

Supplementary Information for

Chromosome-level genome assembly of *Zizania latifolia* provides insights into its seed shattering and phytocassane biosynthesis

Ning Yan ^{1,#}, Ting Yang ^{1,#}, Xiu-Ting Yu ^{1,2,#}, Lian-Guang Shang ^{3,#}, De-Ping Guo ⁴, Yu Zhang ¹, Lin Meng ¹, Qian-Qian Qi ^{1,2}, Ya-Li Li ^{1,2}, Yong-Mei Du ¹, Xin-Min Liu ¹, Xiao-Long Yuan ¹, Peng Qin ⁵, Jie Qiu ⁶, Qian Qian ^{7}, Zhong-Feng Zhang ^{1}

Correspondence should be addressed to N.Y. (yanning@caas.cn), Q.Q. (qianqian188@hotmail.com), or Z.-F.Z. (zhangzhongfeng@caas.cn).

¹ Tobacco Research Institute of Chinese Academy of Agricultural Sciences, Qingdao 266101, China

² Graduate School of Chinese Academy of Agricultural Sciences, Beijing 100081, China

³ Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

⁴ Department of Horticulture, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China

⁵ State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Rice Research Institute, Sichuan Agricultural University, Chengdu, Sichuan 611130, China

⁶ Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University, Shanghai 200234, China

⁷ State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China

[#] These authors contributed equally: Ning Yan, Ting Yang, Xiu-Ting Yu, Lian-Guang Shang

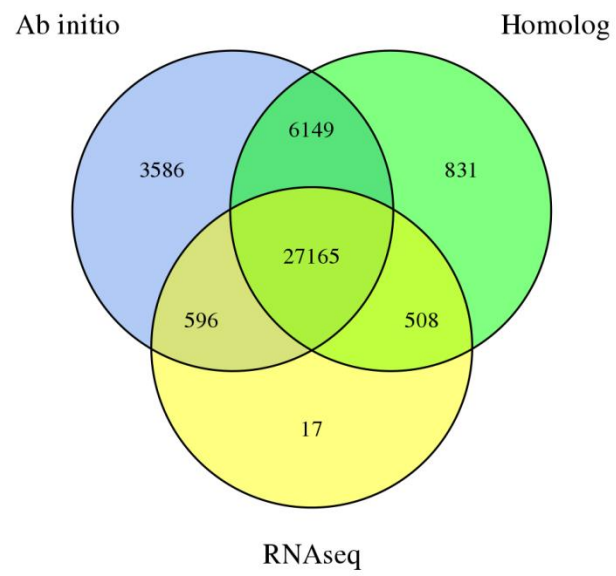
This PDF file includes:

