tcatgcttcactacagcaatgaaagttggtcgccatcaagagcttctagatagccaagtaaggaatgagatgagcgctgaaa

tgctagaagagatcacatacctcctgatgcg**atgcataagca**cattggcgtagctaggggggtagtcctggtgggctcggg

actaccccccgaatttggcccagggagccccaccccccaaaaattttagcccagacttaagcccaagaagccagcccaat

aagaaatgaaaaattagaaaggctagggtttggagttcgctgcaacccgacgctacttcgtttcgcacgcgtgtgacgacc

gcccatgccgccgccgttcctcctttcccggcctcgccagctcgccttgccctcgcattggttccgcagctcccacctctccc

ctgtcgttctttgtccgccgaggcactgagccgctgccgtcgttgacctcccggctgctcaatccaggagaggcggactga

cggaggttagtggcgccgtcacgccgctgaaccgccaagcgccgctggtcgccggcgccgggagcccaaggcccctg

gccggttggccctagttgattgccgctcctCCCCCTTGCCCTGCCTGTGCCAGCGGCTGCCTT

AGTAGCCTATGACGAGgtgagtttaagtgctaactggtgaattgaagctatgagcctatgaaatcaaatagatg

↑

Transcription Start Site of *GN2*

tatatttctaactgtgatttgtgcatcagttagttagtgaatgactgaatgtgtgacgttgacttgctatggaatgtgaattagtga

attagagctatgaactgtgatttgtgatagtgctaattggtgaattagaactatgaactgtgatttgtgaatcactgaatgtgctgt

tgacttgctgccattttattttagtttgtaggatttctttggcaaattaagtagtagttttgataaggttggttcgagaaattggaaca

aatccgacgcacttgacaaacatgtagggttgtaacaagtaaacacaataaaaaagcttcagagaaattatttatgaaagatg

caccaattgctatatttctaattttgtatggccttttctctaaattttgatctgtatatgggatttacggtgattgaaaaattttagtatg

⇓   ⇓   **Alternative Splicing Site of *GN2.1* & *GN2.2***

tgatgtgaacttgtaatccagaagTAGGACTACCCAACAAAAATTTCCTGGCTACGCCAC

TGCATAAGCATGAATGGAGAAGAACGGCCAACGATGAAAGAGGTTGCTG

AGAGGCTGGAGATGTTAAGGAGATACCAGCAGCACCCATGGGTTGAAGCT

GAAGACAATGCAGAGGAGATTGAGAGCTTGCTCGGCAGGGAACAACAGA

ATGCAAATTACCAACTCGAGCAACAAAATGTCTTGTATTTGGAAGAAGGCC

GTAACTATACATTTAGCATGTAGATCACGCTGTTTTGAATTATGTTTTGTAAA

TCAATCAACACAGATATGACTTCTTTTTCTATGTAACATGAAAAAGCAAATT

TATGGCCATTGCAGAAAGGACAAAGAGTTTGGAAATGAAATCTTAACATGT

AAATCTTTATTGGACAATTAATATAGCTCCCATGCAT**AATAAA**TTGCAATGA

AAGAACCA

Figure S5. The sequence at the locus of *GN2* in 'YIL19'. The black vertical arrow indicates the transcriptional start site for *GN2*. The two exon sequences of *GN2* are in capital letters. The adenylation signals (AATAAA) are in bold and double underlined. Alternative splicing sites are indicated by white vertical arrows. The inserted sequence is flanked by the short direct repeats of TGCATAAGCA, which are underlined.