

Chestnut-waxy-gene Supplementary materials

1.Sequences:

>Chinese_Chestnut_Yan_HongGenome_EVM0027964

MATVTTSHFVSSSHVNCHGTSGSESKVNFLRSQAMTHNGLRLLNKLDILKMRTQTKAIA
RPARSTVSKTEKDRPNGKIVCEQGMNLVFGAEIGPWSKTGGLGDVLGGLPPAMAAKGHR
VMTVSPRYDQYKDAWDTSVLVEMKVGDRIVRFFHCYKRGVDRVFDHPTFLEKVVWGKT
GSKIYGPRAGLDYEDNQLRFSLLCQAALEAPRVLNLKSNEYFSGPYGEDVIFIANDWHTA
LLPCYLKTMYSRGLYKNAKVAFCIHNIAYQGRFPFSDFSLLNLPDQFKSSFDFIDGYEK
PVKGRKINWMKGGIIESDRVLTVSPYYAQELVSGIDKGVELDNIRKTGITGIVNGMDVQ
EWNPSRDKYIDIKYDSTTVMDAKPLLKEALQAEVGLPVDQNIPLIGFIGRLEEQKGSIL
VEAISFIGQDVQIIILGTGKKPMEKQIEQLEVLYPNSARGVAKFNVPLAHMVIAGADFM
LIPSRFPCGLIQLHAMRYGTVPIVASTGGLVDTVKEGFTGFHMGAFNVECDVADPADVT
AVATNVKRALATFGTPVLKEMIQNCMAQDFSWKGPARGWEKMLLSLDVAGSEPGTEREEI
APLAKENVATP*

>Chinese_Chestnut_Yan_HongGenome_ManualEdited

MYSNRSSKKEWDFYGRLALEAPRVLALNSNEYFSGPYGEEVAFYIHNIAYQGRFAFTDF
SLLNLPDELKSSFDFIDGYEKPVKGRKINWMKAGILES DKVLTVSPYYAQELFSGEDKGV
ELDNIRKTGIQGIVNGMDVQEWNP LTKFTNVKYDASTVLDAKPLLKEALQAEVGLPVD
RNIPVMGFIGRLEEQKGSILVEAIPHFIKENVQIIVLGTGKKPMEKQLEQLEKIYPDKA
RGVAKFNVPLAHMIIAGADFILVPSRFPCGLIQLHAMRYGTVPIVASTTGLIDTVKEGF
TGFQMGFSNVEVITHFLSHARAIATYGTPAFTEIIQNCMAQDLSWKGPAAKKWEEVLLSLG
VAGSEPGIEGEEIAPLAKENVATP*

>Chinese_Chestnut_Yan_HongGenome_EVM0030994

MATVAGSNFVSTSSRVNYGSVTSGSEAKVALMKFGLINHVTTHSGLRSLNKVDELHMHVRT
MAKLTAQTRSKTFKSGNAWTSRTIVCGSGMNLIFVGTEVGPWSKTGGLGDVLGGLPPAM
AANGHRVMTVSPRYDQYKDAWDTDVVIELKVGDKIEKVRFFHCHKRGVDRVFDHPAFLE
KVWGKTKSKIYGPIAGEDFQDNQLRFSLLCQAALEAPRVLALNSNEYFSGPYGEEVIFIA
NDWHTALIPCYLKTIIYKPKGIYSTARVAFCIHNIAYQGRFAFTDF SLLNLPDELKSSFDF
IDGYEKPVKGRKINWMKAGILES DKVLTVSPYYAQELVSGEDKGVELDNILRKTGIQGIV
NGMDVQEWNP LTKYTNV KYDASTVLDAKPLLKEALQAEVGLPVD RNIPIGFIGRLEEQ
KGSILVEAIPHFIKENVQIIVLGTGKKPMEKQLEQLEKIYPDKARGVAKFNVPLAHMII
AGADFILVPSRFPCGLIQLHAMRYGTVPIVASTGGLVDTVKEGFTGFQMGFSNVECEAV
DPADVTA VATNAKRAIATYGTPAFTEIIQNCMAQDLSWKGPAAKKWEEVLLSLGVAGSEPG
IEGEEIAPLAKENVATP*