Supplementary Figs. 1–18

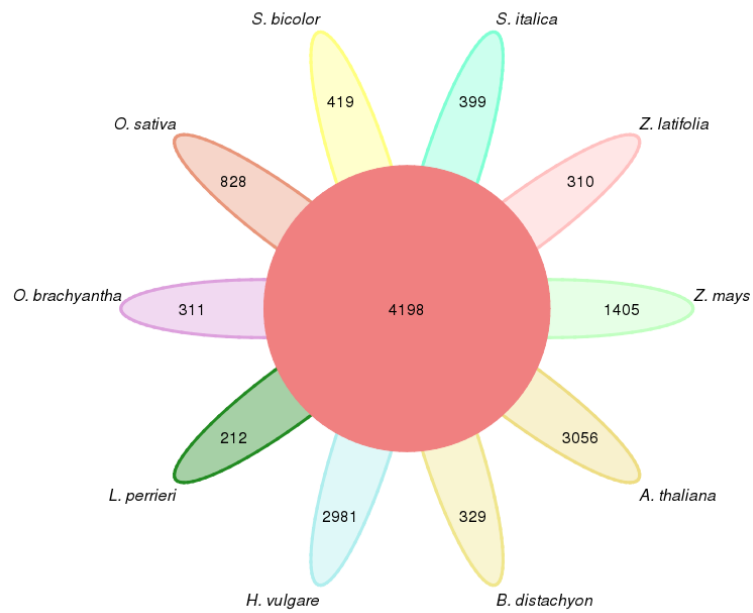
Supplementary Tables 1–11

Supplementary References

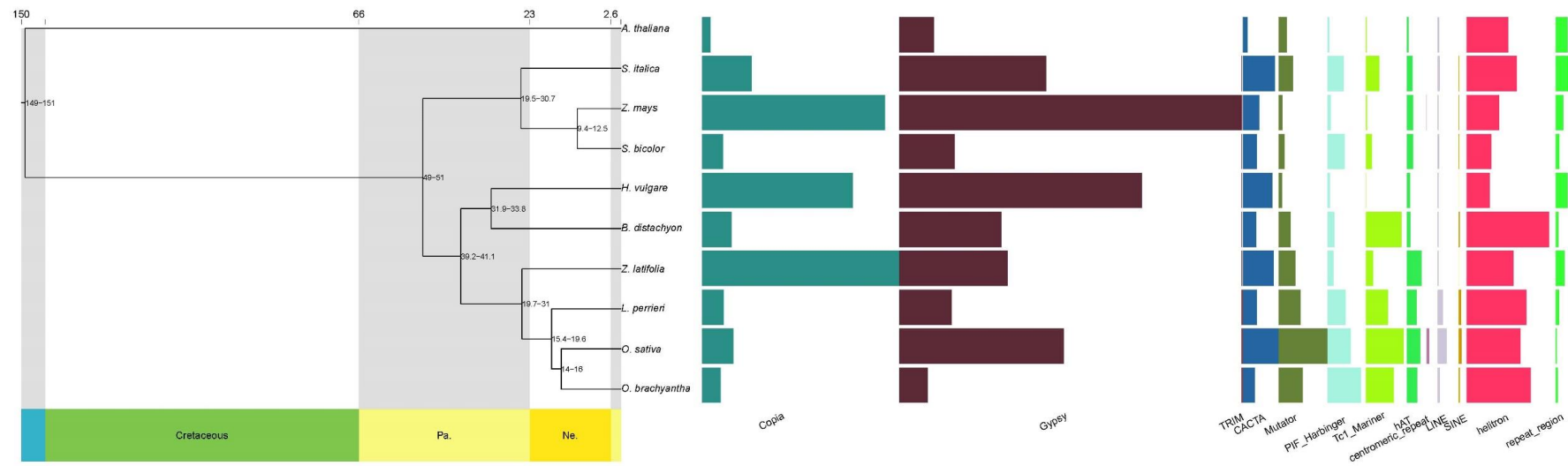
Supplementary Figures



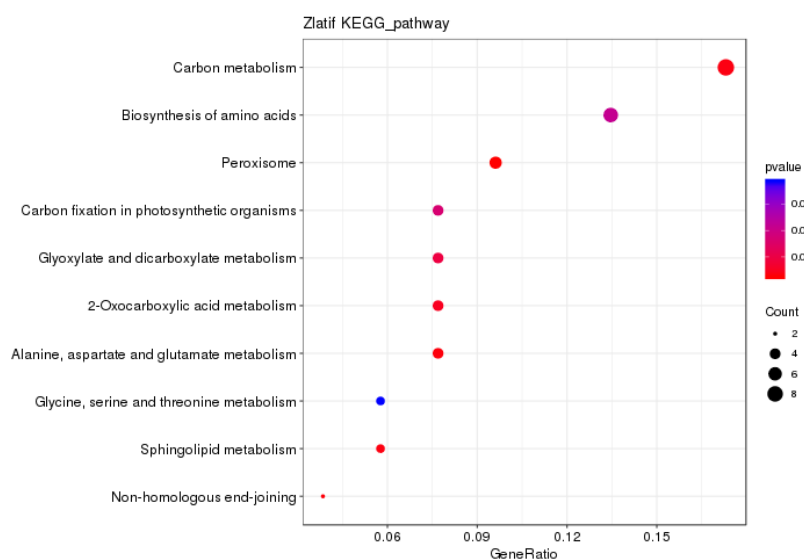
Supplementary Fig. 1. Distribution of Chinese wild rice genes derived from three prediction methods after EVIDENCEModeler integration.



