
Genome-wide comparative analysis of papain-like cysteine protease family genes in castor bean and physic nut

Zhi Zou^{1*}, Qixing Huang², Guishui Xie¹ & Lifu Yang¹

¹ Key Laboratory of Biology and Genetic Resources of Rubber Tree, Ministry of Agriculture, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Baodaoxincun, Danzhou 571737, Hainan Province, China

² Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Xueyuan Road 4, Haikou 570100, Hainan Province, China

*Correspondence and requests for materials should be addressed to Z.Z. (Email: zouzhi2008@126.com or zouzhi@catas.cn)

Supplementary Files:

Supplementary File S1. The gene model for *RcRD21C*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S2. The gene model for *RcXCPI*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S3. The gene model for *RcXBCP3*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S4. The gene model for *RcXBCP3L*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S5. The gene model for *RcPAP3*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S6. The gene model for *RcCTB1*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S7. The gene model for *JcSAG12H8*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated

regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S8. The gene model for *JcRD21B*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S9. The gene model for *JcCEP2*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S10. The gene model for *JcXBCP3L*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S11. The gene model for *JcTH11*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S12. The gene model for *JcSAG12H1*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S13. The gene model for *JcSAG12H2*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary File S14 The gene model for *JcSAG12H7*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The intron sequences are marked with lowercase letters. The start and stop codons are marked with bold letters.

Supplementary Figures:

Supplementary Fig. S1. Detailed information of 15 motifs identified in this study.

Supplementary Fig. S2. Matched positions of 25 RcPLCP genes on physic nut chromosomes. The positions were based on the synteny analysis and only 8 matched chromosomes were shown. RcPLCP genes were shown behind that of physic nut.

Supplementary Tables:

Supplementary Table S1. Detailed information of 31 PLCP genes in *Arabidopsis*.¹ “Yes” represents genes containing alternative splicing isoforms; ² “S, M and C” represent signal peptide, mitochondrial targeting peptide or chloroplast transit peptide, respectively, which was predicted by iPSORT; ³ duplicated modes were determined based on the study of Wang et al. (2013).

Supplementary Table S2. List of orthologous groups of the PCLP gene family based on the analysis of representative plant species. “OG” represents “orthologous group”. “-” represents “Not Detected”.

Supplementary File S1. The gene model for *RcRD21C*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1 tccatztatcctcttataactcctatcttcttcgagtttaggtcactaaattaattaacattt
1                                                                                               M
61 cttgcaaaatttgcattcctcttataaatatacacataccccaatctcccttactcttcA
2   Y N P L L K N Q N S F H N I S T M A L P
121 TGTATAACCCTCTTCTCAAAAACCAAAATTCTTTCCACAACATATCAACCATGGCATTAC
22   I S L S T L L F L F F T L S S A W D M S
181 CCATATCATTATCAACCCTTCTCTCCTTTTCTTCACTCTATCATCTGCCTGGGACATGT
42   I L S H N H G H H H Q S S W R S D N E V
241 CCATCCTCAGCCATAACCATGGCCATCATCATCAATCAAGCTGGAGAAGTGATAATGAGG
62   I S M Y N W W L A K H S K T Y N K L G E
301 TCATTAGCATGTACAATTGGTGGCTGCAAAGCATAGTAAAACGTACAATAAGCTAGGTG
82   R E K R F E I F K N N L R F I D E H N N
361 AGAGAGAGAAGAGGTTTGAGATCTTCAAGAATAACTTGAGGTTTATTGATGAACACAACA
102  S K N R T Y K V G L T R F A D L T N E E
421 ACTCTAAGAATAGAACATACAAGGTTGGATTGACTAGGTTTGCTGATCTTACTAACGAGG
122  Y R A K F L G T K S D P K R R L M K S K
481 AGTACCGTGCCAAATTCTTGGGCACAAAGAGTGATCCTAAGCGTAGATTAATGAAGTCCA
142  N P S Q R Y A F K A G D V L P E S I D W
541 AAAACCCAAGTCAGAGATACGCTTTTAAGCCGGTGATGTGTTGCCGAATCCATTGACT
162  R Q S G A V S A I K D Q G S C G
601 GGAGGCAGAGTGGAGCTGTTAGTGCGATCAAAGATCAAGGAAGCTGCGgtaagcaatttc
661 tatatcaattactaattagaaaactacatgaaaaatgctacatacagttattagatataa
721 aaaataatagaagattaatcaaggtatctcaatgcaatttttttttttttgagttgcttgg
781 aagaagaatggtggttaattttctggtcatgcagaaaggtagtgagaatggtagttact
841 ggatcttctagttagctagtgattcttttcagattacttataatatatgtagggatt
```

901 gtcccataaatgtcacaaaaatataaaaataattataaaaaatattcttgtacatttctgat
961 aataaaaaataactaaaaacagaatttttataaaaaactctaaaatcgagtcgagttt
1021 tatttataaaaaatccgatttgattttttcaaattaaagggtatatgttggatatattt
1081 tgaaaaaaagtatatatattcagtataattttgtaataataaattttaaataatattaa
1141 tttttataaaatattattgtagtttgattttgtaataattaataatgaataaataaa
1201 taaataaattttaaagataatttttaaaatgaagatgcattgagaatattat
1261 atccaaaaactagataaattgctataagataaaaaaaatgtttagtataataatgaaga
1321 aaaaaagtnnnataaaataaataaaatttaagaaaatcgaacgataaaaaatatacat
1381 ttgctaataatagtatatttgaagtatttgaaactttttaaataaaataaatttcaa
1441 ctcaaatgttctaacaatcaaccagtgatgtttcttaaaccatggtatacgttagat
1501 atgtgtaaaatattttaactaatttgaaaaataaaaaagataaaaaagtaaagaat
1561 tggtaattatgaaaagtgaagtacttttataatttttaaaaaagataaaataagaat
1621 ttcctctaataatttaaggtttgagtttgagcccaggatccagagccattaagtcaaac
1681 acccagatctatatctaaacttctaagtttcagtttgagtccagaccaagttaaagtca
1741 acttggactaaggctagtgtcaacttaatgagcgagatagcgacccaccaattgggtg
1801 ccacagttctggattcacggttccattttcattggaaccggatcgaaagaccaattccaa
1861 gccggtttcagtttacagttatggttcgatcagtttagatatatttaagaaaaaaaa
1921 agagtttaactcaattatacaattattaataacttattaataattaattacataattat
1981 taagtattaattacatagcattatattaataaataatttaattgttcttttttaatt
2041 ttaaggggttggaacctaaatggatagttctgattccatctaaaggaaaatcggacttga
2101 accgacctccacagtactgagctgattcaagccagtttcaaaattgaccaatttcaagc
2161 tggttttaaccagctccaatttcgaaccggttcgtggccaccactacacatcaatgttga
2221 gcgagattcaaaattaacttagttcaatctttagccaatctaattatcaaaataaaaa
2281 aattattaggaataactcatgtttttatcaaatttatggttttttttccatttttatt
2341 gtgttaaaaacaattgaaaataattgagaaacaactaataccaacagaaaaagattacc
2401 aactctaaattttttaaaatataaaactgtacttaaaaaatatttatacaataaaaat
2461 gatattttttgttaattctaataatattttgaaagattctctataagatttttcggttta
2521 gatcagcttagttttatcctttttcttgaataatgtgacttaattttgttatcaaatc

2581 caactcatatattggaagtacacacaaaaaattatattgaattgcttatgaactttttttt
2641 tttttttttttgtgttcttatcaaataatttgaattgtttatgaactattctttgtgtt
2701 tctatcaaaaaaaaaattagactgaacttcaaaattgatacatttttagatgatctgatag
2761 gtctagtgtggggcacctaattggagttgcataactcttgctataacaaaaagacttate
2821 aaatTTTTTTTTTAAAAAAAGTTAAACATTATTATGGAAGAGAAATCAGTAAATCAA
178
2881 attcaggtactttcagcacaagaataaaataaatatgcaatgccattgtgttggtgcagG
178 S C W A F S T I A A V E G V N K I V T G
2941 TAGTTGCTGGGCATTCTCAACAATAGCTGCAGTAGAAGGCGTAAACAAGATAGTAACAGG
198 E L I S L S E Q E L V D C D R S Y N A G
3001 GGAGCTAATATCTTTATCAGAACAAGAACTTGTAGACTGCGACAGGTCATACAATGCAGG
218 C N G G L M D N A F Q F I I N N G G I D
3061 TTGCAATGGAGGCTGATGGATAATGCTTCCAGTTCATCATCAACAATGGTGGCATTGA
238 T D K D Y P Y Q A V D G K C D T T K
3121 CACCGATAAAGACTATCCATATCAAGCTGTTGACGGCAAATGTGACACTACCAAGgtatt
3181 tgaagtgtatattactaagggtcattgatcacttctaactttttcattgtttgttttgggc
256 V K N K A V T I D G F E
3241 gtttattatTTTGGATTGTTTCAGGTGAAGAATAAGGCCGTCACCATTGATGGGTTTGGAG
268 D V M A F D E M A L Q K A V A H Q P V S
3301 GATGTTATGGCTTTCGATGAGATGGCCTTGCAGAAAGCTGTGGCACATCAGCCTGTCAGT
288 V A I E A S G M A L Q F Y Q S
3361 GTTGCCATTGAAGCCAGTGGCATGGCTTACAGTCTACCAGTCGgtgagttctttccca
3421 tttctttcaattcttgcagagaaatgagtttcttcttcttatctatTTTCTGAAACAT
3481 tagttatcaaagtaattgcagaaagTTTTTAAATCTCTTTTGGAGAGATAACATTGGT
3541 gagacatggaagtgggtgtcattttgcttgtatgttaaacaccaattgagtagaaacgtt
303 G V F T G E C G S A L
3601 gagttagttttcttttaattttttcagGGTGTATTCACTGGTGAATGCGGATCAGCCCTA
314 D H G V V I V G Y G T E D G I D Y W L V

3661 GACCATGGAGTGGTTATTGTTGGCTATGGAACAGAAGATGGTATAGATTATTGGCTTGTT
334 R N S W G R D W G E N G Y I K M Q R N V
3721 AGGAATTCATGGGGCCGTGATTGGGGTGAAAATGGATATATCAAAATGCAGCGTAATGTG
354 V D T F T G K C G I A M E S S Y P I K N
3781 GTTGACACATTTACTGGCAAATGTGGAATTGCGATGGAGTCTTCGTATCCAATCAAGAAT
374 T Q N P V K I S S V *
3841 ACCCAAAATCCTGTAAAGATCAGCAGTGT**TGA**atattgtccttttggttctctgtaaagc
3901 aaagggcaccaagagaaggaaaaacaagaagatcaaaggtgcaaaatcttatccttctt
3961 cttagtctaaagggcaagaaggcgttatgcgattgtatgaatgattggtgtgtatataa
4021 gctttgaatgtgtcaaacagcaccttaaaaatgagctgatcgtggaatctggattgcat
4081 tttctctgaagttgattgttctaataaaaacttaccgagcacataaatatgcagtcgttat
4141 cttttctgcatttgtaattgcttgctagctgaattgcttaatagaccaaaggtgcctag
4201 gcagtagggccaggctcagtagacaccaagaggagagagttagtgagacagggcacaat
4261 ttcttctctgaaaccaaacaaatcaacaagaagaagaacccaaaagagcgagcaaaga
4321 tctccatcaaaatgcatgatttcttctccaactttctacatgtgcaataaagttaaacia
4381 acttgctatctgtaagcttttagaattcttttttatattgataattgccaactttttcag
4441 tccaaaaacaatatcccatgacttcttttaggcattggctatgttaagttagaaacgca
4501 cacaaaaaaaaaacaatgaagcattaagtttgcttgatagataaggtgcctactatcaat
4561 tccatcatcagactggaagcggctcttacaaggggcaacttttgtttttctgttattat
4621 tatttttcttttaattggttgccgcgcttagagatggat

Supplementary File S2. The gene model for *RcXCPI*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1 tctattattactttatcaatatgcttatgtccctaaagcagaccaacaatcttggggcaa
61 aagcgtacccaaaaagaaccaaccgtaaagattcgggatgcattgaaaattaatttcagct
121 acttcacttttaagcaatttaagaaattaaacctgtatataactgctgcccggtcctcct
1
M A P
181 ccattcttcattcatcctaaccctcccttaccattaagcctcaacctaaataaATGGCTCCC
4 S V S S I F L F T I F T S L V I C S V V
241 TCAGTTTCAAGCATTTTCCTTTTACCATTTTACATCTCTCGTCATCTGTTTCAGTTGTT
24 A H D F S I V G Y S P E H L T S M H K L
301 GCTCATGACTTTTCTATTGTGGGTTACTCGCCGGAGCACTTGACTTCCATGCATAAACTC
44 T E L F E S W M S K H G K T Y E S I E E
361 ACCGAGCTCTTCGAATCATGGATGTCTAAGCACGGCAAAACCTACGAAAGCATTGAAGAG
64 K L H R L E V F K D N L M H I D R R N R
421 AAGCTACACAGACTCGAGGTTTTCAAAGATAACTTGATGCACATCGATCGTAGAAACAGG
84 D V T T Y W L A L N E F A D L S H E E F
481 GATGTTACGACCTACTGGCTTGGCGTGAATGAGTTTGCAGACTTGAGCCATGAAGAGTTC
104 K S K Y L G L N T D A P R G R I S E D L
541 AAGAGCAAGTACTTAGGGTTGAACACCGATGCTCCCAGGGGGAGAATTTCTGAAGACTTA
124 S Y R D V I D L P K S V D W R K K G A V
601 AGTTACAGAGATGTAATAGACTTGCCCAAATCCGTCGACTGGAGAAAGAAAGGAGCTGTT
144 A P V K N Q G S C G S C W A F S T V A A
661 GCTCCAGTCAAGAACCAAGGCTCGTGCGGCAGCTGTTGGGCATTCTCAACAGTTGCAGCG
164 V E G I N Q I V T G N L T S L S E Q E L
721 GTTGAGGGCATAAACCAGATTGTGACCGGAAATTTGACTTCACTGTCCGAGCAGGAGTTG
184 I D C D T S F N S G C N G G L M D Y A F
781 ATCGACTGCGACACGTCTTTAACAGTGGCTGCAATGGAGGCCTCATGGATTATGCATT
```