>Chinese_Chestnut_Yan_HongGenome_EVM0027964

ATGGCAACTGTGACCACCTCACACTTTGTATCCAGTTCACATGTCAATTGCCATGGAACG
TCCGGATCAGAAAGTAAAGTAACTTTGGCCTAAGGAGTCAAGCAATGACTCATAATGGG
TTAAGGTTGTTGAACAAGCTAGATATACTAAAGATGAGAACCCAAACAAAAGCAATTGCT
AGGCCAGCAAGAAGCACAGTGAGTAAACTGAGAAGGACAGACCCAATGGGAAAATTGTA
TGTGAACAAGGAATGAACCTGGTGTGTTGTGGGAGCTGAGATAGGTCCATGGAGCAAGACT
GGTGGACTTGGTGATGTTCTTGGAGGACTTCCACCTGCAATGGCAGCTAAGGGGCACCGT
GTTATGACTGTGTCTCCACGCTATGACCAATACAAAGATGCATGGGACACAAGTGTCTA
GTTGAGATGAAAGTTGGGGATAGAATTGAAACCGTTCGCTTCTTCCACTGCTATAAACGA
GGAGTTGATCGTGATTTCGTGGATCACCCAACGTTCTCGAAAAGGTATGGGGAAAAACT

GGATCCAAAATCTATGGCCCTAGGGCTGGTCTGGATTACGAGGATAACCAGCTGCGATT
AGCTTGTGTGCCAGGCTGCTCTGGAGGCACCAAGGGTTCTGAACTTGAAAAGCAATGAA
TATTTCTCCGACCATATGGGGAGGATGTTATCTTCATTGCCAATGATTGGCACACTGCA
CTTCTTCCATGCTACCTAAAAACAATGTACAAATCAAGGGGACTTTATAAAAATGCCAAG
GTTGCTTTCTGCATTCATAACATAGCTTACCAGGGAAGATTCCCCTTCTCAGACTTTTCA
CTTCTCAATTTGCCTGATCAATTCAAGAGTTCTTTGACTTTATTGATGGCTATGAGAAG
CCAGTGAAAGGAAGGAAAAATCAATTGGATGAAGGGTGGAATAATAGAATCGGACAGGGTC
CTGACTGTGAGCCCATACTATGCCCAGGAACCTGTTTCTGGAATTGACAAGGGTGTTGAA
TTGGATAACATCATTTCGTAAAGACTGGCATCACTGGTATCGTGAATGGGATGGATGTTCAA
GAGTGGAATCCTTCCAGGGACAAATACATTGATATTAATATGATTCCACAACCGTTATG
GATGCAAAGCCTCTATTGAAGGAAGCCCTTCAAGCAGAAGTTGGGTTGCCTGTTGACCAA
AACATCCCTTTGATAGGCTTCATTGGAAGACTAGAAGAGCAGAAAGGTTTCAGATATTCTA
GTAGAAGCTATTTCAAAGTTCATTGGGCAGGATGTTCAAATTATAATCCTTGGAACCTGGC
AAAAAGCCCATGGAGAAGCAAATTGAACAGCTGGAGGTACTATATCCCAACAGTGCCCGG
GGTGTGGCAAAATTTAACGTCCCCTTAGCTCATATGGTGATTGCTGGGGCTGATTTTATG
CTGATCCCAAGTAGATTTGAACCCTGTGGTCTCATTCAAGTTACATGCCATGCGATATGGA
ACGGTACCTATTGTTGCCTCAACCGGTGGACTTGTGATACTGTCAAAGAAGGATTTACA
GGATTCCATATGGGAGCTTTCAATGTTGAATGTGATGCAGTTGATCCAGCTGATGTAAT
GCAGTAGCCACTAATGTCAAGAGAGCTCTTGCAACTTTCGGTACCCCTGTTTTGAAAGAG
ATGATCCAGAATTGCATGGCCCAAGATTTTTCATGGAAGGGACCAGCCAGACAATGGGAA
AAGATGCTGCTGAGCCTGGATGTTGCTGGCAGTGAACCTGGAAGTGAAGGGAGGAGATT
GCTCCTCTTGCCAAGGAAAATGTTGCCACTCCTTGA