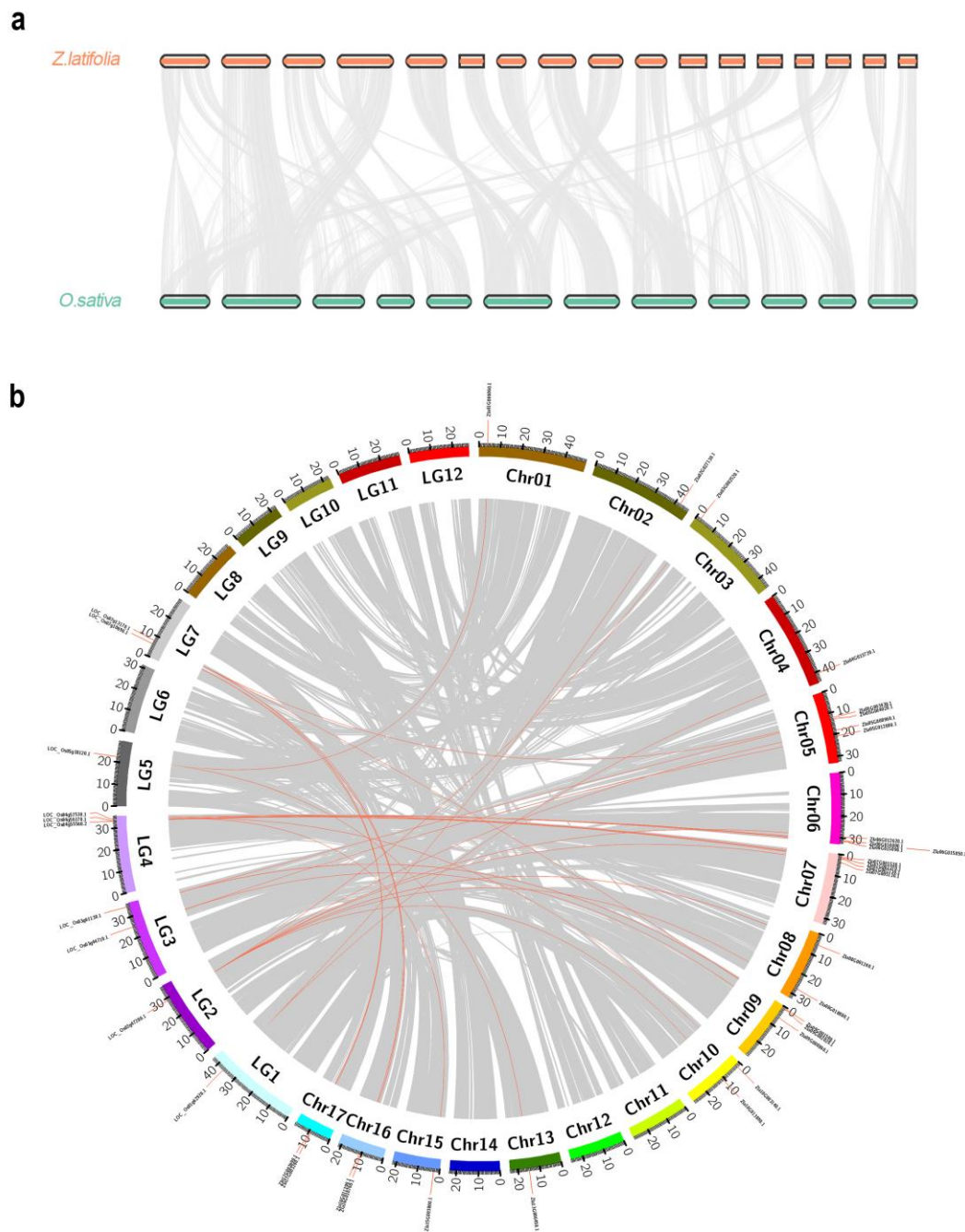
Supplementary Fig. 2. Petal diagram of gene family clustering of *Zizania latifolia* and nine other species. The number of gene families shared by all species is shown in the middle circle and the number of gene families unique to each species is shown in the outer edges.