204 D Y I V N N G G L H K E E D Y P Y L M E
841 GACTATATCGTCAACAATGGAGGACTCCACAAGGAGGAAGACTACCCATACCTAATGGAA
224 E G T C D E K R
901 GAAGGCACTTGTGATGAAAAGCGGgtaaaaggcattgcaaaaaccattgttttacagta
961 taggctaattctaaactggactcaattgtcaaatgactagtaaaattgtgctaacatttc
232 E E M E V V T I S G Y H D V
1021 tatggctattgtcttcagGAAGAGATGGAGGTAGTGACTATCAGTGGTTACCATGATGTG
246 P E N N E E S L L K A L A H Q P L S I A
1081 CCAGAAAACAACGAAGAAAGCCTCTTGAAGGCACTAGCTACCAACCTCTTAGCATAGCT
266 I E A S G R D F Q F Y G R
1141 ATTGAGGCTTCTGGGAGAGATTTCCAGTTCTACGGCAGGgtaagaaaattgacatgcaca
1201 cacacacacatctgataaatttcgtagcgggtaagaccacatttcagaaaaatttctaa
279 G V F N G P C G T D L D H G V
1261 acagagggatctggcagGGGGTATTCAATGGACCTTGTGGGACAGATCTAGATCATGGAG
294 A A V G Y G S S K G L D Y I I V K N S W
1321 TAGCCGCAGTCGGATATGGATCGTCAAAGGGCTAGATTACATCATCGTGAAGAATTCCT
314 G P K W G E K G Y I R M K R N T G K P E
1381 GGGGACCAAAGTGGGGTGAAAAAGGCTACATACGAATGAAGAGAAACACAGGAAAACCAG
334 G L C G I N K M A S Y P T K K K *
1441 AAGGCCTATGTGGGATCAACAAAATGGCTTCATATCCTACTAAAAAGAAGTAAcacagca
1501 gggttgacaacatagatatttcctttctttctgatagctccttcattgatcaaaactttctt
1561 aatccttgttggtactcttatcaataaaatcagatagaaaattcacatccctcaaactct
1621 tgtgctattaaaaagtagataatggcattacacatgtaacaagaatgcagacaagagga
1681 tgtgcatgcttatgtgcacaacaatcatgtttctaacagagagtcaataaaataaaga
1741 tgcagatgagcatgaatagcaaaatgaagttctatataatcagggtttatttaacttcttt
1801 attaattcaatgatcataaattagtagtaatacatgcaagtttaaggtgacaaggacct
1861 aaaacatgtgtatattgatgattggetgctaatttaatggtttaacgatagagcagaaca
1921 agtatcttaactcaaaacctcattgtacttgaaaacacagtcataaagccggcctccaag

1981 aaactgggcaacctctggatgtttcacc

Supplementary File S3. The gene model for *RcXBCP3*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1 ccacgttttttaggcagccaatccaataaaaagaaaatacacgcccatactaaaaaggcaa
1                               M N F L S A
61 gattcccttttatttctccgagagaaaaataagaaaaagaaaaATGAACTTCTTTCTG
7  L F L I T L L F F N L S I S S F S S S S
121 CTCTTTTCTTGATCACTCTCCTCTCTTTAACCTTTCTATTTCTTCCTTTTCTTCTTCTT
27  D I S K L F E S W T K E H G K T Y T S K
181 CTGATATCTCTAAGCTCTTTGAGAGTTGGACCAAAGAACATGGAAAAACATACACTTCAA
47  E D K L Y R F K I F E E N Y E F V K K H
241 AGGAAGATAAGTTGTATAGATTCAAAATCTTCGAGGAAAATTATGAATTTGTAAAGAAAC
67  N S Q G N S S Y T L S L N A F A D L T H
301 ATAACAGCCAAGGTAATTCTTCTTATACCTTTTCTCTCAATGCCTTTGCTGATCTCACCC
87  H E F K A S R L G L S A F S T S G K L S
361 ACCATGAATTCAAAGCTTCTCGCTTGGGCTCTCTGCTTTTCAACCTCCGGGAAGCTTT
107 R R N F P L H D F V G D V P I S I D W R
421 CCCGTCGGAATTTTCCATTGCATGATTTGCTTGGCGATGTTCTATATCAATTGATTGGA
127 K K G A V S Q V K D Q G N C G
481 GAAAGAAAGGGGCTGTTTCTCAGGTCAAAGATCAAGGGAATTGCGgtgcgtgcatcctttt
541 atttggattcctttatcttttttatgcttaatttttgatgagttttgaaactactctagt
601 ctttctttatattgatccatgagcatcgtgcatacttagttttgaaactcttgtgtgatc
661 tttattatcatttattagtagtaatctcttaactctgtttttccttctactattaaact
721 tgggaatcatgtaagaattgcagccatcttattttattagatttcgtaatatagtctagta
781 cctcagcatgctgtataaccggcatttgattttgattgttttgcataatgaaatatacatct
841 ctctatatttcttaacgtattccaaaaggatttgccacctttataattttaaatcc
901 tcatcaaagagatggtggcttgtgaaaaatgctttcatttatagatttgttttaccttga
961 aagttttctccttttttagttttgattgcagtgaagttcagttcttcacaattttttctca
```

1021 tgcactatatttttcttctgatattcttagttaagacttccatcttgactgaaatcatca
1081 agttagtactattttcttggtgcgttcttagttatcaagactttcactttgactgaaatca
1141 tcaagttctacaattgactgtaatctggttaaattcattctactatgactaaatgatatat
1201 atatggccataattgtagctgtttttgactggtaaattgcagatacatctatttcatgg
142 A C W S F S A T
1261 tgcttaagaataaattgtttactgtattacgcagGTGCTTGTGGTCGTTCTCAGCCACA
150 G A I E G I N K I V T G S L V S L S E Q
1321 GGAGCTATTGAAGGCATAAATAAGATTGTTACTGGATCTCTCGTTAGCCTCTCGGAGCAG
170 E L V D C D R S Y N N G C E G G L M D Y
1381 GAGTTAGTTGATTGCGACAGATCTTACAATAATGGTTGTGAGGGTGGACTAATGGACTAT
190 A Y Q F V I E N N G I D T E E D Y P Y Q
1441 GCCTACCAGTTTGCATAGAAAACAATGGGATTGACACTGAGGAAGATTATCCATATCAA
210 A R E K T C N K E K
1501 GCACGGGAAAAGACCTGCAACAAGGAAAAGgtaattattttatttccatttgccttctt
1561 cacctataaactctcaattcagttcccactcttccgtgttcacatcattgtctatgttga
220 L K
1621 ttcactctctgtttctcgtttcattccggtagtcctaatttatacacattgcagTTGAAA
222 R H V V T I D G Y T D V P Q N N E K E L
1681 AGGCATGTGGTAACCATTGATGGATACTGATGTGCCCAAACAATGAGAAAAGAGTTA
242 L K A V A A Q P V S V G I C G S E R A F
1741 TTGAAAGCGGTTGCAGCTCAACCTGTAAGTGTCGGTATATGCGGCAGTGAGAGAGCATT
262 Q L Y S K
1801 CAGTTGACTCAAAGgtagtttttgatagttaatatcaatgaccatttactatagacag
267 G I
1861 caaaaaaaaaattggtagtgatttttaagtggtttaattatttgtatgatagGGAATT
269 F T G P C S T S L D H A V L I V G Y G S
1921 TTCACTGGTCCATGTTCAACTTCTTTGGATCATGCTGTATTGATTGTAGGATATGGGTCA
289 E N G V D Y W I V K N S W G T H W G I N

1981 GAAAATGGTGTAGATTATTGGATTGTGAAGAACTCATGGGGAACACACTGGGGAATAAAT
309 G Y M Y M L R N S G N S Q G L C G I N M
2041 GGTTATATGTACATGCTACGCAACAGCGCAATTCTCAAGGGCTCTGTGGTATCAACATG
329 L A S F P V K T S P N P P P P A P P G P
2101 CTGGCTTCATTTCCAGTGAAGACCAGCCCTAATCCCCCTCCTCCAGCTCCACCAGGACCC
349 T K C D L F T R C G E G E T C C C T R R
2161 ACTAAATGTGATCTTTTCACTCGATGCGGAGAAGGGGAAACTTGCTGTTGTACTCGCCGC
369 I F G L C F S W K C C E L D S A V C C K
2221 ATTTTTGGATTATGCTTTTCGTGGAAGTGTGCGAATTGGATTGAGCTGTGTGTTGCAAG
389 D G L H C C P H D Y P V C D T K R N M C
2281 GACGGGCTCCACTGTTGCCCCATGACTATCCAGTTTGCACACAAAAAGGAACATGTGC
409 L K
2341 CTTAAGgtcagcatattctctgccttcaatctgctagctgtttgattgtgtgatgcacat
2401 aaaactatcaaagctagggcatgattatactgtcgtttgaatttcttctgaacattaaatt
2461 gcttccttgacattcttccttttatTTTTgtaacttcaaagatcttactctctggtggaa
2521 atgatagaattgcatgggttgctttttgtttatcccacgtgatgttcaatactgtggat
2581 aaaaccacaaatttgatagagttgatttatcatttttagctaaaatctatctgatgttggt
2641 attctataaagtgaattaaagcttgaattaaagtttgcttccagtagactgatactgct
411 F P G N A T R M E T V A K K S S S
2701 tatgcttgcagTTCTCTGGTAATGCCACAAGAATGGAAACAGTTGCAAAGAAAAGCTCCT
428 G M F G S W N S L L E G
2761 CTGGGATGTTTGGCAGTTGGAATCCCTTCTTGAAGgttgattctgtaaaaccattttc
2821 atttattctggcgagtgatttaattttttcagaaaaacccttgtagaagttgcatgttt
2881 cacaattgaaggaggagactttctagatttagagcgtaaaattttgtgtacaccccaga
2941 agttgtattgtttagcttgtgtgtagagagattttctggtgcatggcattgttttacaat
3001 ggggggatggatttaagaaggaattaaggaggtgagcttggcattgtttcttttcta
3061 gaggcatgaatttaagaattgattaggattttccttatcctagatataagactgcttt
3121 tagcatatgtctttcgtatgtctgtacgtatattacgtctcccagctgcttacctgcat

3181 ctttatttgtgtgcatcaattgcttgaatattagatataaagatatectatcagctagtt
440 T A I A K G
3241 ttgatatgtaagcgagtagataagcaaacctctatgaacagGTACAGCTATTGCAAAAGG
446 H P R I A G V I V C Q D N I I S F R H Q
3301 CCACCCTAGGATTGCTGGAGTAATTGTGTGCCAAGACAATATCATATCCTTCAGGCACCA
466 N *
3361 AAATTAAaacacacgcaaaaaaaaaatgaaaaagaacaccagaaaagaggtgagaacatca
3421 ttctgcctggataattaggggttgctgcagggaccttgatgatttctctctagacctgctc
3481 atgggcatgaaaaacaagccctgtgcacctagacagaactcttttaacaactggcagtg
3541 tgaatgtgaattaccataagagtcceaattggaactgttgacttcttcagtatcgaa
3601 ctcttcacccaaatattatcaggcactctgtccagttctgccattgatgatgagtctctgc
3661 cattgatgatcgtcgtttatcaaatttgtaaactgccaagtagtaaatgagaatatga
3721 tgctgttcagcctgagacataattcaatcattcacataactcttcattaa

Supplementary File S4. The gene model for *RcXBCP3L*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1                               M G S Q K I
1 ggtagatcaaagaaatcatcaccggaacccaatttgaggaaaATGGGTCCCAAAAGAT
7  Q F L I I F L L V G P L T C L S F T L P
61 CCAATTCCTGATCATTCTTGCTGTAGGTCCACTAACATGCCTATCTTTACCTCCC
27  D E Y S I V G N D L H E L L S E E R V K
121 AGATGAATACTCTATAGTGGGCAATGACCTTCATGAGCTGCTATCTGAGGAACGTGTCAA
47  E L F Q Q W K E K H R K V Y K H V E E A
181 AGAACTCTTTCAGCAATGGAAAGAAAAGCACAGAAAGGTTACAAGCATGTAGAGGAAGC
67  E K R L E N F R R N L K Y V V E K N Q K
241 AGAAAAGAGGCTAGAAAACCTTCGAAGGAACTTGAAGTATGTAGTGGAGAAGAATCAGAA
87  K K N L G S A H T V G L N K F A D M S N
301 GAAGAAAAACTTGGGCTCAGCGCATAACAGTTGGATTGAATAAGTTTGCTGATATGAGCAA
107 V E F R Q K Y L S K V K K P I K K R N N
361 CGTGGAGTTTAGGCAAAAGTATCTTTCAAAAGTGAAAAACCCATCAAGAAACGGAACAA
127 N L M T S R Q R N L Q S C V A P S S L D
421 TAATTTAATGACCAGCAGGCAGAGAAATTTACAATCATGCGTTGCTCCTCAAGCTTGGA
147 W R K K G V V T P V K D Q G D C G
481 TTGAGAAAGAAAGGAGTTGTTACTCCAGTTAAGGATCAAGGAGATTGTGgtatgtttct
541 ttaatttctttctttgttattttctttataattatattatctgttttattttcagttgc
601 tgttgtattggacttagtggataactaattaatttgaaaaaagaattcaaatc
661 ttttgaatgcaacgaagatcataataaattaagtaataagattccttgaatgtcatgttgac
721 agataatcttaatttttaatggaatatccaaagatttagtctgtttttattagtaaaagtt
781 ttattttaaattgcaatgctgctttataagtaatttcatataattattatgagaaagagag
841 tatctaattgcttaaatttaaggttgcttagatgaataaagatggtttataatttcagat
901 ctaaaggaaagaagagtctggtttgactggttaggtacgcactagggttgaatcagtgctc
```