>Chinese_Chestnut_Yan_HongGenome_ManualEdited

ATGTATAGTAACCGAAGTAGTAAGAAAGAATGGGACTTCTATGGTAGACTGGCAGCTCTA
GAGGCACCAAGGGTTCTAGCACTTAACAGCAATGAATACTTCTCTGGACCATATGGTGAA
GAGGTTGCCTTTTACATTCACAACATTGCTTACCAAGGCAGATTTGCCTTCACAGATTTT
TCACTTCTCAATCTCCCAGATGAATTGAAAAGCTCTTTTGATTTTATTGATGGGTATGAA
AAGCCAGTTAAGGGAAGAAAAATCAATTGGATGAAAGCTGGAATTTTAGAATCAGACAAG
GTTTTAACCGTGAGCCCATACTATGCCCAAGAACTTTTTTCTGGAGAAGACAAAGGAGTG
GAATTGGATAACATCATTTCGTAAAACCTGGCATTCAAGGAATTGTGAATGGCATGGATGTC
CAGGAATGGAACCCCTTAAGTACAAATTTACAAACGTCAAATATGATGCTTCAACTGTG
TTGGATGCAAAGCCTCTTTTGAAGGAAGCCCTCCAAGCTGAAGTTGGATTGCCAGTCGAT
AGAAACATCCCTGTCATGGGCTTCATTGGTAGACTTGAAGAGCAGAAAGGTTTCAGATATT
CTTGTAGAAGCCATTCCCCATTTTATCAAAGAGAATGTTCAAATAATAGTCCTTGGGACT
GGCAAAAACCAATGGAGAAGCAGCTTGAACAGCTGGAGAAAATATACCCTGACAAGGCC
CGAGGAGTGGCAAAATTCAACGTTCCCCTGGCTCATATGATAATAGCTGGTGCTGATTTT
ATACTGGTTCCTAGTAGATTTGAGCCCTGTGGTCTCATTCAATTACATGCCATGCGTTAT
GGAAGTGTGCCTATTGTTGCCTCACTACTGGATTAATTGACACGGTTAAAGAAGGCTTC
ACAGGATTTCAAATGGGAAGCTTCAATGTTGAAGTAATAACTCATTTCTTTTACATGCT
CGGGCCATCGCAACCTATGGAACCTCAGCTTTCACTGAGATAATACAGAATTGCATGGCT
CAAGATCTCTCATGGAAGGGACCAGCTAAGAAGTGGGAGGAGGTGCTGCTAAGTTGGGG
GTGGCAGGAAGTGAACCTGGAATCGAGGGTGAGGAAATAGCTCCACTCGCAAAGGAAAAC
GTTGCAACACCTTAA

