

Figure S6: 5'ETS variation in additional species of Brassicaceae Lineage I, Clade A.

(A) 5'ETS genotyping in Cv, *Chrysochamela velutina*, Dm, *Descurainia millefolia*, Cp, *Cardamine parviflora*, Bs, *Boechera stricta*, Lc, *Lepidium campestre* compared to At, *Arabidopsis thaliana*, Rs, *Raphanus sativus* and Br-r, *Brassica rapa ssp rapa*. * uncloned amplified PCR products. Sizes in base pairs are labeled alongside the lane of the 1kb⁺ DNA ladder. **(B)** Schematic representation of variants containing an insertion splitting away the conserved U3 snoRNP binding site A¹²³B from the primary processing site P. **(C)** Schematic representation of insertion free variants. Dotted lines indicate deletions and a number of base pairs between brackets indicates the size of an inserted sequence without any homology with known D or C repeats. **(D)** Multiple Sequence Comparison by Log- Expectation (MUSCLE) alignments of 5'ETS insertion free variants. *Brassica oleracea* (Bo), *Arabidopsis cebennensis* (Ac2) and *Lepidium campestre* (Lc2). **(E)** Fasta sequences of 5'ETS variants.

Brassicaceae Lineage I, Clade A, Camelinae

Cv1 and Cv2 are supported by two independent clones whereas Cv3 sequence was only cloned once for *Chrysochamela velutina*.

Brassicaceae Lineage I, Clade A, Boechereae

Bs is the 100% identity sequence supported by two independent clones for *Boechera stricta*.

Brassicaceae Lineage I, Clade A, Cardamineae

Cp is the common sequence shared by two independent clones for *Cardamine parviflora*.

Brassicaceae Lineage I, Clade A, Descurainieae

Dm is the 100% identity sequence shared by three independent clones for *Descurainia millefolia*.

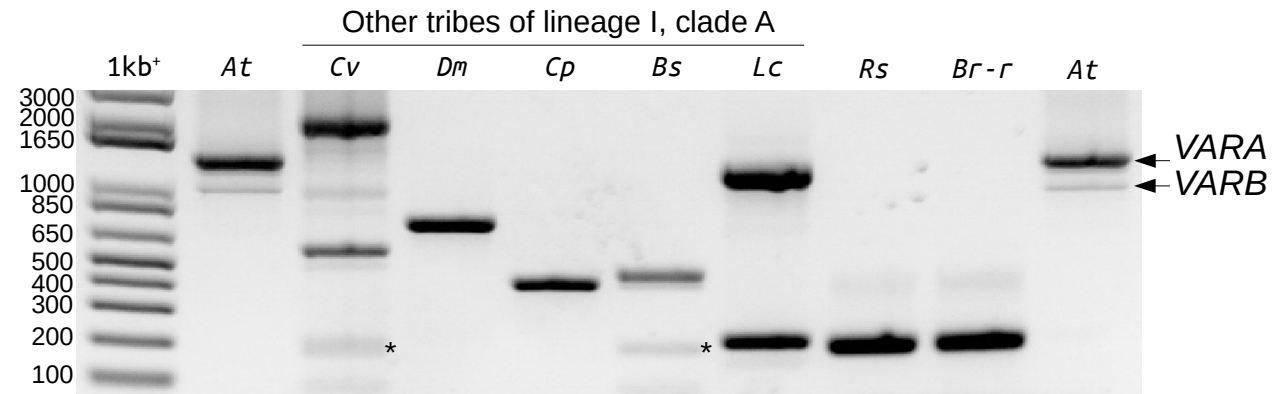
Brassicaceae Lineage I, Clade A, Lepidieae

Lc1 and Lc2 are both only supported by one clone for *Lepidium campestre*.