961 tattagattccttgctggggtgtgtataaaaagtaatttgagtgaaagtaataataa
1021 taataattaatcggatgggtgtatctgtaagactttttaactgtaaatthaagaatattat
1081 ttttagaaatttcgtttctttataattaagtgattattacactgttttaatttaa
1141 aaaacatttataaagtgaagtaacaaaagaacagacatttaattgcaaatttaagcat
1201 tttagttttagtaattttatttttagactttatcattaagtgattattactgtttttagtt
1261 taaaaaacatttataagcgaaagataataactgctcaattggctatatgcttccatcc
1321 acctgcatatTTTTGTGTGGTCAAAAAGGAAAACATTTCAAAAAAAGGTCCAAC TAGA
1381 tatagagaaattctaaagaaaaagcagatatgtttctcattgcacctaaagtatatatta
1441 ataagtatgagttgtttggataaaataaatttttaaacttttttgataagttatgt
1501 agtttatgatttgataaatttttttagataaaataaaatacattaaaaaattttacta
1561 ttaaaaaatatcaatttctagagatttaattttttactttaatcaactattattatca
1621 tattaaatatatttgctttaacaattcaaaaatattaccgtgtgattgtcaagttggtct
1681 cgtcagaccaatgacttcttaaatcacaatttgacaattaattaataaaattatgaatta
1741 gtttctaaactattaaacatgtgaaaacataataataataataataataaacct
1801 gattaaagtgaaaatagtttcttcttgctatccaaagtttttaacagtaaaacacat
1861 ttactgctatacattataaccacatttttcttttaactaagaaaaatacattaaaaat
1921 cataaaacagaacaagtacagtttaatactttatctacattattaattttatgaagttaa
1981 aaattacagtttaaatcttaagtgacaagaaaatttgatgtggaaagtaaacgttacc
2041 ttttgcatcacctatgttattgcttttatcatgagtatacagtttaatactttatctaca
2101 ttaggtgaattattaattttatgagcctctttatataatctcgtaattattgggatttcat
2161 gaactaaaaaaataaaaagtaatagaaatatctatttgctttattttttaatcaaattta
2221 caatattttttattttattatgttaaggtattaaaaaatatgatttttttaaaaattag
2281 aattttttatttcttttgtttgttttgagcaaaataattaaaaataaccgaaaaataat
2341 tatactataaatttgnnn
2401 nnn
2461 nnn
2521 nnn
2581 nnnnnnnnnnnnnnnnnnattatactataaatttgaaaaaggagacagtattttaaaag

2641 aaaattatacagtaaaaacaatattctttttattgattttaaggtattttaaatatTTTTG
2701 aaaaaattcctaaaaaacatacagcttaaatttttgaaatcaccggattgatcctggcag
2761 ctcaacgaacaagtctaataacacttcagcatcctaactaaattattgaaacttgcaagt
2821 accaaaatactaaattagtggtgctcctgcttcattttaactaaaactgatcaatattcg
2881 gatattcttactcgtccacaagcttattaacagtttcatttttttaagtgtttgtttt
164 S C W A F S
2941 aggtgttttgtaattgaggtccttggttggttaattttgcagGAAGCTGTTGGGCATTCTC
170 S T G A I E G I N A I V T G D L V S L S
3001 TTCCACTGGAGCAATCGAAGGAATAAATGCCATTGTCACCGGAGACCTCGTTAGCCTTTC
190 E Q E L M D C D T T N Y G C D G G Y M D
3061 TGAACAAGAACTTATGGATTGTGATACAACCAATTACGGTTGTGATGGTGGCTATATGGA
210 Y A F E W V I N N G G I D T E I D Y P Y
3121 CTACGCTTTTGAATGGGTTATAAACAATGGTGGGATTGATACAGAAATCGATTATCCTTA
230 T G V D G T C N I A K
3181 TACAGGTGTGGATGGTACCTGCAACATCGCCAAGgttcgtttccttcgtgttctatata
3241 acccactagaagatgattataatatcttttcttcgtcttacttgtggtgttcttgtgttt
241 E E T K V V S V D G Y E D V A E S D
3301 gatcagGAAGAAACAAAAGTTGTATCCGTTGATGGATACGAAGATGTAGCTGAATCAGAT
259 S A L L C A T V Q Q P I S V G I D G S A
3361 AGTGCTCTCTTGTGTGCTACTGTTACAGCAGCCTATTAGTGTGGTATCGATGGTTCTGCA
279 I D F Q L Y T S
3421 ATAGATTTTCAACTGTACACAAGTgtaagtcctcttttcattttatgagtagcaccctaaa
3481 aaaaaaaaaaaaaaatcaaaagatacttaagaagtgattatgacatgaaagatgatgtg
287 G I Y N G S C S D N P N D I D H A
3541 atttaaatcagGGAATCTACAATGGTAGTTGCTCTGATAATCCAAATGACATTGATCATGC
304 V L I V G F G S E D G E D Y W I V K N S
3601 TGTTTTGATCGTTGGATTTGGCTCTGAAGATGGTGAAGATTATTGGATCGTGAAGAATTC
324 W G T E W G I N G Y F Y L L R D T N L P

3661 ATGGGGCACAGAATGGGGAATAAACGGATATTTCTATTTATTAAGAGACACTAATTTGCC
344 Y G V C A V N A M A S Y P T K E S S A P
3721 ATATGGTGTGGTGCAGTCAATGCTATGGCCTCTTATCCCACTAAAGAATCCTCTGCACC
364 S P T S P P S P P S P P P P P P P P T P V
3781 ATCCCCGACTAGTCCTCCATCACCACCATCACCCCTCCACCTCCACCCCAACACCAGT
384 P P P P S P S P S E C G D F S Y C P T D
3841 GCCCCACCTCCCTCACCTTCACCAAGTGAATGTGGCGACTTCTCTTATTGTCCAACCTGA
404 E T C C C L Y E F F D F C L V Y G C C P
3901 CGAGACATGTTGTTGCCTTTATGAATTCTTTGATTTCTGCTTAGTCTACGGTTGCTGTCC
424 Y E N A V C C T G T E Y C C P S D Y P I
3961 GTACGAGAATGCTGTTTGCTGCACTGGCACAGAATATTGTTGCCCCAGTGATTACCCAAT
444 C D I K E G L C L Q
4021 TTGTGATATTAAGAAGGGCTCTGTCTCAGgtgagaactcatagccattaatatctgag
4081 aattgcctctttctagcgtgaataacgagctatttagttcaaaatacttgcaagttacgaa
4141 ttattaagccattattcaatcagttgattgaatggttaacctttctttttccaaaattgt
454 N Q G D Y L G V A A T K K H M A K
4201 gtatgcaacagAACCAAGGAGATTACTTGGGAGTAGCTGCAACGAAGAAGCATATGGCTA
471 H K L P W S K L E E S K R E R T Y Q P L
4261 AACACAAGCTTCCCTGGAGCAAGCTAGAAGAAAGCAAGCGAGAGAGAACATACCAACCTC
491 M W K R N P F A A I R *
4321 TCATGTGGAAGAGGAACCCCTTTTGCTGCAATTCGCTGAaagaagaaaagctgatggccccg
4381 ttaagtaattgaaggtcttcgcaattattttaattaaatttaccgcataggcagacattc
4441 agtaatcctgctacaattattggtttatgaacattaaacgaatgtagcaagaaaaactag
4501 aagtaaagaagaaaaggggaaattgagtatggcgttctcttgatgatcgactggatgga
4561 aactggatttgtagtaagacttctttcatgtatctttctcatggcttacagetcatttca
4621 agccatgctctgtattggtgtccaggacttgtaataacttcagtatatggttttaaaat
4681 aagagattctgtttggaagtttacttgatttcaatttcagaaaactgatctttac

Supplementary File S5. The gene model for *RcPAP3*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

1 M A L S L E K K L A I A L L V V F S T W
1 **ATG**GCTCTATCACTTGAAAAGAAACTCGCGATTGCCTTGTGGTGGTTTTCTCGACATGG
21 A S Q A M A R Q L I N E D A L V E K H E
61 GCATCTCAAGCCATGGCCCGACAGTTGATCAATGAAGATGCCCTCGTTGAGAAGCACGAG
41 Q W M A R H G R T Y Q D S E E K E R R F
121 CAATGGATGGCTCGCCATGGGCGCACTTATCAGGACAGTGAAGAAAAAGAGAGACGGTTT
61 Q I F K S N L E Y I D N F N K A S N Q T
181 CAGATATTCAAGAGCAACTTAGAGTACATTGACAACCTCAACAAGGCCCTCCAATCAGACT
81 Y Q L G L N N F A D L S H E E Y V A T Y
241 TACCAGTTAGGTCTCAATAATTTTGCAGACTTATCCCATGAAGAATACGTTGCTACTTAC
101 T A R K M P V L P T A N T T T K T T L F
301 ACTGCACGCAAGATGCCCGTACTCCCTACAGCAAACACCACAACCAAGACAACGCTATTC
121 N S D D I N Q V E V P E S I D W R D H G
361 AATAGTGATGATATTAACCAGGTTGAGGTTCCGGAGAGTATTGACTGGAGAGACCATGGC
141 A V T P I K N Q Y Q C G
421 GCTGTCACCCCTATTAAGAACCAATACCAATGTGgtaacatctcatctctctttacgtta
481 tatatatatataaaattaagtgggaattattaatgaattcaacgggtttatttagttgag
541 aagtgaatttgcttatacaaaaatagtaaaatatatgtaaaactttctaaaagaatcactg
601 acttggagatgtcaatctagtatttgaaatgggtaccaactaaataaagagtttctgtct
153 C C W A F
661 ttctttgactgattaccctccttatatattgtgtattgcacgcagGGTGTGCTGGGCAT
158 S A A A A V E G I V A N G V S L S A Q Q
721 TTTCAGCGGCGGCAGCCGTGGAAGGGATCGTAGCTAATGGTGTCTCACTGTCAGCTCAAC
178 L L D C V S D N Q G C K G G W M N N A F
781 AATACTAGACTGTGTGAGTGATAATCAAGGTTGTAAGGGGGTTGGATGAATAATGCTT

198 N Y I I Q N Q G I A L E T D Y P Y Q Q M
841 TTAATTACATAATACAAAACCAGGGCATAGCCTTAGAAAACGGATTACCCGTATCAACAAA
218 Q Q M C S S R M A A A Q I S G F E D V T
901 TGCAACAAATGTGCAGTTCAAGAATGGCTGCAGCCCAGATAAGTGGTTTTGAAGATGTAA
238 P K D E E A L M R A V A K Q P V S V T I
961 CTCCAAAAGACGAGGAAGCCCTGATGAGGGCAGTAGCAAAGCAACCTGTATCGGTCACCA
258 D A T S N P N F K L Y K E G V F T A A G
1021 TTGACGCTACTTCTAATCCAAACTTCAAATTGTACAAAGAGGGAGTGTTCCTGCAGCAG
278 C G N G H S H A V T L V G Y G T S E D G
1081 GCTGCGGGAATGGACATAGCCATGCTGTGACCTTAGTTGGGTATGGGACAAGTGAAGATG
298 T K Y W L A K N S W G E T W G E S G Y M
1141 GGACCAAGTATTGGCTAGCCAAGAATTCATGGGGTGAGACTTGGGGAGAAAGTGGTTACA
318 R L Q R D I G L E G G P C G I A L Y A S
1201 TGAGGTTGCAGAGAGATATTGGTTTAGAAGGAGGTCCTTGTGGCATTGCATTGTATGCTT
338 Y P T I N *
1261 CGTATCCTACCATTAATTAGaccgctctatctaagaatattacaacatgcataaaataa
1321 aagaaaagac