>Chinese_Chestnut_Yan_HongGenome_EVM0030994

ATGGCAACTGTGGCTGGTTCAAACCTTTGTATCAACAAGTTCTCGTGCTCAACTATGGATCA
GTAACCTCTGGATCAGAGGCTAAAGTAGCACTGATGAAATTTGGATTAATAAATCACACC
GTGACTCACAGTGGGTTAAGATCTTTGAACAAAGTGGATGAACTAATGCATGTCAGAACC
ATGGCAAAATTAAGTCTAGGCAAACAGGAGTAAACATTCAAGAGTGGGAATGCATGG
ACTTCACGGACCATTGTTTTCGGAAGTGGGATGAATTTGATCTTTGTGGGGACAGAAGTA
GGTCCCTGGAGCAAAACCGGTGGACTTGGAGATGTTCTTGGAGGTCTGCCACCAGCAATG
GCGGCCAACGGGCATCGTGTTATGACTGTCTCTCCACGTTATGATCAGTACAAAGACGCG
TGGGATACAGACGTTGTAATTGAGCTTAAAGTAGGAGATAAAATCGAAAAGGTTGCTTC
TTCCATTGCCACAAAAGAGGAGTTGATCGCGTTTTTGTGGATCACCCAGCGTTTCTTGAA
AAGGTGTGGGGGAAAACCAAATCCAAATTTATGGGCCTATTGCTGGAGAGGATTTCCAG
GACAACCAACTTCGATTACAGCTTATTATGCCAGGCAGCTCTAGAGGCACCAAGGGTTCTA
GCACTTAACAGCAATGAATATTTCTCTGGACCATATGGTGAAGAGGTCATATTCATTGCC
AACGACTGGCACACTGCCTTGATCCCATGCTACCTGAAAACCTATATACAAACCCAAAGGC
ATATACAGCACTGCCAGAGTTGCCTTTTGCATTACAACATTGCTTACCAAGGCAGATTT
GCATTCACAGATTTCTCACTTCTCAATCTCCCAGATGAATTGAAAAGCTCTTTTGATTTT
ATTGATGGGTATGAAAAGCCAGTTAAGGGAAGAAAAATCAATTGGATGAAAGCTGGAATT
TTAGAATCAGACAAGGTTTTAACTGTGAGCCCATACTATGCCCAAGAACTTGTTTCCGGA
GAAGACAAAGGAGTGAATTGGATAACATCCTTCGTAAAACCTGGCATTCAAGGAATTGTG
AATGGCATGGATGTCCAGGAATGGAACCCCTTAAGTACAAATACACAAACGTGAAATAT
GATGCTTCAACTGTGTTGGATGCAAAGCCTCTTTTGAAGGAAGCCCTCCAAGCTGAAGTT
GGATTGCCAGTGGATAGAAACATCCCTATCATAGGCTTCATTGGTAGACTTGAAGAGCAG
AAAGGTTCAAGATATTCTTGTAGAAGCCATTCCCCATTTTATCAAAGAGAATGTTCAAATA
ATAGTCCTTGGGACTGGCAAAAACCAATGGAGAAGCAGCTTGAACAGCTGGAGAAAATA
TACCTGACAAGGCCCGAGGAGTGGCAAAATTCATGTTCCCCTGGCTCATATGATAATA
GCTGGTGCTGATTTTATACTGGTTCCTAGTAGATTTGAGCCCTGTGGTCTCATTCAATTA
CATGCCATGCGTTATGGAAGTGTGCCTATTGTTGCCTCAACTGGTGGATTGGTTGACACG
GTTAAAGAAGGCTTCACAGGATTTCAAATGGGAAGCTTCAATGTTGAATGTGAAGCTGTT
GATCCTGCCGATGTGACTGCAGTGGCTACAAATGCGAAAAGGGCCATCGCAACCTATGGA
ACTCCAGCTTTCACTGAGATAATACAGAATTGCATGGCTCAAGATCTCTCATGGAAGGGA
CCTGCTAAGAAGTGGGAGGAGGTGCTGCTAAGTTTGGGGGTGGCAGGAAGTGAACCTGGA
ATCGAGGGTGAGGAAATAGCTCCACTCGCAAAGGAAAACGTTGCAACACCTTAA

2. Gene distribution and gene length

GeneID	Chr	Gene Range	mRNA Range	No. Exon	ORF Len
EVM0027964	Chr02	26150031:26155577	26150031:26155577	14	1836
EVM0030994	Chr08	50511551:50515651	50511551:50515651	13	1854
ManualEdited.gene	Chr08	50521016:50523688	50521016:50523688	10	1155

