacaattccattgtaatatttataatttttctggttgtagtt
U2	ttct <mark>ucactautttctttatucttautccttautccttautccAATTCTCAACAATTTCATA</mark> tutccttcautaaaataaaattttaututttuaa <mark>u</mark> tt
U3	<mark>tacgaaactttgtatattgatcaatgatcttcacttcttAATTCTCAACTTACT</mark> tacagcatagtttaagtactttgtgccgattttatcaatgtt
U4	<mark>ggcaactctggcgctcgtttcgtgcacagtcatctcgcatAATTCTCAAACGTTTTATA</mark> aatatttggaatttaataaaatttcttgg <mark>cataccaaaatc</mark>
U4	tttcattgctacacacttcaatttttgttttgcattcaA <mark>GTTCTCAAACATTTTATA</mark> cacaaacgtatttaagtagttaatattgtgggtgatgttattt
U4atac U5	ttcacacaaataatattttcttaaccttaacatcacca ttttaaaataatttaaaca acacaataataaacactcatttactttaacaatca
U5 U5	acatttcacttattgccttaatttgtgtttaattttagatcatACTTCTCAAACTETTCATAaactotcgtaataaatactttagtaaaactaaaaaaaaa
U5 U5	
U5 U5	
U11	
U12	
OrCD1	Ctgladataattaataataacagtticatggitt AATTCCAAACAATTAATA caagctatagttaaattattittucudattatti.
	5-3

Upstream promoter alignment showing conserved regions. Sequences include Pol III promoters (noeCR34335, Arthropod_7SK, 7SL, RNase_MRP, RNaseP_nuc, U6, U6atac, tRNAsec, snoRNAMe18SA1806) and Pol II promoters (U1, U2, U3, U4, U4atac, U5, U7, U8, U11, U12, OrCD1). Conserved positions are highlighted in the alignment.