Supplementary File S6. The gene model for *RcCTB1*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1 cttgctttgtactatctatctgtatctaccaataaattaatcaaagcctcattaaccgaa
61 cacactttccttattcgacgtttactatatagctccttctgacctgttgatactgcggc
1 M A A S I L S S F A L L L
121 agttgctttggagaagcaaaaATGGCAGCTTCTATCTTATCCTCCTTCGCTCTTTGCTG
14 F L V A L S S F H S R
181 TTTCTTGGCTCTCTCCTCTTTTCACTCCCGGgtacctattattggtgttattagatt
241 tcttttgtctggttataaagtttctgatcgtgctgattaaattagttttggatttgaa
301 attaagaatgacgagattgttttaactcatcgatttcgatttctgtagcttgaatttaag
361 agttgatgaagtattttcaacggattgatttcgattttgtacctatgtcatctgtgccga
421 aagtaaaatttcatgtaaccgaagttttatttgaactaaaaatgtaaaggaacgtacct
481 tttatgctactgtgatgatcacctctctaataagcactgtaaagccacatgaatttatct
541 tatataggagtgttagagtttggatttatgttgcgctttttgttgcgtgttgttct
601 tgtgcttgctcctttgcataccctgataagaggaggatgttcattcttgaatgttcttca
661 aatcttgttctgagttttgattcttttggagttactgttgactttttaacatectaa
25 V I S T E L D S K L K L N
721 atgctgttgacactgtcaacagGTGATCTCAACAGAACTAGATTCCAAGCTGAAGCTGAA
38 S R I L Q
781 CTCTAGGATCCTTCAGgtatTTTTCTTgctgccgatcattggaatccaaaagctttatt
43 E S I I K
841 ctgtagagcaagtcatttaattgcttttctgacttcgacgagcagGAGTCAATAATTAA
48 K V N E N P D A G W E A A M N P Q L S N
901 GAAAGTAAATGAAAACCTGATGCTGGATGGGAAGCTGCCATGAATCCTCAACTTTCAA
68 F T
961 CTTCACTgtgagctttaactcatcagaggccaacatcttttctaagtccatttgtctac
70 V G Q F K Y L L G
```

1021 agttaacacatacttctatattttgctgcgtagGTTGGCCAATTCAAGTATCTTCTTGGA
79 A K P T P K K E L M G V P M I S H P K T
1081 GCCAAGCCAAC TCCAAAAAGGAACTGATGGGTGTTCCCTATGATAAGTCATCCAAAAACC
99 L K L P K E F D A R T A W P H C S T I G
1141 TTAAAGTTGCCAAAAGAATTTGATGCAAGAACAGCTTGGCCACATTGCAGTACCATTGGA
119 K I L D
1201 AAAATTCTAGgtcagctgctttctttctataatataattttcaattttcttcttcttttt
1261 cttgaggtgctgatgtttccactatgtatgccaaatgcttgattctggatcaatttcctc
123 Q
1321 atcaaagATCAGgtaaaatctggttctattgttttctcagtcactcactcaacatgagt
124 G H C G S C W A
1381 tctacatattctgagttcgtctgtatatgacaatccagGGTCATTGTGGATCCTGTTGGG
132 F G A V E S L S D R F C I H F G M
1441 CATTGGTGCTGTTGAATCACTATCTGATCGATTTTGTATCCATTTTGGCATGgtaacta
1501 gctagctatattcactaaatctcctgttaaacttcactagctgtgttgcattacaaac
149 N I S L S V N D L L A C C
1561 atatcttttgtttggccacagAATATATCATTGTCTGTCAATGATCTCTTGGCATGCTGT
162 G F L C G D G C D G G Y P M Y A W R Y F
1621 GGCTTTTTGTGTGGAGATGGTTGTGATGGAGGGTATCCGATGTATGCATGGAGATACTTT
182 V H H G V V T E E
1681 GTCCACCACGGGGTTGTCACTGAGGAGgtacaaatttctgtgtttgcctgtttccttgg
1741 tcttctgttttcccctgcactggagattgttaaattggaagctaagggttcttghtaatgt
1801 catgtttctttttgatgacttcatgaatataattatttgatgcaatggetttcttctt
191 C D P Y
1861 gtttttaaatatttcttcttaaaaatctgcaaactgtttacatctgacagTGTGACCCATA
195 F D N I G C S H P G C E P G F P T P K C
1921 TTTTGATAACATTGGCTGTTCCACCCTGGTTGTGAGCCTGGATTTCCAAC TCCAAAGTG
215 V R K C I D K N Q L W R Q S K H Y S V N

1981 TGTGAGAAAGTGCATTGATAAGAATCAGCTCTGGAGGCAGTCAAAGCATTACAGCGTTAA
235 A Y R I S S D P H D V M A E V Y K N G P
2041 TGCATATAGAATTAGTTCTGATCCCCATGATGTTATGGCTGAAGTTTATAAGAATGGACC
255 V E V S F T V Y E
2101 AGTAGAGGTTTCTTTCACCTGTTTATGAGgtaacatatcaattcgtgtcttctgtttgct
2161 tgcattaacttctttgtatgtgaatataagcttatttcttttgctgttctatccctgctg
2221 taatccttaaccataattaagtgagctgaagaaatcttgagccagcttcacatgtagatg
2281 taggaagaacctcaatgggagttttcatatattggatgcctgaaatatgattatcttttc
2341 catgtatattagatactcgaacttttcatgataatagctcttggtagaattacttttttg
264 D F A H Y K S G V Y K H
2401 aagctactgctttgcgttctacagGATTTTGCTCATTATAAATCTGGAGTTTATAAGCAC
276 I T G E V M G G H A V K L I G W G T S D
2461 ATTACAGGTGAAGTAATGGGAGGTCATGCAGTTAAGCTTATTGGGTGGGGACTTCTGAT
296 N G E D Y W
2521 AATGGAGAGGATTATTGGgtatgcgtgattagctaactttttgcagatatttagctat
302 L L A N Q
2581 gaccttataaggatgttggtataattgtgcacttttcttcttgcagCTTCTTGCAAACC
307 W N R G W G D
2641 AGTGGAATAGAGGCTGGGAGATgtagttcctataataagtttcttgctgatgaatcac
2701 gctactttcccttcttttgttttctcttcatecttttattcccctgctetcaaattttc
314 D G Y F K I R R G T N E
2761 acaaatttttatttcgaatgacagGATGGTTACTTTAAGATAAGGAGAGGAACAAATGAG
326 C G I E D D A V A G L P S A R N L D L V
2821 TGTGGTATTGAAGATGATGCGGTTGCTGGTTGCCTTCAGCCAGAAATCTTGATCTTGTT
346 R E V A S M D A L E D A F A *
2881 AGAGAGGTTGCCAGCATGGATGCTCTTGAAGATGCTTTTGCT**TGA**tagtagatcattgta
2941 cttgcaagttctgtaatttacgcatcagtcacagagatatgttatgggtgtgt
3001 ctgctgttcagcatgaaacgaaacatatttatgaatccatgctgtgttggtgcttata

3061 tcaactagacatttgttgtgaatgcaagaaaggaacaaatctgtaaataatgtaattgtc
3121 acacatagtataagagtcaactgctttctgctttcacattgcttttacattaggcactta
3181 tgaaatgtttggaaggatgtccaaattaataagatagtcctgctttatgaagaaaatggg
3241 aagtttctgcagaaatagctctaattgtcagtagcataagccctaaccgagagatggaa
3301 tctaatacgtggtacgccaacagaaggagtacatgtggcggctagcaagcgagtcaaat
3361 actgtttaaggataaggtagtgactcttccttttgctttttcccgagggtgggggcta
3421 tataggaaaacattgtaggtaataatattttgttttgcacttacagccagcatgttca
3481 ttgtaggcttagtattccgttttataatgttatattttcccctcatcattccaa

Supplementary File S7. The gene model for *JcSAG12H8*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1 gaaactgaattttggttgccatatacacaactaacgggattaactcaacataagac
61 tgcaatttcagacatctcttgtaaggaagccaatttatatggtctataaatacatgacag
121 ctctttaaggactagtgtgccaaaaatacacatatacctacattatatactctctagtctt
1 M A S I L E N K L V F V T
181 tacttctctctccaccaagaagaATGGCTTCAATTCTGAAAACAACTGGTCTTTGTGA
14 V L V L G L F A S Q A F A R P L Q N E T
241 CAGTGTTGGTGCTGGGGCTCTTTCATCTCAAGCCTTGCACGTCCTCTACAAAATGAAA
34 M K E R H E M W M A K Y G R V Y K D S A
301 CCATGAAGGAGAGGCATGAAATGTGGATGGCTAAATATGGACGTGTTTATAAAGACAGTG
54 E K E K R F N I F K N N V E F I E S F N
361 CAGAGAAGGAGAAACGCTTTAACATATTCAAGAATAATGTGGAGTTCATTGAATCTTTTA
74 K D G N K L Y K L D I N G F A D L S N E
421 ATAAGGATGGGAATAAGCTTTACAAGCTAGATATCAATGGATTTGCAGACTTAAGTAATG
94 E F K A S R N G Y K R S S I A K S S E T
481 AAGAGTTTAAGGCCTCTAGAAATGGATATAAAAGATCTTCTATTGCTAAGTCATCTGAGA
114 L S F K Y E N V T A V P T S M Y W R N K
541 CGCTGTCGTTTAAGTATGAAAATGTCACCGCTGTCCCAACTTCCATGTACTGGAGGAACA
134 G A V T P I K D Q G Q C G
601 AAGGAGCTGTACTCCCATCAAGGACCAAGGCCAATGCCgtaaggaacacaatgtttctt
661 atttttcttctactatgtactagctatagtaaattaatctaataatgtattttcccttt
147 C C W A F S A V A A M E G I
721 tggaacacacaaatgtagGATGTTGCTGGGCATTTTCTGCTGTGGCTGCCATGGAAGGAA
161 T K L S T G K L I S L S E Q E L V D C D
781 TTACAAAGCTCTCAACAGGAAAATTGATTTCTTTCTGAGCAAGAACTGTTGATTGTG
181 T S G E D Q G C E G G L M D D A F E F I
841 ACACAAGTGGAGAAGATCAAGGCTGCGAAGGGGTCTTATGGATGATGCATTTGAATTCA
201 K K N G G L T T E A N Y P Y Q G T D D S
901 TAAAAAAAATGGAGGACTAACAACTGAAGCCAATTACCCTTACCAAGGAACTGATGATT
```

221 C N K R K A V D H A A K L T G Y E D V P
961 CTTGCAATAAAAAGAAAGGCAGTTGATCACGCAGCAAAGCTTACTGGCTATGAAGATGTGC
241 A N S E D A L L K A V A N Q P V S V A I
1021 CTGCTAATAGTGAAGATGCCTTATTGAAGGCAGTAGCCAACCAACCAGTCTCTGTAGCCA
261 D A S G S A F Q F Y S G G V F T G D C G
1081 TTGATGCAAGTGGCTCTGCTTTCCAGTTCTACTCAGGTGGAGTATTTACAGGAGATTGCG
281 T E L D H G V T A V G Y G T S S D G T K
1141 GAACTGAACTCGACCATGGAGTTACTGCAGTTGGGTATGGCACAAGCAGTGATGGAACCA
301 Y W L V K N S W G T S W G E N G Y I R M
1201 AGTATTGGTTAGTGAAGAACTCATGGGGAACTTCTTGGGGTGAAAATGGATACATTAGAA
321 E R D I D A S E G L C G I A M E P S Y P
1261 TGGAAAGAGATATTGATGCTAGCGAAGGTCTTTGTGGAATTGCGATGGAACCTTCTTACC
341 T A *
1321 CAACTGCATAAattaaaattaagtggcagaaaagaaaatatgggtcaatcttaaaatataa
1381 atacattgtccaaaa