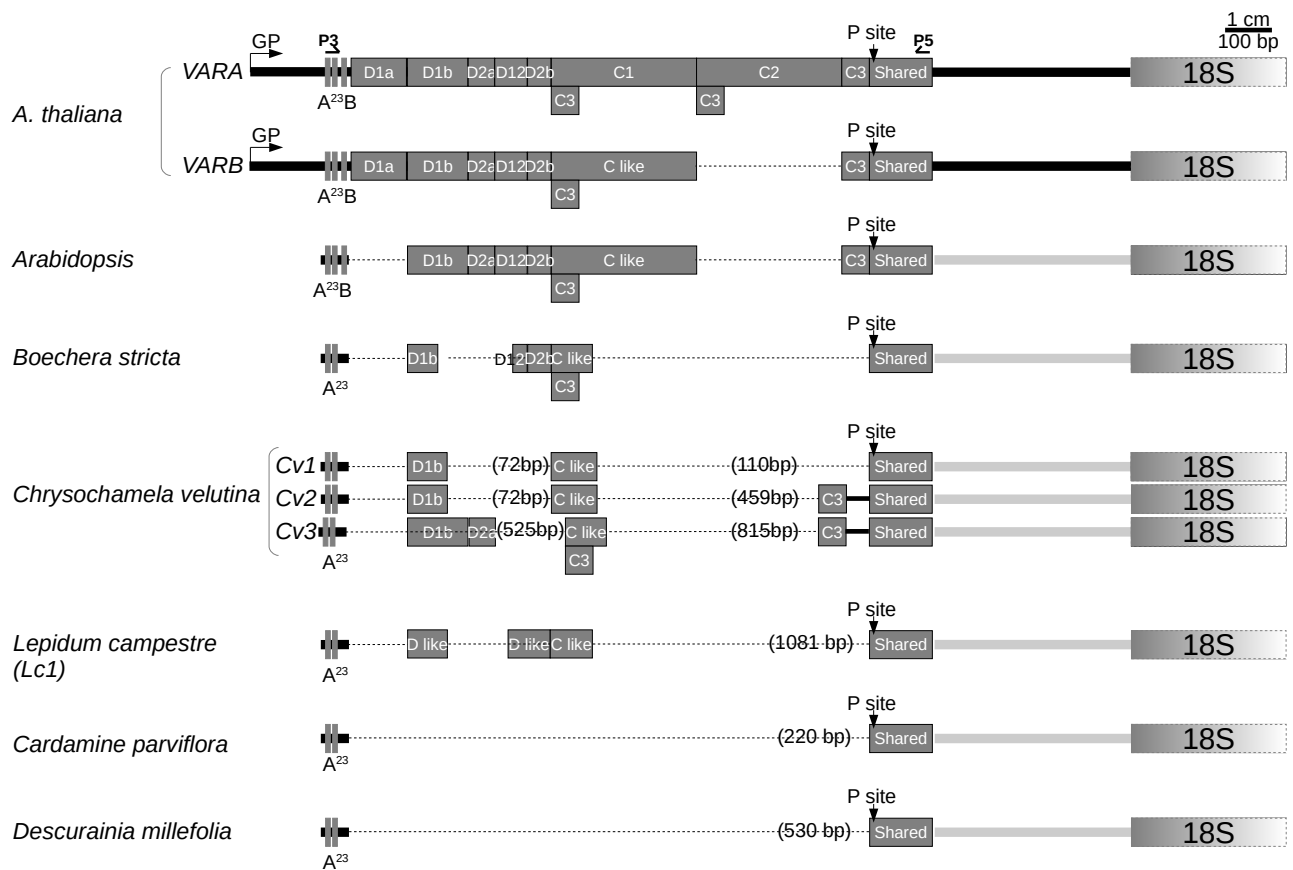
Supplementary result description

Longest variants were found in *Chrysochamela velutina* and *Lepidium campestre*, intermediate variants were identified in *Chrysochamela velutina*, *Descurainia millefolia*, *Cardamine parviflora*, *Boechera stricta*, insertion free variant was abundant in *Lepidium campestre* and only faintly detectable in *Boechera stricta* and *Chrysochamela velutina*. In addition to the deletion of the D1a repeat identified in the *Arabidopsis* genus, the major variant identified in *Boechera stricta* has a truncated version of D1b, complete loss of D2a, partial loss of D12 but still D2b and a truncated C-like repeat. When present in Brassicaceae species, the insertion either retains partially some homologies with some D and C repeats (*Boechera stricta*, *Lepidium campestre*, *Chrysochamela velutina*) and/or interrupt them by insertions (*Chrysochamela velutina*) or share no homology with any of them (*Cardamine parviflora*, *Descurainia millefolia*). Illustrating some of these statements for *Chrysochamela velutina*, we identified two major variants of 1912 bp (Cv1) and 552 bp (Cv3) while genotyping and an additional rare variant of 946 bp (Cv2) after cloning and sequencing. Surprisingly, all Cv variants have an insertion separating D repeats from the C like repeat, the length of this insertion vary from 72 bp in Cv1-2 to 525 bp in Cv3. Cv1-2 variants have a truncated D1b repeat while Cv3 retained both D1b and D2a. Contrary to Cv1, Cv2-3 retained the additional third occurrence of the C3 repeat but an insertion of 51 bp separate it from the P site. There is an additional insertion separating the C-like repeat from either the primary processing site (Cv1, 110 bp) or the last C3 repeat occurrence (Cv2, 459 bp ; Cv3, 815 bp).