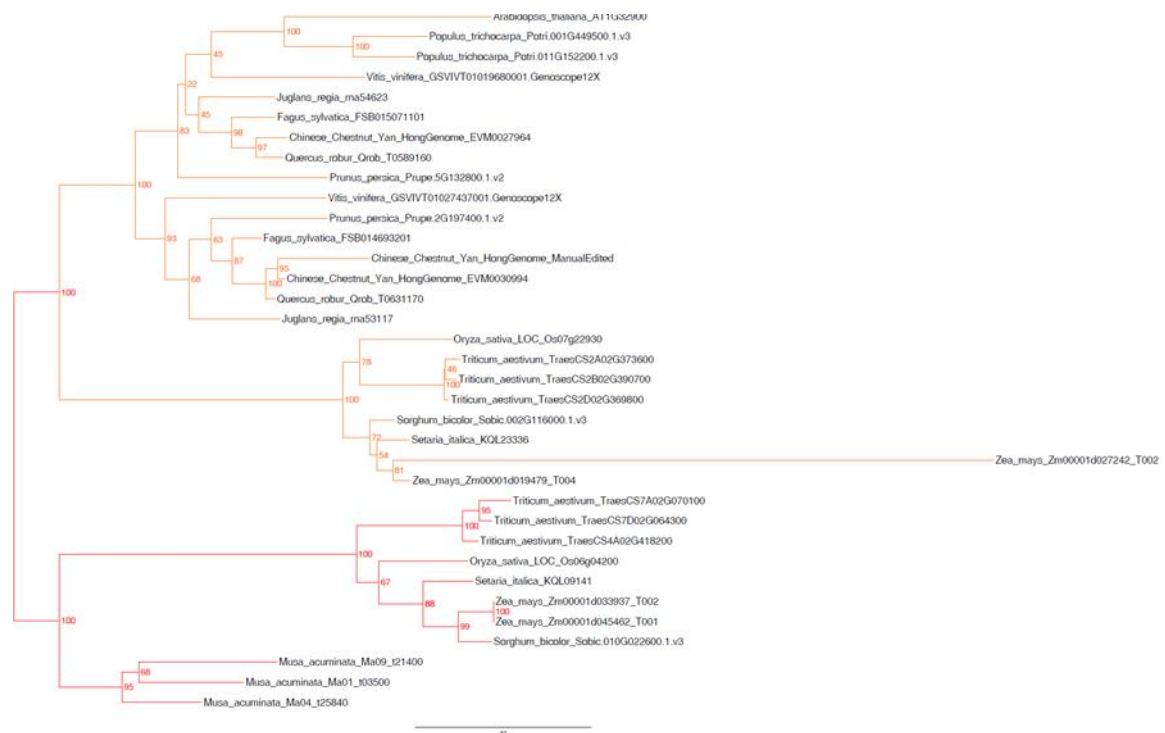
Note: two of the three members are tandem repeats

3. Subcellular localization prediction

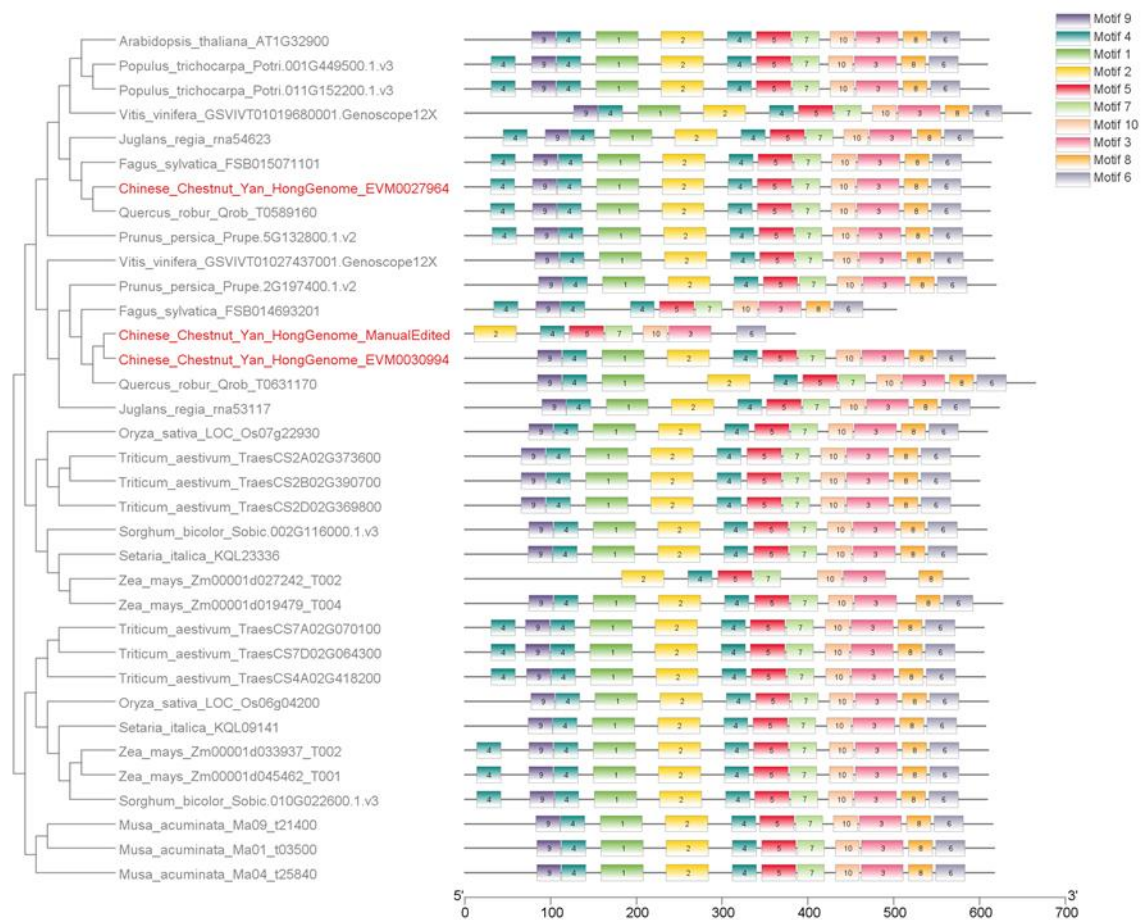
	pI	Mw	Localization
EVM0027964	7.94	67564.8	Chloroplast
ManualEdited	5.21	42587.89	Cytoplasmic
EVM0030994	6.66	67692.85	Chloroplast