Supplementary File S8. The gene model for *JcRD21B*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1 ctttaccgtaacctgtctttgtacatgtagtaatccgattctaaaaatataaatttctc
61 cctatatatatacgtcaaaagcacaagcaaaatcacccatataccatccgacccaac
1      M A S P G R S L P F F I L C G F L
121 cctacagccATGGCTTCTCCTGGAAGATCTCTCCCTTTTTTCATCCTATGTGGGTTCTT
18 F F S S L A L D M S I I D Y N I R H G Q
181 TTTTTCTCTCTTTAGCGCTGGACATGTCGATTATTGATTACAATATCAGGCATGGACAA
38 N P E E R T D A N V R R I Y E M W L V K
241 AACCCAGAAGAAAGAACAGACGCTAATGTTTCGGAGAATCTATGAGATGTGGCTCGTCAA
58 H G K A Y N A L G E K E K R F Q I F K D
301 CATGGGAAAGCTTATAATGCTTTGGGCGAGAAGGAGAAGCGGTTTCAGATCTTTAAAGAT
78 N L K F I D E H N S V N R S Y K L G L N
361 AATCTTAAATTCATCGATGAGCATAACTCGGTTAATCGGAGTTATAAGCTCGGATTGAAC
98 R F A D L S N E E Y R A M F L G T R M E
421 CGGTTCCGGATCTGAGTAATGAGGAGTACCGGGCGATGTTCTTGGGTACAAGGATGGAA
118 R K N R L G A S M G T E S R Y L Y K E G
481 AGGAAAAATAGATTGGGGGCATCGATGGGACTGAAAGTAGGTATTTGTATAAAGAAGGT
138 D D L P E K V D W R E K G A V V P V K N
541 GATGATTTGCCGAAAAGGTTGATTGGAGAGAAAAAGGGGCTGTTGTTCTGTCAAGAAT
158 Q G Q C G
601 CAAGGACAGTGTGgtaagtttgtttcttagcgatttttccttttctttttgagttttg
661 ttggttttgattgcttgtgaagtgggacctggaagcatggcatgttcttccaaggatttt
721 cctggaaactagcgacaccaattgatttttgatttgtagtgttatctccttagcaggga
163                                                                 S C
781 tatttcttgatttttcgtgaaagaaaataaattgtttgttttatttattgaagGGAGTT
165 W A F S T V A A V E G I N Q I V T G D L
```

841 GCTGGGCTTTCTCAACTGTTGCTGCAGTAGAAGGGATTAACCAGATTGTAAGTGGTGATC
185 I S L S E Q E L V D C D R S Y N Q G C N
901 TAATCTCTCTATCAGAGCAGGAGTTGGTAGATTGTGATAGATCATATAACCAGGGATGCA
205 G G L M D Y A F E F I I Q N G G I D T E
961 ATGGAGGTCTTATGGACTATGCCTTTGAATTTATTATCCAAAATGGTGGTATTGATACTG
225 Q D Y P Y E A V D N V C D P Y R
1021 AACAAAGATTACCCATATGAGGCTGTTGATAATGTTTGTGATCCATACAGGgtatgctaaa
1081 tttattttacttaatcctttgttctgtagttacattggtatattgatcgttctaat
241 K N A K V V T I D G Y E
1141 tatttgttcttgcacatcttttagAAAAATGCTAAGGTAGTTACTATTGATGGATATGA
253 D V P E N D E K S L K K A V A N Q P V S
1201 AGATGTTCTGAAAATGATGAGAAATCCTTGAAGAAGGCTGTGGCAAATCAGCCAGTTAG
273 V A I E A G G R A F Q L Y Q S
1261 TGTTGCCATTGAAGCTGGTGGCAGGGCTTTCCAACCTTACCAATCGgtaattactcttta
1321 ttgccaaatgaaattatctcatttgcagtttaacaagagcagtgttatattttgtgagta
1381 ctagttatacagaaactattataattcctattcttttggtttggcataattttgttaact
1441 caactcaatatatgttatgcatcagaatcatataacaagttgaatatataaagtacaat
288 G V
1501 ctttgctggcattgattgttttgtacttgtctttaattttacttctcaatgcagGGTGT
290 F T G R C G T Q L D H G V A A V G Y G T
1561 TTTTACTGGTAGATGTGGTACACAGCTTGACCATGGTGTGGCTGCTGTTGGATATGGCAC
310 E N G V D Y W L V R N S W G P D W G E S
1621 AGAGAATGGTGTAGATTACTGGTTGGTGGAGAACTCATGGGGTCTGATTGGGGTAAAAG
330 G Y I K L E R N L A T T T T G K C G I A
1681 TGGTTATATCAAGTTGGAGCGTAATTTGGCCACCACCACAACCTGGAAAATGTGGAATTGC
350 V E A S Y P T K K G P N P P N P G P T P
1741 AGTGGAGGCTTCATATCCAACCAAGAAGGGTCCAAATCCCCCTAATCCTGGCCCAACCCC
370 P S P V N P P P P K P S T Q C D D Y F S

Supplementary File S9. The gene model for *JcCEP2*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1 ctttgattgtctgaagtcacgttttctttctgttgcatagccgtaaccttttctgctca
61 atttttggttaaagctctcaaaaagcgtataaccctcatctcttctccatggetatata
121 acaattcaacacatccttgcaatttcattgctaagccaacaagaaaatcaagaattccc
1
M G A R N I
181 gactttgtaacccttttcttgatttatagtaggttagatattATGGGTGCAAGAAATA
7 I L F A V S L V L V F K I A E S L D Y Q
241 TAATCTTGTTTGCCGTTTCATTAGTGCTAGTTTTTAAGATAGCTGAAAGTTTGATTACC
27 E A D L A S E E S L W N L Y E R W R S H
301 AAGAAGCAGACTTAGCGTCGGAAGAGAGTCTTTGGAAGTGTACGAGCGGTGGAGAAGCC
47 H T I S R S L T E K K Q R Y N V F K E N
361 ATCACACAATCTCTCGTAGCCTGACAGAGAAAAAGCAGCGATAACAATGTGTTCAGGAAA
67 L K H I H N V N Q K D K P Y K L R L N K
421 ATTTGAAGCATATTCATAACGTTAACCAGAAGGATAAGCCTTATAAATTGAGATTGAATA
87 F A D M T N H E F L Q H Y G G S K V S H
481 AGTTTGCGGATATGACTAATCATGAGTTCTTGCAACATTATGGAGGCTCTAAGGTTAGCC
107 F R M L H G S R Q K T G F S H H N T D L
541 ACTTCAGGATGCTACATGGGTCTCGTCAGAAAAGTGGGTTTAGCCACCACAACACTGATC
127 P P S V D W R K K G A V T G V K D Q G K
601 TTCCACCTCCGTTGATTGGAGAAAGAAAGGAGCAGTCACTGGTGTAAGGATCAAGGCA
147 C G
661 AATGTGgtgagtttctcaagtcataaatttcagctttcagtatTTTTtagtaataattttg
721 aacttataatcactagatctttaaatgacactcttaactcttggtttataaatatttaa
149
S C W A F S S V A A V E
781 atcaatgatttttatgctatacagGTAGCTGTTGGGCATTTCAAGTGTGGCAGCAGTGG
161 G I N K I K T G E L V S L S E Q E L V D
```

841 AGGGTATTAACAAAATCAAAACAGGAGAATTAGTTTCTCTATCAGAGCAAGAGTTGGTTG
181 C S S A N H G C D G G L M E Q A F K F I
901 ATTGCAGCTCTGCCAACCATGGCTGTGATGGAGGATTAATGGAACAGGCATTCAAGTTCA
201 K K I G G I T A E T T Y P Y T A R D G S
961 TTAAGAAAATTGGTGAATAACCGCAGAGACTACCTATCCTTATACTGCTAGAGATGGAT
221 C D S N K
1021 CCTGTGACTCAAACAAGgttaactgagtttaatatattgatctataactcaactttttcagt
226 M
1081 gtacgtatatgtttctatgcacaagtgaggattatatacttttcttggtgtttagAT
227 N A P V V Q I D G Y E M V P E N D E N A
1141 GAATGCTCCTGTTGTGCAAATTGATGGATACGAAATGGTGCCAGAAAACGACGAGAATGC
247 L M K A V A N Q P V S I A I D A G G K D
1201 ATTGATGAAAGCTGTTGCTAACCAACCAGTATCTATTGCAATAGATGCAGGCGGAAAAGA
267 M Q F Y Y E
1261 TATGCAATTCTATTATGAGgcaagttttttattcaataggaaactgtacagctgaatca
1321 aacagaccctatgatatagttttttcattccctgaaaattttgtttctcttgttgttga
273 G V F T G D C G T E L N H G V A L V G Y
1381 gGGAGTATTCAGTGGAGATTGCGGGACAGAGCTAAACCATGGGGTAGCACTTGTGGGATA
293 G T T L D G T K Y W I V K N S W G S D W
1441 TGGCACAACACTAGATGGCACAAAGTATTGGATAGTGAAGAATTCATGGGGAAGTGATTG
313 G E N G Y I R M E R G I D A E E G L C G
1501 GGGAGAGAATGGTTACATAAGAATGGAAAGAGGGATTGATGCAGAAGAGGGACTGTGTGG
333 I T L E A S Y P I K N S S D N K K S P D
1561 TATAACTTTGGAGGCTCTTACCCTATCAAGAATAGCTCAGACAACAAGAAATCACCCGA
353 S R R D E L *
1621 TTCTCGCAGAGATGAACTCTAAatttgattatatactatctatatatatatgcctcaat
1681 aacttaaggacctaaaaggagagtttttaatttaattaccgcttttctcttttagtatta
1741 attaagttgaagttggatttctttgcatctcttttgttatatatagttcaatgatttga

1801 attccaataaaaccaatgctttctctc

Supplementary File S10. The gene model for *JcXBCP3L*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1  gaggatatggttgcctctgactggttaggccgggagtaagaccatgtttcgaaacttataaac
1      M T D Q N P I G H C L V I K R K M S
61  actataATGACAGATCAAAACCCAATCGGCCACTGCTTAGTGATAAAGAGAAAAATGAGT
19  F Q K L G L A I V F L V L S L V P L T C
121 TTCCAAAAGTTGGGATTAGCAATCGTTTTTCTGGTTTTATCTTTAGTGCCACTAACCTGC
39  F S F S L P S E Y S I V G N D L H E L L
181 TTCTCTTTTAGCTCCCAAGCGAGTACTCTATAGTTGAAAACGACCTTCATGAGCTGCTT
59  S H E R V I E L F Q Q W K L E H K K I Y
241 TCTCACGAACGGGTGATCGAGCTGTTCAACAATGGAAACTGGAGCACAAAAAGATTTAC
79  K H A E E A E K R L E N F R R N L K Y V
301 AAGCACGCAGAAGAGGCAGAAAAGAGGCTCGAAAATTTTCGAAGGAATTTAAAGTATGTG
99  I E K N A K N A E M G T G F K V G L N K
361 ATAGAGAAGAATGCGAAGAATGCAGAGATGGGTACAGGATTTAAAGTTGGATTGAACAAA
119 F A D L S N E E F K Q T Y L S K V K K P
421 TTTGCAGATTTGAGTAATGAGGAGTTTAAGCAAACGTATCTGTGCGAAAGTGAAAAACCT
139 I S K K K N T M M T T G Q R N L Q A C E
481 ATCAGTAAAAAGAAAAATACAATGATGACTACCGGACAGAGAAAATTTGCAGGCTTGTGAG
159 A P S S L D W R K K G V V T P I K N Q G
541 GCGCCTTCTAGCTTGGATTGGAGGAAGAAGGGAGTTGCTACTCCAATAAAGAACCAAGGA
179 N C G
601 AATTGTGgtaattttctataatttctccctcttatatatattacatatgtacatatatgc
661 atacttgatgtatatgtgaaggcataagacaccataaatgatatgacaggttttagataa
721 cttaaaatctaaagtaaaaagtactttgatgctaagagaataaaaaaaaaacgtaaatat
781 attcataggtataggttgaagtatacaacatacttgagttaacatatgatttaacgt
841 tcataaaaaataattgaaagaaaagaaaaaaggcagatgcccttttattattcataaac
```

901 tggctttcctgttattattcacttaccagacgtaatcctagttaacattaggataataa
182 S
961 tgtgggaaattatcttcattttgggtgacttcaactagattgatatttgtttacagGAAG
183 C W S F S V T G A I E G I N A I V T G D
1021 TTGTTGGTCATTCTCTGTAAGTGGCGCCATAGAAGGAATAAACGCCATAGTTACTGGTGA
203 L I S L S E Q E L V D C D T T N F G C D
1081 CCTAATTAGCCTTTCAGAACAAGAACTGTAGATTGTGATACCACCAATTTCCGGATGTGA
223 G G Y M D Y G F E W V I N N G G I D T E
1141 TGGAGGCTATATGGACTACGGTTTTGAATGGGTTATTAACAACGGTGGGATTGACACAGA
243 A N Y P Y T G V N G T C N I T K
1201 AGCGAATTATCCATATACTGGTGTAAATGGGACCTGCAATATCACCAAGgtttctttctt
1261 atctaaaatcttattttacccatataccaatgcgaacatgatttaaaccttaactctgtt
259 E E M K I V T V D
1321 gtgtgtctcattaatgatctttttgtttgatcagGAGGAAATGAAAATTGTTACCGTTGA
268 G Y E D V A K S D S A L F C A T A Q Q P
1381 TGGGTACGAGGATGTAGCTAAATCAGACAGTGCCCTGTTCTGTGCTACAGCTCAGCAACC
288 V S V G M D G S A E D F Q L Y T S
1441 TGTTTCTGTGGGTATGGATGGTTCTGCAGAAGATTTCAATTGTATAACAAGTgtaagtct
1501 ttttgttcgtttcatgagtacaattaaaaagataattactatagaagtgtttgataaaat
305 G I Y D G N C S S N P D D
1561 ggaaatgatgtgatatgatcagGGGATTTATGATGGTAATTGCTCCAGTAATCCAGACGA
318 I D H A V L I I G Y G S E G D E D Y W I
1621 CATTGATCATGCCGTTTTGATCATTGGGTATGGATCTGAAGGTGATGAAGATTATTGGAT
338 V K N S W G T D W G I D G Y F Y L R R N
1681 AGTGAAAAATTCGTGGGGTACTGACTGGGGTATCGATGGATACTTTTATTTAAGAAGGAA
358 T D L T Y G V C A V N A M A S Y P T K Q
1741 TACTGATTTAACATATGGCGTTTGTGCTGTTAATGCAATGGCTTCTTATCCTACTAAACA
378 S S A P S P T S P P S P L S P P P P P S

1801 ATCCTCTGCTCCATCTCCAACAAGTCCTCCTTCACCACTATCACCTCCTCCACCACCATC
398 L P P P P P P T P V P P P P P P S P S E
1861 ACTACCACCACCACCTCCTCCGACACCAGTGCCTCCACCTCCTCCTCCTCACCAAGTGA
418 C G D F S Y C P S G E T C C C F L E I F
1921 ATGTGGAGATTTCTTACTGTCCAAGCGGTGAAACATGTTGTTGCTTTCTTGAAATCTT
438 D I C W A Y S C C P F E N A V C C P D S
1981 TGATATCTGTTGGGCATATAGTTGCTGTCCTTTTGAGAATGCTGTATGCTGCCCTGATTC
458 L D C C P S D Y P I C D V V E G L C L Q
2041 ATTAGATTGCTGCCCTAGTGATTACCCTATTTGTGATGTTGTGGAAGGCCTCTGCCTCCA
478
2101 Ggtaagaatcgtcaatttaataactccccatttcacatatataagtcggttagagtaatt
2161 ttattgtctgtaaataaagtagtttttaaatattgtcaattcaaacagtagcattaat
478 N P G D Y L G V
2221 acaccttagttaattgtttattgaaattgtgtaacagAACCCAGGAGATTACTTGGGAGT
486 A S R K R H M A K H K L P W S K K E Q S
2281 TGCTTCCAGGAAGCGGCATATGGCTAAGCACAAGCTTCCCTGGAGTAAAAAGAGCAAAG
506 K R Q T Y Q P L V W K R N P F A A I R *
2341 CAAACGACAGACTTACCAACCTCTCGTGTGGAAGAGGAACCCCTTTCGCTGCAATTCGCTG
2401 Aaaaataatgtgatttttcttattaatttaaagttgtttcattgcaactttttgaatgatc
2461 ctattacaaaaaaacttatgttttataaatattattaggggaaaatttgcaatgtgtgt
2521 tccatttgattatggttttttgatgtttgaattaaagattggtttcccttattatatatg
2581 tcttgcagctttttgcaagccaaaaattcagtattgttgtcaaggccttgtaataaacct
2641 tactaaatgggttttttaattaaaataaacaatcatttttccatttgaaa

Supplementary File S11. The gene model for *JcTH11*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1  ttccccagaaaaatggacagcaaagaagacatgaatgaaacaagaagtccaactgaaa
61  taacttgtgcaaaagtatgatctacgtaaataactgcattccgtaagcaaccca
121 tatatcattcctgggatgtactaaaattgtcatttataagctcccattctccatagacat
1      M E S H I N L K K A G L I L I S L
181 cgtcagtagccATGGAGTCGCATATTAATTTAAAGAAAGCTGGCTTGATTCTTATATCTC
18   C I L L I H T M A D L E I H S L P R Y N
241 TCTGCATTCTATTGATACATAACAATGGCAGATTTAGAAATCCATAGTCTGCCAAGATATA
38   P K A M R K R Y D R W L K H H G R K Y H
301 ACCCAAAGCCATGAGAAAGAGGTATGACAGGTGGCTGAAACATCATGGTCGAAAATATC
58   N K D E Y Y L R F G I Y Q S N I Q F I D
361 ATAACAAAGACGAATATTATCTGCGGTTTGAATATATCAATCAAACATTCAGTTCATTG
78   Y I N A Q K L P Y E L R D N K F A D L T
421 ACTACATCAATGCTCAGAAATTACCATACGAACTCAGAGATAACAAGTTTGCAGACTTGA
98   N D E F K S I Y L G F Q A N K H G R K K
481 CTAATGATGAGTTTAAATCTATCTACCTGGGTTTTTCAGGCTAATAAACACGGTAGAAAGA
118  P S H E H G N S S D L P T S V D W I K E
541 AGCCTAGCCATGAGCATGGAACTCCTCAGACTTGCCAACCTCAGTAGATTGGATAAAGG
138  G A V T P V K D Q G Q C G
601 AAGGCGCTGTAACCCCGTTAAAGATCAAGGTCAATGTGgtatgaagctttgttcctt
661 ttctttttcattttttcttcctgagaaatgtgagccaagtgcactttttagaagtgc
721 ttttgttttatgccaattgtcaattaatcattttgttttattctaagaattttggcatg
151                                     S C W A F S A V A A
781 acaagatctaaagtgacagttctatacagGAAGTTGTTGGGCATTCTCTGCAGTGGCAGC
161  V E G I N K I K T G K L V S L S E Q E L
841 TGTTGAAGGCATTAACAAAATCAAGACAGGAAAATTAGTATCTCTGTCAGAACAAGAGCT
```