(A) 5'ETS genotyping in additional species of Brassicaceae Lineage I, Clade A



(B) Schedules of variants containing an insertion splitting away the conserved U3 snoRNP binding site A123B from the primary processing site P



(C) Schedules of insertion free variants



(D) MUSCLE alignment of 5'ETS insertion free variants

```

      A1      A2      A3      B
Ac2      CTTTTCCGGGCAC-----TTTTCCGGTGGACATTTTGCCCC
Lc2      CTTTTCCGGGCACTTTTCCGGCGACTTTCCGGCGAGTTTTCCGGTGGTTCATTTTCTTCC
Bo       CTTTTCCGGGCACTTTTCCGGCGAC-----TTTTCCGGTGGACATTTTGCCCC
          *****
          P-site & P motif
Ac2      TAAAATCAAAATTTTCGCGCTTGTGTGGTCTTGGCCTGGTTTCATCCGTCTTCCAGCTGCT
Lc2      C---TTTATTTTTCGCGCTTGTGCGGCTTTGGACTGGTTTCATTTGTCTTCTTATGCA
Bo       TAAAATCAAAATTTTCGCGCTTGTGTGGTCTTGGCCTGGTTTCATCCGTCTTCCAGCTGCT
          * * ***** ** *** ***** ***** ***
          CTTTTGATTACATCTCGAGAGTGGTTGGAAAGATTGATGTTAGCGGGGCAA-TGAACGTT
          TTTTCGATTATATCGAAAGAGTGGTTGGAAAGATTGATGTTAGCGGGCCAAC-TGAACGTT
          CTTTTGATTACATCTCGAGAGTGGTTGGAAAGATTGATGTTAGCGGGGCAATTGAACGTTG
          *** ***** ** ***** ***** *****
          CGGCGTATGAGTGGTGATCGG
          CGGCGTATGAGTGGTGATCGG
          CGGCGTATGAGTGGTGATCGG
          *****
```

(E) Fasta sequences of 5'ETS variants

>Cv1

```
CTTTTCCGGGCACTTTTCCGGCGGTCTGGTTTTTTTGCCTTTTCCGATTTTTTCCGGATTAATCGACGAAAGA
TGGTGAATGGCCGAGGAGAATGGGCGTGTCATGCATGGGCTGACATGGATTCTCCGAGGCCGAGGGTG
GAGGTATGTTCCGATGGGCCCCACGACATTACCTGATGTGCCATGCGCATCGTTACACCCCAAGCCGAA
CTGGATGGGCGTGCTCTGTTGGGGAGCATGGATCCAGTTTTTGGCCGTCCCGGGGGGATCAAGGTGGCG
ATGTCGCCCATGGGCAACGATCACCTTGTGGAAGGGTTCGGTAACGGAAATGGTTGCGGTAATGAAAAA
AATGGAAGCTGGCTCCCGTGTTTCGACAATCATCGATGTTTCGTCAATGGACATCGAACATTCTTAGGATTTG
TTTCTCGCGCTTGTGCGGCTTTGGCTTGGTCCGTCCGTCTTCACTTTGCTCGTCCGAATATTTCGGAAGG
GTGGGGGAATGATTGGTGGTTTTGGGTCAATTGAACGTTCCGGCGTATGAGTGGTGATCGG
```

