

**Additional File 4: Occurrence of identical tags in related and unrelated contigs.**

>CTG1027103 (F2009.210 12S cruciferin seed storage protein, At4g28520, 1.5732e-45)  
GTTTCTGTCTCGACGTCAGCAATGAGGCCAACCAACTTGATTTCCAACCCAGAAGATTCTATCTTGCCGGCGACCCA  
CAAAATGAATTTTCAGCAGCGACAGCAACAACCACAAGAACAACAACAAGGAAGCGAGGGCCAGCAGCAACANCNNGAA  
GGGGGCGGCAGTGAGGGCAGATNNNNNNNNNGNNNNNGNAGACAACATCTTCAGTGGTTTTCGATGCACAGCAGTTGGCT  
GAGGCATTTAACGTGGATGTCCAGCTCATAAGGAAGCTTCAGGGCCAAAATGACAGGAGGGGCAACATAGTGAGAGTG  
GAGGGAGGCCCTACAGGCACTGTTACCACCAAGGGGACAACAGGAGCGAGGAGAGCAACAGCAAGATCACCTCCATGCA  
CGCGGAAATGGCTATGAAGAGACCATATGCAGCTTGAGGCTCAAACAAAACATCGGTGATCCATGGAGAGCAGATGTC  
TACACACCTAGAGGTGGACACAGGAGCAGCGTCACCGGCTATGATCTTCCAGTTCTCCAAAAGCTTGTTAAATTAAGT  
GCTCATAAAGGACGTCTATATCAGGGCGCATTGGGGCTACCATACTACAACGTGAATGCCAATAGTGTAAATCTACGCT  
ATAAGAGGCAGTGCCAGGATTCAGTAGTGCAACAGCAGNNNCAGACAGTGGCCAATGAGGAAGTCCAGCAGGGTCCAG  
GTGTTGGTTATTTCCACAGAACTTCGCTGCGTTGATAAAAAGCAAGGGATTTCAGGCTTTGAGTATGTGGC**GATCAAGACT**  
**GATGAAA**ATGCCATGATTAACACTTTAGCANNCAATCTTTTCGCTGAT

>CTG1027117 (F2009.210 12S cruciferin seed storage protein, At4g28520, 1.15902e-40)  
ACATATGCCCGGGATTTTCGGCCATNACNNNNNGGGGGGACAACAGGAGCGAGGAGAGCAACAGCAAGATCACCTCCA  
TGCACGCGGAAATGGCTATGAAGAGACCATATGCAGCTTGAGGCTCAAACAAAACATCGGTGATCCATGGAGAGCAGA  
TGTCTACACACCTAGAGGTGGACACAGGAGCAGCGTCACCGGCTATGATCTTCCAGTTCTCCAAAAGCTTGTTAAAT  
AAGTGCTCATAAAGGACGTCTATATCAGGGCGCATTGGTGCTACCATACTACAACGTGAATGCCAATAGTGTAAATCTA  
CGCTATAAAGAGGCAGTGGTACAACGCAGAGTGGCCATTACGGCCGGGGTGAACAGCAGGGCCAGACAGTGGCCAATG  
AGGAAGTCCAGCAGGGTCCAGGTGTTGGTTATTCCACAGAACTTCGCTGCGTTGATAAAAAGCAAGGGATTTCAGGCTTG  
AGTATGTGGC**GATCAAGACTGATGAAA**ATGCCATGATTAACACTTTAGCCGGCAATCTTTTCGCTGATGCGGGCTATGC  
CAGTCCAAGTCATAGCCAGCGCTACCAGGCGTCCAATAACGAGGCTAAGCAGCTGAAGCATAACAGGCAGGAGAGCA  
CTATAGGGGCTCCGGGAAGTCTAGGTCTGAGTAGAGGGCTCTGCTTAATGCGTCAACTTTGTTTTATCAGGTTGTT  
GTATAGGTTTGGAAAATAATAAGANNCCATATTACTTGGCTTTTTTTGTC