181 V D C D T H N D S Q G C N G G F M E T A
901 TGTAGACTGTGATACTCACAATGATAGCCAGGGATGTAATGGTGGATTTCATGGAAACAGC
201 F T Y I Q K Q G L S I E D D Y P Y E G K
961 GTTTACATACATTCAAAAGCAGGGGCTCTCCATTGAAGATGATTATCCCTATGAAGGAAA
221 D G N C H K N K N K N Q T V T I S G Y E
1021 AGATGGCAACTGCCACAAAAACAAAAATAAAAATCAGACAGTACTATTAGTGGTTATGA
241 A V P A N D E K S L Q T A V A K Q P V S
1081 AGCAGTACCTGCCAATGATGAGAAAAGCCTACAACTGCAGTTGCCAAGCAACCTGTATC
261 V A I D A G G F S F Q F Y S K G I F S G
1141 TGTTGCAATTGATGCCGGTGGTTTTTCATTTCAATTCTACTCAAAAGGAATCTTCAGTGG
281 F C G T Q L N H G V T A V G Y G E T D G
1201 CTTCTGTGGAACACAACCTCAACCATGGAGTCACAGCAGTCGGCTACGGAGAACTGATGG
301 K K Y W L V K N S W G T E W G D S G Y I
1261 TAAAAGTACTGGCTCGTTAAAAATTCATGGGGTACAGAGTGGGGTGATTCTGGCTATAT
321 K L R R G S K G K K G I C G I A M E P S
1321 AAAACTGAGGCGTGGTTCCAAGGGCAAGAAAGGTATCTGTGGCATTGCCATGGAGCCTAG
341 Y P V K E Q E *
1381 CTACCCTGTCAAGGAACAGGAATAAgaatcattttatatgtgcttggctgatcctggctt
1441 tagggtgacttgacaatctcttaacaagatgtcttccatattagtaattacatacaaatc
1501 caagtaaaagatattgctcatggaaattaatttaacaatagttaagctttgcaggaagta
1561 tggtagggaaaatgcatgccataagcataagaagatgggtaattccttcgatcaaaatg
1621 gatattacatcattgattggagatccccaattcttgatataaaatgcttaacataacttt
1681 catctatctcatttgtcatttggcaaagcccgaaaaagaaccgcataccaatcacaagt
1741 ctttcgagcaactaaaaggtgtaataatctttactttgttctggagtgaaaatgagactta
1801 aagtctgtaggtatatttaaggtttgaatgtatcagaatatggatgagcctcaagagcca
1861 tcattgcaacaaaaaatttgcacatgatttagcatcaattgactgattgaggttaaggctt
1921 tgcttactttgtttatataattgaaccaaatgtgtaagtacaacattaatcagtttagca
1981 gaaatcagaattgtaaactaac

Supplementary File S12. The gene model for *JcSAG12H1*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1                               M T K K Q S
1 caaagattccttctgtatagttatagatactctcttgatactATGACTAAAAACAAAGC
7 K S I F L V F V L N I L T I W A T H T V
61 AAATCCATATTTCTGGTATTTGTGTTGAACATATTAACCATATGGGCTACACATACGGTT
27 C R P L N E E Y M L K R H E E W R A Q H
121 TGTCGTCCTCTTAACGAAGAATACATGTTAAAGAGGCATGAAGAATGGAGAGCCCAACAT
47 G R V Y K D T A E K Q K K Y L V F K D N
181 GGACGTGTCTACAAAGACACAGCAGAGAAAACAGAAAAAATACCTGGTTTTTAAGGACAAC
67 L E R I E S F N N G V D R G Y K L G L N
241 CTTGAACGTATTGAATCCTTTAACAATGGTGTGGACCGTGATACAAGCTAGGACTCAAC
87 K F A D L T D E E F Q A M H L G Y K S P
301 AAATTTGCAGACTTAACAGATGAGGAATTTTCAGGCCATGCACCTTGTTACAAGAGCCCA
107 P S K L M S T S K S R S F R Y R N V T S
361 CCCTCCAAATTAATGTCCACTTCAAAGTCCAGATCCTTTAGGTACAGAAATGTAACTTCT
127 V P T T I D W R K A G A V T P V K D Q G
421 GTGCCAACTACTATAGATTGGAGAAAGCCGGTCTGTGACCCCTGTCAAAGATCAAGGC
147 S C G
481 TCCTGCGgtgagtgctgatataatcctaataatgtaaggcagaagtagacctaggatttag
541 aaccttaacttgcatactatgagacagactcaaccactgtcctattgctaaataaatac
601 acatgaaattgaacaatgattttgatcttaaaaatcatgctattgatgcataacagaaaa
661 gagcaatccccgcaccacaaatgaaaattcgtttctaataaatcattttttaatgtttaa
150           S C W A F S A V A A M E G I T K L
721 tgcatagGAAGTTGCTGGCATTCTCAGCAGTGGCAGCAATGGAAGGGATCACAAAACTC
167 K T G N L I S L S E Q E L V D C D V A G
781 AAAACTGGCAATTTAATATCTTTATCAGAGCAAGAGCTCGTAGATTGCGACGTAGCAGGT
187 E D Y G C D G G F M D T A F Q Y I L K N
841 GAGGATTATGGTTGTGACGGAGGTTTCATGGACACTGCTTTCCAATATATCCTAAAAAAT
207 G G L T S E A N Y P Y Q G E D G I C S K
```

901 GGAGGTCTCACGAGTGAGGCTAATTACCCCTACCAAGGAGAAGATGGCATCTGCAGCAAG
227 K K T A T S M A K I T G Y E D V P S N S
961 AAGAAGACAGCAACTTCTATGGCTAAAATAACTGGATATGAAGATGTGCCATCTAACAGT
247 E K A L L Q A V A N Q P V S V A I D A S
1021 GAAAAGGCTCTCTTGCAAGCTGTGGCAAACCAACCAGTTTCTGTTGCTATTGATGCTAGT
267 G Y D F R F Y S S G V F K G D C P T D L
1081 GGGTATGACTTCAGATTTTACTCTTCTGGTGTTTTTAAAGGGGACTGTCCTACCGATCTC
287 N H A V T V I G Y G S S S D G T K Y W L
1141 AACCATGCTGTTACTGTAATTGGGTATGGTAGTAGCAGTGATGGTACTAAGTATTGGTTG
307 L K N S W G T G W G E N G Y M R M Q R E
1201 CTAAAGAATTCATGGGGCACCGTTGGGGTGAGAATGGGTATATGAGGATGCAAAGGGAA
327 I S A N E G L C G I A M K A S Y P T A *
1261 ATTAGTGCAAACGAAGGCCTCTGTGGCATTGCCATGAAAGCTTCGTATCCAACCTGCT**TGA**
1321 aattgaacaaggagaagcataatggcaataaatcctatttgatgtgtagtatcagaaa
1381 tttataggctgtaaaatgttctgctatctgtgctcattctatctttcatgcagttgtata
1441 atttcagcaatatgtgtttactatccatgcataaaggttcattttattagctatcaagt
1501 accaac

Supplementary File S13. The gene model for *JcSAG12H2*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The transcribed untranslated regions, including 5' UTR, intron and 3' UTR sequences, are marked with lowercase letters. The start and stop codons are marked with bold letters.