>Cv2

```
CTTTTCCGGGCACTTTTCCGGCGGTCTGGTTTTTTTGCCTTTTCCGATTTTTTCCGGATTAATCGACGAAAGA
TGGTGAATGGCCGAGGAGAATGGGCGTGTCATGCATGGGCTCACATGGATTCTCCGAGGCCGAGGGTGG
AGGTATGTTCCGATGGGCCCCACGACATTACCTGATGTGCCATGCGCATCGTTACACCCCAAGCCGAAC
TGGATGGGCGTGCTCTGTTGGGGAGCATGGATCCAGTTTTTGGCCGTCCCGTGGGGATCAAGGTGGCGA
TGTCGCCCATGGGCAACGATCACCTTGTGGAAGGGTTCGGTAACGGGAATGGTTGCGGTAATGAAAAAA
ATGGAAGCTGGCTCCCGTGTTTCGACAATCATCGATGTTTCGTCAATGTAATCGATCCGGTCTGTTAGCCGATG
TTGGAGTAATTGCCGTAGCCTTCGACGTTTCTGTTCTGGGAACCATCGTTTGATGGGAGTCCGGGTTTTG
CAAAACCGGCCAAGGGTAATGAACGATTGTGAACCTTGTGCGCTAGTCGGATGGGCGTGCTTAGGTAG
CATGGATCCGGTTAAGCTTGCTGGCCCGTGAGGGGTCTCTTCAATCGGACTTCAGGCGTGCCGTAAGT
CGCACACAAGCATGTTGGACGTTGATTCGGAGATACCTACCGCAGCTCTCATAGATATTGCAGTCTATCGCC
TAGGTCGGATGGGCGTGCTTCGATGGAAGCATGGATCCGCCTAGGCTGTCCCGACGGTTACACAAGAT
GACGGTGTCTCCCATGGACATCGAACATTCTTAGGATTTGTTTCTCGCGCTTGTGCGGCTTTGGCTTGGT
CCGTCCGTCTTCACTTTGCTCGTCCGAATATTTCGGAAGGGTGGGGGAATGATTGGTGGTTTTGGGTCAA
TTGAACGTTCCGGCGTATGAGTGGTGATCGG
```

>Cv3

```
CTTTTCCGGGCACTTTTCCGGCGGTCTGGTTTTTTTGCCTTTTCCGATTTTTTCCGGATTAATCGACGAAAGA
TGGTGAATGGCCGAGGAGAATGGGCGTGTCATGCATGGGCTGACATGGATTCTCCGAGGCCGAGGGTG
GAGGTATGTTCCGATGGGCCCCACGACATTACCTGATGTGCCATGCGCATCGTTACACCCCAAGCCTAAC
TGGATGGGCGTGCTCTGTTGGGGAGCATGGATCCAGTTTTTGGCCGTCCCGGGGGGATCAAGGTGGCGA
TGTCGCCCATGGGCAACGATCACCTTGTGGAAGGGTTCGGTAACGAAATGGTTGCGGTAATGAAAGAAAT
GGAAGCTGGCTCCCGTGTTTCGACAATCATCGATGTTTCGTCAATGTAATCGATCCGGTCTGTTAGCCGATGTT
GGAGTAATTGCCGTAGCCTTCGTCTGTTGCTGAACCTGGGAACCATCGTTTGATGGGAGTCCGGACTTTCC
AAAGCCGGCCAAGGGTAATGAACGAATTGTGAACCTTGTTCATCGCCAAGTCGGATGGGCGTGCTTTGT
TGGAGAGCATGGATCCGGTTAGGCTGTCCGATGATTACTTAGTTGCGCGTGGTATTGCATGTCGAAGTGGC
TGGCCCGCGAGGTGTCTTCTTCAATCTCCCTTCAGACGTTCTGTTTCACGTACGTTTTGTTGGGAGTTGAT
CCGAGGATAACTACCGCAGCTCTCCTAGATATTGCACTCTATCGCCTAGGTTGGATGGGCGTGCTTTGTTG
GGAAGCATGGATCCGCCTAGGCCGTCCCGGGGGAAGCAAGGTGGCGATGTCGCCCATGGGCAAGATC
```