Note: Two proteins were localized on chloroplasts, and one protein was localized in the cytoplasm.

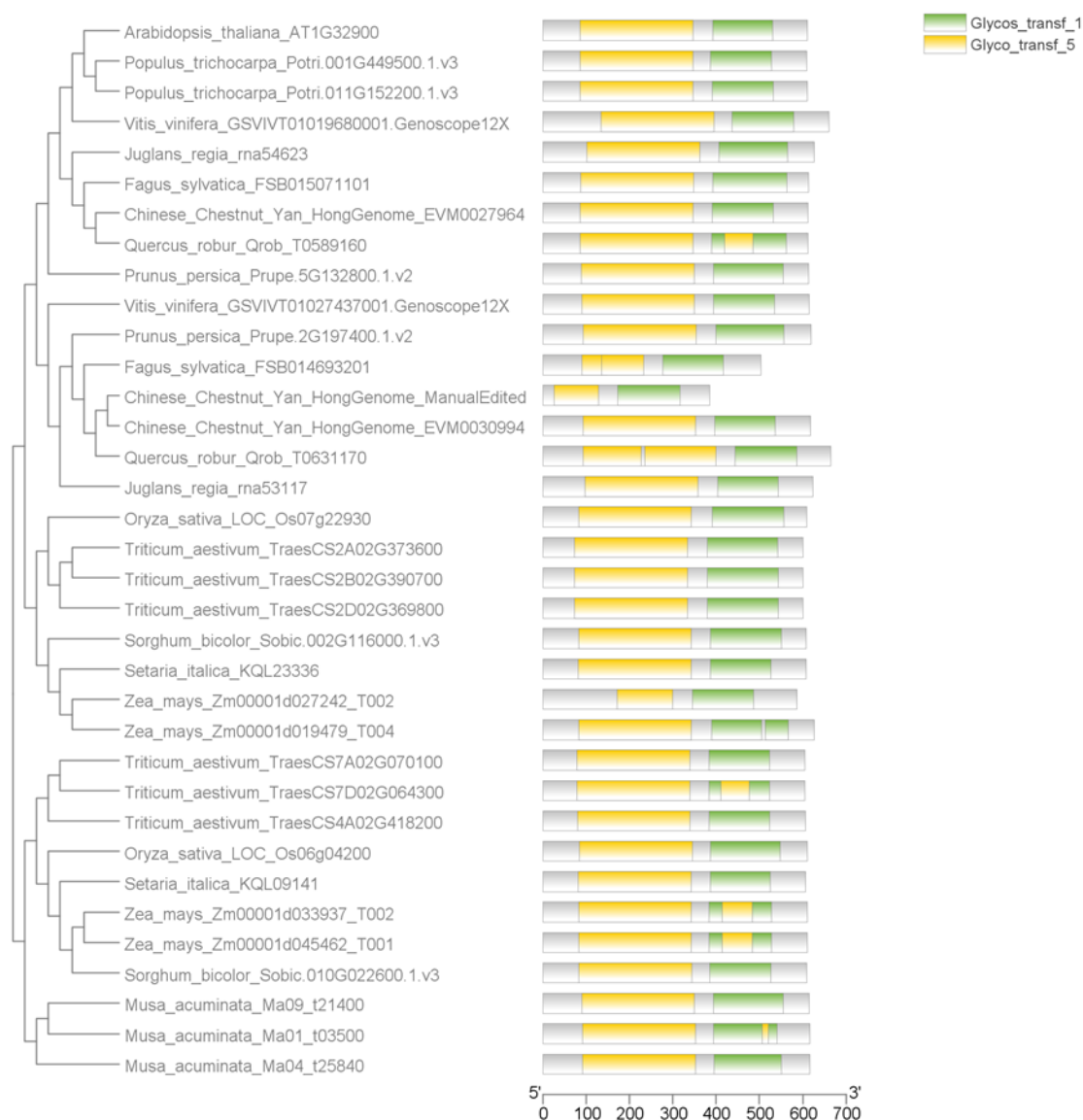
4. Evolutionary tree



5. Motifs analysis



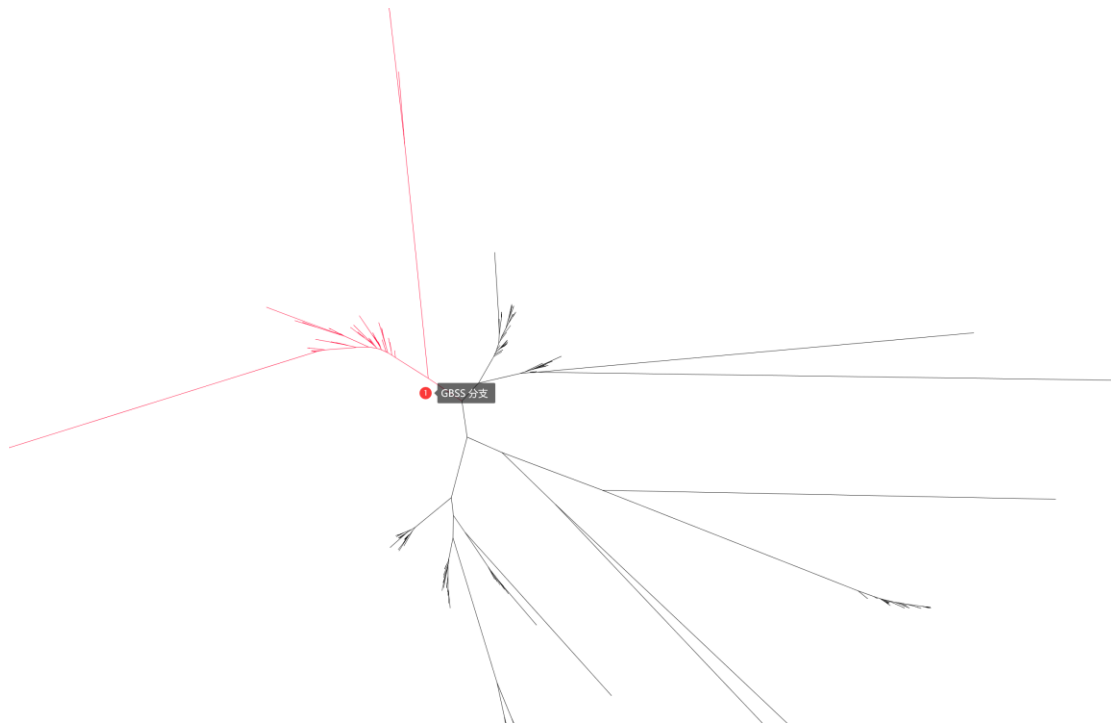
6 . Conserved domain prediction



Note : All sequences have the starch synthase catalytic domain Glyco_transf_5 and glycosyltransferase domain Glycos_transf_1 both domains.

7. Is waxy (GBSS II) family expansion present in chestnut?

A clear GBSS clade was identified by phylogenetic tree construction of all GBSS and their close family members, as follows



GBSS II in dicotyledons can be divided into two branches, and most species have only one GBSS II-b member in each branch



Gene structure annotation information in IGV reveals the proximity of the two GBSS II genes on chromosome 8 within a 14 Kb region, so that the source of the expansion of GBSS II genes in chestnut is tandem duplications.