```
1                               M T K K Q S
1 caaagattccttctttatagttatagatactctcttgatactATGACTAAAAACAAAGC
7 K S I F L V F V L N I L T I W A T H T V
61 AAATCCATATTTCTGGTATTTGTGTTGAACATATTAACCATATGGGCTACACATACGGTT
27 C R P L N E E Y M L K R H E E W R A Q H
121 TGTCGTCCTCTTAACGAAGAATACATGTTAAAGAGGCATGAAGAATGGAGAGCCCAACAT
47 G R V Y K D T A E K Q K K Y L V F K D N
181 GGACGTGTCTACAAAGACACAGCAGAGAAAACAGAAAAAATACCTGGTTTTTAAGGACAAC
67 L E R I E S F N N G V D R G Y K L G L N
241 CTTGAACGTATTGAATCCTTTAACAATGGTGTGGACCGTGATACAAGCTAGGACTCAAC
87 K F A D L T D E E F R A M H L G Y K S L
301 AAATTTGCAGACTTAACAGATGAGGAATTTTCGGGCTATGCACCTTGTTACAAGAGCCTA
107 P S K L M A T S K S R S F R Y R N V T S
361 CCCTCCAAATTAATGGCCACTTCAAAGTCCAGATCCTTTAGGTACAGAAATGTAACTTCC
127 V P T T I D W R K A G A V T L V K D Q G
421 GTGCCAACTACTATAGATTGGAGAAAGGCTGGTGTGTGACCCTTGTCAAAGATCAAGGC
147 S C G
481 TCCTGCGgtaagtgctgatataatcctaataatataaggcagaagtagacctaggatttag
541 aaccttaacttgcatactatgagacagactcaaccattgtcctattgctaaataaatac
601 acataaaaacaatgattttgatcttaaaaatcatgctattgatgcataacagaaaagagc
661 aatccccgtaccagaaatgaaaattcatttctaataaatcattttttaatgtttaatgca
150   C C W A F S A V A A M E G I T K L K T
721 tagGATGTTGCTGGGCATTCTCAGCAGTGGCAGCAATGGAAGGAATCACAAAACCTAAAA
169   G K L I S L S E Q E L V D C D I A G E D
781 CTGGCAAGTTAATATCTTTATCAGAGCAAGAGCTCGTAGATTGTGACATAGCAGGTGAGG
189   Y G C D G G F I D T A F Q Y I L K N G G
841 ATTATGGTTGTGACGGAGGTTTCATAGACACTGCTTTCCAATATATCCTAAAAAATGGAG
209   L T S E A N Y P Y Q G E D G I C S K K K
```

901 GTCTCACGAGTGAGGCTAATTACCCCTACCAAGGAGAAGATGGCATCTGCAGCAAGAAGA
229 T A T S T A K I T G Y E D V P S N S E K
961 AGACAGCAACTTCTACGGCTAAGATAACTGGATATGAAGATGTGCCATCTAACAGTGAAA
249 A L L Q A V A N Q P V S V A I D A S G Y
1021 AGGCTCTCTTGCAAGCTGTGGCAAACCAACCAGTTTCTGTTGCTATTGATGCTAGTGGGT
269 D F R F Y S S G V F Q G D C T T Y L N H
1081 ATGACTTCAGATTTTACTCTTCTGGTGTCTTTCAAGGGGACTGTACTACCTATCTAAACC
289 A V T V I G Y G S S S Y G T K Y W L L K
1141 ATGCTGTTACTGTAATTGGGTATGGTAGTAGCAGTTATGGTACTAAGTATTGGTTGCTAA
309 N S W G T G W G E N G Y M R M Q R E I S
1201 AGAATTCATGGGGCACCGTTGGGGTGAGAATGGGTATATGAGGATGCAAAGGAAATTA
329 A N E G L C G I A M K A S Y P T A *
1261 GTGCAAATGAAGCCTCTGTGGCATTGCCATGAAAGCTTCGTATCCAAGCTGCT**TGA**aatt
1321 gaacaggagaagcataatggcaataaatcctatttgatgtgtagtatcaggaaatttat
1381 aggctgcaaatgttctgctatctgtgctcattctatctttcatgcagttgtataatttc
1441 agcaatatgtgtttaatatccatgtataaaggttcatttatttttagctatcaagtaccaa
1501 t

Supplementary File S14. The gene model for *JcSAG12H7*. The coding region is marked with uppercase letters, above which are its deduced amino acids. The intron sequences are marked with lowercase letters. The start and stop codons are marked with bold letters.

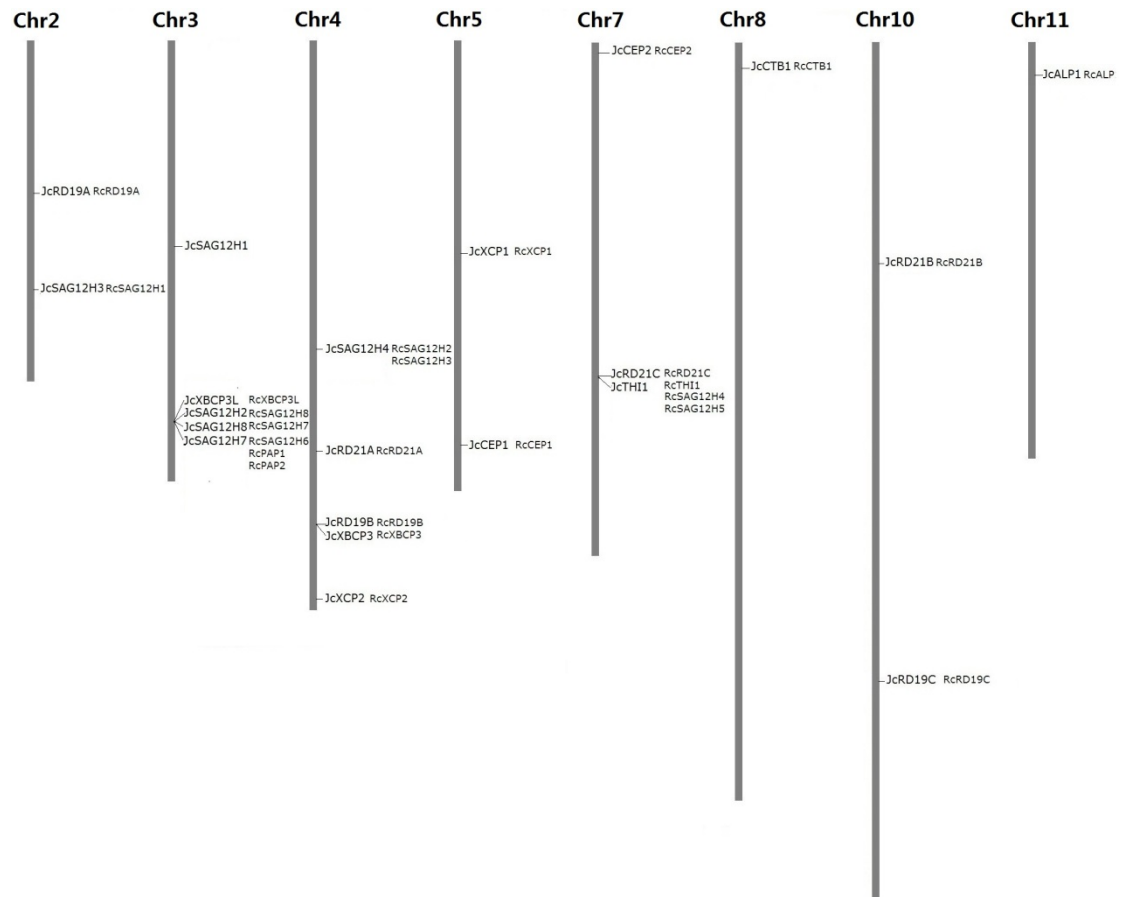
1 M A V V V L L E N K L I F V A V L V M G
1 **ATGGCTGTAGTAGTACTACTT**GAAAACAAACTAATCTTTGTTGCAGTGT**TAGT**GATGGGG
21 L C A C Q G W S R S L Y E A S L E E R H
61 TTGTGTGCTTGCCAAGGCTGGTCACGTTCCCTCTATGAAGCTTCGTTGGAGGAGAGGCAT
41 E I W M A Q H E R V Y K D A T E K E M R
121 GAAATTTGGATGGCCCAACATGAACGTGTTTATAAAGACGCGACAGAAAAGGAGATGCGC
61 F Q I F R E N V K F I E S F N K A G N K
181 TTTCAGATATTCAGGGAGAATGTGAAATTCATTGAATCTTTCAACAAGGCTGGGAATAAG
81 S Y K L G I N A F T D L T N K E F R A S
241 TCTTATAAGCTAGGGATTAATGCATTACAGACCTACTAATAAGGAATTCGTGCATCG
101 R N G Y K R S S Y S S S S G T N S F K Y
301 CGAAATGGATACAAGAGGTCTTCTTATTCAAGCTCATCAGGAACAACTCATTCAAATAT
121 E N V T T I P S S L D W R S K G V V T P
361 GAAATGTCACTACAATTCCATCTAGCTTGGATTGGAGAAGCAAAGGAGTCGTTACACCC
141 I K D Q G Q C G
421 ATCAAAGACCAAGGTCAATGTGgtaagtaacttggacttcatgtaaacaagtcagggttt
481 caaatcctgatagtccatccctttaagaattgtgataaaagaaaaaatagaaaagga
149 C C W
541 aaaaagaaagaaacctcttaacattatatgaattgcatgcaaaaatgcagGATGTTGTTG
152 A F S A V A A M E G I T K L S T E K L I
601 GGCATTCTCTGCGGTGGCAGCTATGGAAGGGATCACAAACTCTCAACTGAGAAATTGAT
172 S L S E Q E L I D C D T S G I D Q G C E
661 CTCCCTTTCTGAGCAAGAATTGATTGACTGTGATACAAGTGAATAGACCAAGGCTGTGA
192 G G L M D D A F E F I I Q N N G L T T E
721 GGGTGGTCTTATGGATGATGCCTTTGAATTCATTATTCAAATAATGGCCTTACGACTGA
212 A N Y P Y Q A E D G T C N T E K A A N H
781 AGCAAATTATCCTTACCAGGCAGAGGATGGAAGTTCGCAACACCGAGAAAGCAGCAAATCA
232 A A T I T S Y E D V P E N N E E A L R M
841 TGCTGCCACAATACAAGTTATGAAGATGTGCCAGAAAACAATGAGGAGGCCTTACGCAT

252 A V A N Q P V S V A I D A G E S A F Q H
901 GGCAGTGGCAAACCAACCAGTTTCTGTTGCCATTGATGCTGGTGAATCTGCCTTCCAACA
272 Y S S G I F T G D C G T E L D H G V T V
961 CTA~~CT~~CGAGTGGAAATATTTACTGGAGATTGTGGA~~ACT~~GAATTAGACCACGGTGTCACTGT
292 V G Y G T S D D G T K Y W L V K N S W G
1021 AGTCGGTTATGGGACAAGTGATGATGGGACTAAGTATTGGTTAGTAAAGA~~ACT~~CATGGGG
312 T S W G E D G Y I R M Q R D I D A K E G
1081 AAC~~CAG~~CTGGGGTGAGGATGGATACATTAGAATGCAAAGAGACATTGATGCTAAAGAAGG
332 L C G I A M Q P S Y P T A *
1141 CCTATGTGGAATTGCCATGCAGCCTTCCTATCCA~~ACT~~GCATAA

Supplementary Fig. S1. Detailed information of 15 motifs identified in this study.

	Logo	E-value	Sites	Width
Motif 1.		1.3e-1107	49	30
Motif 2.		1.5e-881	36	37
Motif 3.		1.6e-875	44	31
Motif 4.		1.8e-638	44	21
Motif 5.		1.0e-565	49	21
Motif 6.		1.2e-597	47	29
Motif 7.		3.4e-445	47	21
Motif 8.		1.7e-342	48	15
Motif 9.		6.1e-390	28	29
Motif 10.		2.8e-273	39	21
Motif 11.		7.3e-234	48	11
Motif 12.		5.5e-194	8	50
Motif 13.		9.9e-141	26	15
Motif 14.		2.6e-110	5	50
Motif 15.		3.6e-062	17	11

Supplementary Fig. S2. Matched positions of 25 RcPLCP genes on physic nut chromosomes. The positions were based on the synteny analysis and only 8 matched chromosomes were shown. RcPLCP genes were shown behind that of physic nut.



Supplementary Table S1. Detailed information of 31 PLCP genes in *Arabidopsis*.

Subfamily	Gene name	Other names	TAIR ID	Chr	Nucleotide length (bp, from start to stop codons)		Intron NO.	AS ¹	AA	MW (kDa)	p I	GRAVY	iPSORT ²	Duplicate	Mode ³
					CDS	Gene									
RD21	<i>AtRD21A</i>	RD21	AT1G47128	Chr1	1389	2471	4	-	462	50.97	5.26	-0.446	S	AT1G47128	Beta
	<i>AtRD21B</i>		AT5G43060	Chr5	1392	2334	4	-	463	51.20	5.86	-0.457	S		
	<i>AtRD21C</i>	CP1	AT3G19390	Chr3	1359	1745	4	-	452	49.31	5.60	-0.368	S	AT4G11310	Proximal Tandem Transposed
	<i>AtRDL1</i>		AT4G36880	Chr4	1131	1489	3	-	376	41.64	7.54	-0.505	S		
	<i>AtRDL2</i>		AT3G19400	Chr3	1089	1369	3	Yes	362	40.29	4.77	-0.322	S		
	<i>AtRDL3</i>		AT3G43960	Chr3	1131	1507	3	-	376	41.54	5.08	-0.235	S		
	<i>AtRDL4</i>		AT4G11310	Chr4	1095	1725	3	-	364	39.92	5.69	-0.326	S		
	<i>AtRDL5</i>		CP2	AT4G11320	Chr4	1116	1492	3	-	371	40.71	5.58	-0.349		
<i>AtRDL6</i>	AT4G23520	Chr4	1071	1763	3	-	356	39.60	5.65	-0.378	S				
CEP	<i>AtCEP1</i>		AT5G50260	Chr5	1086	1258	2	-	361	40.71	7.15	-0.613	S	AT5G50260	Beta Tandem
	<i>AtCEP2</i>		AT3G48340	Chr3	1086	1336	2	-	361	40.37	5.99	-0.617	S		
	<i>AtCEP3</i>		AT3G48350	Chr3	1095	1619	2	-	364	40.97	5.87	-0.576	M		
XCP	<i>AtXCP1</i>		AT4G35350	Chr4	1068	1347	3	Yes	355	39.62	5.60	-0.449	S	AT1G20850	Gamma
	<i>AtXCP2</i>		AT1G20850	Chr1	1071	1330	3	-	356	39.71	5.16	-0.478	S		
XBCP3	<i>AtXBCP3</i>		AT1G09850	Chr1	1314	2028	4	-	437	48.07	6.02	-0.309	S		
THI	<i>AtTHI1</i>		AT1G06260	Chr1	1032	1137	1	-	343	37.77	7.51	-0.225	S		
SAG12	<i>AtSAG12</i>		AT5G45890	Chr5	1041	1460	2	-	346	38.23	7.98	-0.321	S	AT1G29080	Tandem Alpha Transposed Transposed
	<i>AtPAP1</i>		AT2G34080	Chr2	1038	1347	1	-	345	38.27	6.90	-0.350	S		
	<i>AtPAP2</i>		AT1G29090	Chr1	1068	1283	1	-	355	39.94	6.12	-0.363	S		
	<i>AtPAP3</i>		AT1G29080	Chr1	1041	1181	1	-	346	38.79	5.74	-0.295	S		
	<i>AtPAP4</i>		AT2G27420	Chr2	1047	1209	1	-	348	38.74	5.32	-0.351	C		
	<i>AtPAP5</i>		AT3G49340	Chr3	1026	1231	1	-	341	37.69	4.56	-0.307	C		
RD19	<i>AtRD19A</i>	RD19	AT4G39090	Chr4	1107	1501	2	-	368	40.42	6.75	-0.306	S	AT4G39090	Alpha Gamma
	<i>AtRD19B</i>		AT2G21430	Chr2	1086	1338	2	-	361	39.82	7.08	-0.252	S		
	<i>AtRD19C</i>		AT4G16190	Chr4	1122	1366	3	-	373	41.26	6.50	-0.252	S		
	<i>AtRD19D</i>		AT3G54940	Chr3	1104	1726	3	-	367	40.13	6.79	-0.198	S		
ALP	<i>AtALP</i>	AALP, ALEU, SAG2	AT5G60360	Chr5	1077	2109	7	Yes	358	38.96	6.26	-0.186	S	AT5G60360	Alpha
	<i>AtALP2</i>		AT3G45310	Chr3	1077	1770	7	Yes	358	39.54	6.22	-0.271	S		
CTB	<i>AtCTB2</i>	CATHB2	AT1G02305	Chr1	1089	2059	10	-	362	40.03	6.45	-0.217	S	AT1G02305	Tandem Alpha
	<i>AtCTB1</i>		CATHB1	AT1G02300	Chr1	1140	2089	9	-	379	42.26	6.74	-0.058		