>CTG1027135 (MLN1.4 legumin-like protein, At5g44120, 3.62753e-86)  
CAGCTTAGTAGTCTTATCTGCAGTGTACCATGGCTAAGAGTCTTTTGTTCCTTTTGGCTTTGCTTTCTTATTTTCTTC  
AATTACTGCTCGGCTCGGCAGCAAAAGTCTAGGCAGCAAAATGAGTGCAGCATCTCTAGACTCACTGCTCAGAAACCT  
TCCAACCGGATCCAGTCCAGAGCCGGTGTGACCGAGGTCTTTGATCATAACAATGAGCAATTCCAGTGCAGGGGTGTG  
GCTGTTGTTTCGCTATACTATTGAACCAACCGCTTGCTTTTGCCTTCCCTACGTCAATGCGCCTCAACTCTTGTACTTC  
GTCCAAGGTAGAGGCTTGCAGGGAATCCTGATCAGTGGTTGTCTGAAACCTTCCAATCATTCCAGGAATCTCAGGAA  
GGGCATCAAGGAAGAGAACAGGAAGGGCAGCAAGGACAACAGGGACAACAAGGCCAGCAAGGACAACAAGGCCAGCAA  
GGACAACAAGGCCAGCAATTCCTTGGAGACCAACATCAGAAGATCCGAGAGGTGCAAGAGGGTGTGCTTTTGCAGTG  
CCCACCGGATTTGGGCATTATATCTATAACAATGGTAACAGACAGCTTGTGTAGTTTCTGTCTCGACGTCAGCAAT  
GAGGCCAACCAACTTGATTTCCAACCCAGAAGATTCTATCTTGCCGGCAACCCACAAAATGAATTCAGCAGCAACAG  
CAACAACAACAAGGAAGCGAGGGCCAGCAGCAACAGCAAGAAGGGGGCGGCAGTGAAGGTAGAGGGCAAGAGAGTTCC  
GGCGACAACATCTTCAGTGGTTTCGATGCACAGCAGTTGGCTGAGGCATTTAACGTGGATGTCCAGCTCATAAGGAAG  
CTTCAGGGCCAAAATGACAGGAGGGGCAACATAGTGAGAGTGGAGGGAGGCCTACAGGCACTGTTACCACCGAGGGGA  
CAACAGGAGCGAGGAGAGCAACAGCAAGATCACCTCCATGCACGCGGAAATGGCTATGAAGAGACCATATGCAGCTTG  
AGGCTCAAACAAAACATCGGTGATCCATGGAGAGCAGATGTCTACACACCTAGAGGTGGACACAGGAGCAGCGTCACC  
GGCTATGATCTTCCAGTTCTCCAAAAGCTTGTTAAATTAAGTGCTCATAAAGGACGTCTATATCAGGGCGCATTTGGTG  
CTACCATACTACAACGTGAATGCCAATAGTGTAAATCTACGCTATAAAGAGGCAGTGCCAGGATTCAGTAGTGCAACAG  
CAGGGCCAGACAGTGGCCAATGAGGAAGTCCAGCAGGGTCCAGGTGTTGGTTATTCCACAGAACTTCGCTGCGTTGATA  
AAAGCAAGGGATTTCAGGCTTTGAGTATGTGGC**GATCAAGACTGATGAAA**ATGCCATGATTAACACTTTAGCCGGCAAT  
CTTTCGCTGATGCGGGCTATGCCAGTCCAAGTCAATGAGGAGGCTTACCAGGCTCCAATAACAGGCTTAAGCAGCTG  
AAGCATAAGGAGGAGGAGCACTATAGGGGCTCCGGGAAGTCTTAGGCTGAGTAGAGGGCTTACCTGCTTAATGCATC  
AACTTTGTTTTATCAGGTTGTTGTATAGGTTTGGAAAATAATAAGAGGCCATATTACTTGGCTTTTTTTGTACCCCATG  
ATGTACAATGAGAGTTGAGTTTTTTGCTCTTTAATAATAATAATAATAACATCAATAATGCTTATGCAGTAAA