ACCTTGTGGCTAGGTTCAAGTAACGGGAAAGGTTGCGGTAATGAAGAAAATGGAAGCTGGCTACCGTGTTG
GACAATCATCGATGTTTCGTCAATGTAATCGATCCGGTCGTTAGCCGATGTTGGAGAAAATTGCCGTAGCCTT
CGTCGTTTCTGTTCCCTGGGAACCATCGTTTGATGGGAGTCCGGACTTTCCAAAGCCGGCCAAGGGTAATG
AACGAATTGTGAACCCCTGTTTATCGCCAAGTCGGATGGGCGTGCTTTATTGGAGAGCATGGATCCGGTCA
GGCTGTCCGATGGTTACTTAGTTGCGCGTGGGGATTTTCATATCGCCTAGGTTGGATGGGCGTGCTTTGTT
GGGAAGCATGGATCCGCCTAGGCCGTCCCGGGGGAAGCAAGGTGGCGATGTCGCCCATGGGCAACGAT
CACCTTGTGGCTAGGTTCAAGTAACGGAAAGGGTTGCGGTAATGAAGAAAATTGAAGCTTGCTCCCGTGTT
CGACAATCATCGATGTTTCGTCAATGTAATCGATCCGGTCGTTAGCCGATGTTGGAGTAATTGCCGTAGCCT
TCGACGTTTCTGTTCCCTGGGAACCATCGTTTGATGGGAGTCCGGGTTTTGCAAAACCGGCCAAGGGTAAT
GAACGTATTGTGAACCCCTGTCGCCTAGTCGGATGGGCGTGCTTAGGTAGCATGGATCCGGTTAAGCTTGC
TGGCCCGTGAGGGGTCTCTCTTCAATCGGACTTCAGGCGTGCCGTACTGCGCACACAAGCATGTTGGACG
TTGATTCGGAGATACCTACCGCAGCTCTCATAGATATTGCAGTCTATCGCCTAGGTGCGGATGGGCGTGCTTC
GATGGAAAAGCATGGATCCGCCTAGGCTGTCCCGACGGTTACACAGGATGACGGTGTCTCCCATGGACAT
CGGACATTCTTAGGATTTGTTTTCTCGCGCTTGTGCGGCTTTGGCTTGGTTCCGTCCGTCTTCACTTTACTCG
TCCGAATATTTTCGGAAGGGTGGGGGGAATGATTGGTGGTTTTGGGTCAATTGAACGTTCCGGCGTATGAGT
GGTGATCGG

>Bs

CTTTTCCGGGCACTTTTCCGGCGGTCGGTTTTGTTGCCTTTTCTGGATTTTTTCGGCATTCTCGACGAAAG
ATGGGGAGTGGCCGAGGATGCATGTGCGGACATGGATTCTTCGAGGCTTTGGGTGGCGGTACGTTCCGA
GGGCAAACTAAATACCGAGATGTCCCATGGTCAACGTTTCCACCCGTGCATAGGTCGGATGGGCGTGCT
TCGATGGAAAGCACGGATCCGCCTAGGTTGCCCGACGGTTAAACAAGATGACGGTGTCTCCCATGGGCA
TCGGTCAAATCTTTTGGCTAGTTTTTCGCGCTTGTGCGGCTTTGGCTTGGATTCTGTCGGTCTTCTCTCAGCT
CAACCAAATACTTTGGAAGAGCTGTGGGAACGATTGAAGTTTTTGGGTCAACTGAACGTTCCGGCGTAGAG
TGGTGATCGG

>Cp

CTTTTCCGGGCACTTTTCCGGCGAGCGTTTTGATGACTTTCTCTGATTTTTTCAGCATGTCTCCACGAAAG
ATGGTGATATAGCGGCCTAGTCTGATAGCGAGCTTGTAGATTCCAAAGGTGCATGTCCGAGTGAAATGATG
GAGCTGGCCCGGGAGTGGGGTAGAGGAGCCTTAACGGGTAAACCACTACTTTTCGTCCCCCGCGGCGAA
GTTGTTTCCGTGACATGCATTGGAGGAAATTGTTTGAATCGATTTGGGCTAGTGTTTCGCGCTTGTGCGG
CTTTGGCTTGGATTGCGCTGTCTTCTCTCGCTCTTTTGATTTCCTCGGAAGGGTTGTGCGAAAGATTGTCTG
CTTCGGGTCAACTGAACGTTCCGGCGTATGAGTGGTGATCGG