	<i>AtCTB3</i>	CATHB3	AT4G01610	Chr4	1080	2081	10	Yes	359	39.42	5.79	-0.235	S		

¹ “Yes” represents genes containing alternative splicing isoforms; ² “S, M and C” represent signal peptide, mitochondrial targeting peptide or chloroplast transit peptide, respectively, which was predicted by iPSORT; ³ duplicated modes were determined based on the study of Wang et al. (2013).

Supplementary Table S2. List of orthologous groups of the PCLP gene family based on the analysis of representative plant species.

Subfamily	OG	<i>R. communis</i>	<i>I. cretacea</i>	<i>A. thaliana</i>	<i>M. esculenta</i>	<i>P. trichocarpa</i>	<i>C. papaya</i>	<i>T. cacao</i>	<i>V. vinifera</i>	<i>A. lyrata</i>	<i>B. oleracea</i>	<i>R. rapa</i>	<i>A. coarctata</i>	<i>G. sativa</i>	<i>A. trichopoda</i>	<i>S. moellendorffii</i>	<i>C. reinhardtii</i>	
RD21	I	<i>R-RD21A</i>	<i>J-RD21A</i>	<i>AR21A</i>	Manes.18G041100	Posit014G041100	evm.TU.supercontig_21.158	Thecc1HE003775	GSVVG0100940001	ALIG5250	Bo037029	Brara.F00502	-	LOC_Os09g57440	-	-	270344	Cce09.g407700
		<i>R-RD21B</i>	<i>J-RD21B</i>	<i>AR21B</i>	-	Manes.10G077200	Posit00G020100	evm.TU.supercontig_300.16	Thecc1HE007113	GSVVG01003179001	ALIG3640	Bo032703	Brara.A00157	-	-	-	-	-
		<i>R-RD21C</i>	<i>J-RD21C</i>	<i>AR21C</i>	-	Manes.13G119500	Posit00G041000	evm.TU.supercontig_55.102	Thecc1HE000063	GSVVG0102063001	ALIG3640	Bo032703	Brara.H01714	-	-	-	-	-
		-	-	-	-	-	-	-	-	-	ALIG32800	Bo018095	Brara.H02264	-	-	-	-	-
		-	-	-	-	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-
IV	-	-	-	<i>AR21C</i>	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-	
		-	-	-	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-	
		-	-	-	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-	
		-	-	-	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-	
		-	-	-	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-	
V	-	-	-	<i>AR21D</i>	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-	
		-	-	-	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-	
		-	-	-	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-	
		-	-	-	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-	
		-	-	-	-	-	-	-	-	ALIG32810	Bo013106	Brara.H02264	-	-	-	-	-	
CEP	I	<i>R-CEP1</i>	<i>J-CEP1</i>	<i>ACEP1</i>	Manes.06G13000	Posit01C090000	evm.TU.supercontig_4.28	Thecc1HE001457	GSVVG01007583001	ALIG24300	Bo015079	Brara.C02533	-	-	-	-	-	
		<i>R-CEP2</i>	<i>J-CEP2</i>	<i>ACEP2</i>	Manes.14G087000	Posit01S0487400	evm.TU.supercontig_33.75	Thecc1HE001054	GSVVG01000582001	ALIG24300	Bo015079	Brara.C02533	-	-	-	-	-	
XCP	I	<i>R-XCP1</i>	<i>J-XCP1</i>	<i>AIXCP1</i>	Manes.10G097200	Posit00G020100	evm.TU.supercontig_232.4	Thecc1HE003454	GSVVG0102383001	ALIG3810	Bo020880	Brara.A00297	AqcoeSG11800 LOC_Os01g73980	evm_27.TU.AmTr.v1.0_scaffold00022.170	-	-	-	
		<i>R-XCP2</i>	<i>J-XCP2</i>	<i>AIXCP2</i>	Manes.15G113300	Posit00G020100	evm.TU.supercontig_286.15	Thecc1HE003454	GSVVG0101342001	ALIG3810	Bo020880	Brara.H02309	AqcoeSG16420 LOC_Os01g01810	evm_27.TU.AmTr.v1.0_scaffold00022.173	-	-	-	
XBCP3	I	<i>R-XBCP3L</i>	<i>J-XBCP3L</i>	<i>AIXBCP3L</i>	Manes.18G019400	Posit00G022000	evm.TU.supercontig_698.4	Thecc1HE003454	GSVVG0100979001	ALIG20490	Bo036791	Brara.F00644	AqcoeSG18200 LOC_Os01g4220	evm_27.TU.AmTr.v1.0_scaffold00002.349	-	-	-	
		<i>R-XBCP3L</i>	<i>J-XBCP3L</i>	<i>AIXBCP3L</i>	Manes.16G080500	Posit01G066800	evm.TU.supercontig_7.76	Thecc1HE002966	GSVVG0102122001	ALIG20490	Bo036791	Brara.F00644	AqcoeSG18200 LOC_Os01g4220	evm_27.TU.AmTr.v1.0_scaffold00002.349	-	-	-	
TH	I	<i>R-TH1</i>	<i>J-TH1</i>	<i>AITH1</i>	Manes.13G139700	Posit01G066900	evm.TU.supercontig_209.7	Thecc1HE003574	GSVVG01020657001	ALIG16260	Bo023341	Brara.H02966	AqcoeG24200	-	-	-	-	
		<i>R-SAG12</i>	<i>J-SAG12</i>	<i>AISAG12</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.270	-	-	-	
SAG12	Ia	<i>R-SAG12H1</i>	<i>J-SAG12H1</i>	<i>AISAG12H1</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.271	-	-	-	
		<i>R-SAG12H2</i>	<i>J-SAG12H2</i>	<i>AISAG12H2</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.272	-	-	-	
		<i>R-SAG12H3</i>	<i>J-SAG12H3</i>	<i>AISAG12H3</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.273	-	-	-	
		<i>R-SAG12H4</i>	<i>J-SAG12H4</i>	<i>AISAG12H4</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.274	-	-	-	
		<i>R-SAG12H5</i>	<i>J-SAG12H5</i>	<i>AISAG12H5</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.275	-	-	-	
		<i>R-SAG12H6</i>	<i>J-SAG12H6</i>	<i>AISAG12H6</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.276	-	-	-	
		<i>R-SAG12H7</i>	<i>J-SAG12H7</i>	<i>AISAG12H7</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.277	-	-	-	
		<i>R-SAG12H8</i>	<i>J-SAG12H8</i>	<i>AISAG12H8</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.278	-	-	-	
		<i>R-SAG12H9</i>	<i>J-SAG12H9</i>	<i>AISAG12H9</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.279	-	-	-	
		<i>R-SAG12H10</i>	<i>J-SAG12H10</i>	<i>AISAG12H10</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.280	-	-	-	
		<i>R-SAG12H11</i>	<i>J-SAG12H11</i>	<i>AISAG12H11</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.281	-	-	-	
		<i>R-SAG12H12</i>	<i>J-SAG12H12</i>	<i>AISAG12H12</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.282	-	-	-	
		<i>R-SAG12H13</i>	<i>J-SAG12H13</i>	<i>AISAG12H13</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.283	-	-	-	
		<i>R-SAG12H14</i>	<i>J-SAG12H14</i>	<i>AISAG12H14</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.284	-	-	-	
		<i>R-SAG12H15</i>	<i>J-SAG12H15</i>	<i>AISAG12H15</i>	Manes.16G038400	Posit00G0485900	-	-	-	ALIG12660	Bo006318	Brara.B02856	AqcoeG32500 LOC_Os01g12660	evm_27.TU.AmTr.v1.0_scaffold00061.285	-	-	-	
RD19	I	<i>R-RD19A</i>	<i>J-RD19A</i>	<i>ARD19A</i>	Manes.14G288600	Posit00G0260000	evm.TU.supercontig_3257.1	Thecc1HE000619	GSVVG01024210001	ALIG11020	Bo008027	Brara.A00027	AqcoeSG18100 LOC_Os01g27030	evm_27.TU.AmTr.v1.0_scaffold00001.80	266883	-	-	
		<i>R-RD19B</i>	<i>J-RD19B</i>	<i>ARD19B</i>	Manes.18G018400	Posit00G0234000	evm.TU.supercontig_150.51	Thecc1HE003440	GSVVG0100987001	ALIG38000	Bo037199	Brara.A00297	AqcoeSG17600 LOC_Os01g36400	-	-	-		
		<i>R-RD19C</i>	<i>J-RD19C</i>	<i>ARD19C</i>	Manes.10G080900	Posit00G0228400	evm.TU.supercontig_53.142	Thecc1HE003315	GSVVG01016190001	ALIG36050	Bo045388	Brara.B03759	AqcoeSG13100 LOC_Os01g29760	evm_27.TU.AmTr.v1.0_scaffold00022.391	-	-		
		<i>R-RD19D</i>	<i>J-RD19D</i>	<i>ARD19D</i>	Manes.16G095100	Posit00G041700	evm.TU.supercontig_28.31	Thecc1HE007524	GSVVG01013078001	ALIG36970	Bo008941	Brara.B01397	AqcoeSG147600 LOC_Os01g28100	evm_27.TU.AmTr.v1.0_scaffold00004.285	439207	-		
		<i>R-RD19E</i>	<i>J-RD19E</i>	<i>ARD19E</i>	Manes.10G040700	Posit00G040700	-	-	-	ALIG1640	Bo004850	Brara.H03120	-	-	-	-	-	
		<i>R-RD19F</i>	<i>J-RD19F</i>	<i>ARD19F</i>	Manes.10G040700	Posit00G040700	-	-	-	ALIG1640	Bo004850	Brara.H03120	-	-	-	-	-	
		<i>R-RD19G</i>	<i>J-RD19G</i>	<i>ARD19G</i>	Manes.10G040700	Posit00G040700	-	-	-	ALIG1640	Bo004850	Brara.H03120	-	-	-	-	-	
		<i>R-RD19H</i>	<i>J-RD19H</i>	<i>ARD19H</i>	Manes.10G040700	Posit00G040700	-	-	-	ALIG1640	Bo004850	Brara.H03120	-	-	-	-	-	
		<i>R-RD19I</i>	<i>J-RD19I</i>	<i>ARD19I</i>	Manes.10G040700	Posit00G040700	-	-	-	ALIG1640	Bo004850	Brara.H03120	-	-	-	-	-	
		<i>R-RD19J</i>	<i>J-RD19J</i>	<i>ARD19J</i>	Manes.10G040700	Posit00G040700	-	-	-	ALIG1640	Bo004850	Brara.H03120	-	-	-	-	-	
ALP	I	<i>R-ALP1</i>	<i>J-ALP1</i>	<i>AALP1</i>	Manes.10G040700	Posit00G040700	-	-	-	ALIG1640	Bo004850	Brara.H03120	-	-	-	-	-	
		<i>R-ALP2</i>	<i>J-ALP2</i>	<i>AALP2</i>	Manes.10G040700	Posit00G040700	-	-	-	ALIG1640	Bo004850	Brara.H03120	-	-	-	-	-	
CTB	I	<i>R-CTB1</i>	<i>J-CTB1</i>	<i>AICTB1</i>	Manes.10G069000	Posit01G0409800	evm.TU.supercontig_57.55	Thecc1HE000487	GSVVG0102827001	ALIG1630	Bo011427	Brara.B00106	AqcoeSG28000 LOC_Os01g24850	evm_27.TU.AmTr.v1.0_scaffold00048.14	143272	Cce02.g119600		
		<i>R-CTB2</i>	<i>J-CTB2</i>	<i>AICTB2</i>	Manes.10G069000	Posit01G0409800	evm.TU.supercontig_57.55	Thecc1HE000487	GSVVG0102827001	ALIG1630	Bo011427	Brara.B00106	AqcoeSG28000 LOC_Os01g24850	evm_27.TU.AmTr.v1.0_scaffold00048.14	143272	Cce12.g498850		

“OG” represents “orthologous group”. “-” represents “Not Detected”.