>Dm

CTTTTCCGGGCACTTTTCCGGCGGACGGTTTTAGAGGATTTTTGGGTTTTTCGACCGTGCTGATGGGTTTTG
AAAATGAAAGCTACGGAATTCCACGACTGTAAGTCGACATAGTAGCCGTTTGGGTCCGTCTAAGCTTGGAG
GCAAATAATTGGTCGTTTCGATTAGCCGTAAAGATTGGTGCGGATCGTTTCTGGGTGTCGTAGTGACAGCC
TCTGGCGTAGGGCCAATGTGAACCTAACTGACATTATGTTAATATTGTTACAGCCCAGCTACGCGATCAC
TAATACCCCCAGAGTTTCGAGCCGTGGCGCCTGCTTTCGGTGAAAACCGATGTGCGTTTGCTTTCAGGTTT
CCATGTCCTAGCAGTAGGTATTATGATATCGGATAAAGGATGGCCCTGTTCCGCAGGCGCAATTGTTCCGAT
TCGATATCATAACCGAAAATGGGAACCTGGGTACGATGGTCCTATATTGGTGAACCGTTTGCTTACTTTTGGC
GGCGTCAATAGTTGAGTTTTCTTCCCGAAGGCTTATATTCCCTTCCCACGGCCCCCTATTTGTGTGCCCTT
GGTGTTTTTCGCGCTTGTGCGGCTTTGGCTTGGTTATGTCCGTCTTCTCTCTGCTCATCCAAATCCTTGGACG
AGTGGTGGAATGATTGATGTTTTCGGGTCAATTGAACGTTCCGGCGTATGAGTGGTGATCGG

>Lc1

CTTTTCCGGGCACTTTTCCGGCGGCGGTTTTAAGCGATTTTTGGCTCTTCTGGGGGGGATTATGACGAAA
GATGGTGTTTCCGTTGAGCTTGTGGTTGGCCTAGAGGAATGGGCGTGCTAGCGTGGAAGCATGGATTCT
TCTAGGCTCCAACGGGTGGCGTTTTTTCAGATCGAGGTGTCCCCATGGCATCGAACCACCCTTTCGCCAG
GCGGCTGGGCGTGCTTCGTTGGAAGTACGGATCCGTCTGGGCTGTCCCGATCGGAAAAGCAGGGGC
GTTTCGCCCATGGAAATCGTTCACCCTGCGTCGGGAACCTGATGCGGAACCTGGGCTCTAGGCGGCGATTTA
AAGTGGCATATGAAAGCGGGTTCGAGAAAAATCTGATCTCTTCGGTAGTTAGCTCACACTGTCCGGGTGAA
CCAACAAAACGAGATTCGATTTCGATTGTCCCCGAACGATTTTGCTCAACTTAAAATCGTCGTTGATATCCCT
CTCCGCAGATGTTCTTGGAACCATCCCGCGATGGAAGTCCGGCCATTTGAAGGGTCGGCCGTAGGTGCT
GATGGGCGACGGTGCCGAAGGAAATTTGCCAACGGGTTCCGAAGCGATCACTCTGTCCGAAGTCTCTTT
GGTTTGGCGAAGAATATCCAATTATTGCGGTGAGCCATTTGCATAATCCTGACCACAGTGGACGCTAGTA
ACCTTTTGATGTCGAAAATGGATGGCACTGGCATGCGTAACGCGATAAGCTTTCTTAGAGATTACGTCCCC
AGGTGCAGTTGTTCCCTCTTAATGATGTCATCGGGCGAAGGCCGAAATGTGCGACGGTTTGTGGGTTTGG
TGAGATCTCAGTTTTGGTGTTTCCATAGGCAGATCATCGAGAACAGTAAGATCTTTGTGACCGTAGACGTTT
TCCAAAGGCGATGACCGAAGTATCAGCCTGATGATGTAGCTACGTTACTGCCCGGGCGGATGGGCGTGCT
TTTATTAATTTATGTGGTGGGAAAGCACGGATCCGCCTAGGCTGTGTCCCGACGGGTTAAAGAAGGTGGT
GGTGTCATCCCATCGGCATCGTTCCATCTTTTTCGTGCTTGTGCGGCTTTCCGGCTTGGTTACGTCCGTCTTC
CCGTGCTCAACCGAGTTCAACCTCGGAAGGGTTGCGAGAAAGATTGATGTTTTCCGGTCAACTGAACGT
TCGGCGTATGAGTGGTGATCGG

>Lc2

CTTTTCCGGGCACTTTTCCGGCGACTTTCCGGCGAGTTTCCGGTGGTTCATTTTCTTCCCTTTATTTTTTCG
CGCTTGTGCGGCTTTGGACTGGTTTCATTTGTCTTCCTATTGCATTTTCGATTATATCGAAAGAGTGGTTGG
AAAGATTGATGTTAGCGGGCCAACGAACGTTCCGGCGTATGAGTGGTGATCGG