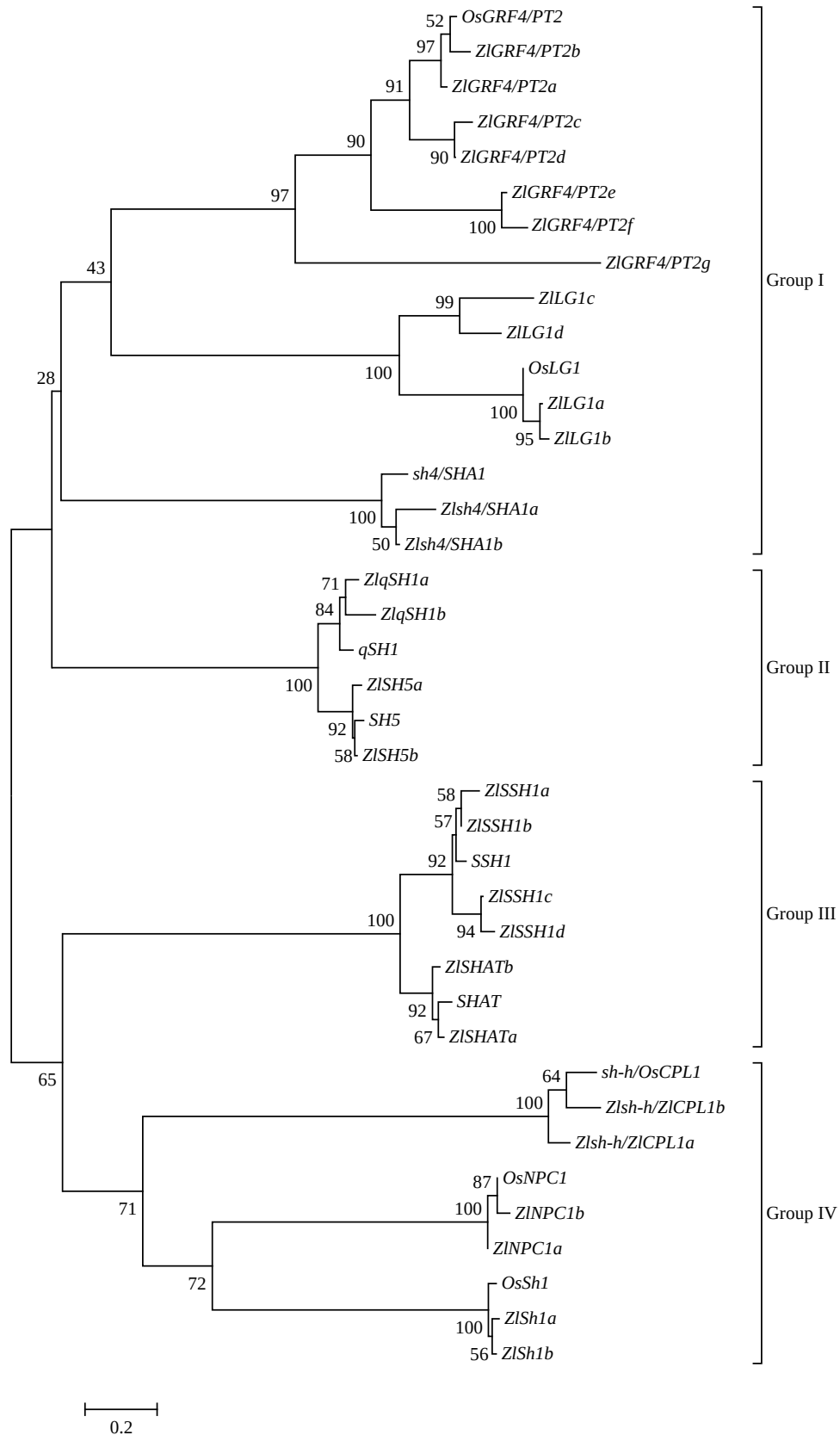
Supplementary Fig. 3. Evolutionary trees and transposable element (TE) of Chinese wild rice and other representative plant species with differentiation time. Time on the evolutionary tree represents the divergence time supported by 95% of the highest posterior density (HPD). At the bottom of the tree is the geological era, and at the top of the tree is the absolute age, in millions of years, defined by the shadow of each geological period. The geological periods include Cretaceous, Paleogene (Pa.), and Neogene (Ne.). Each coloured bar represents the proportion of each TE superfamily.



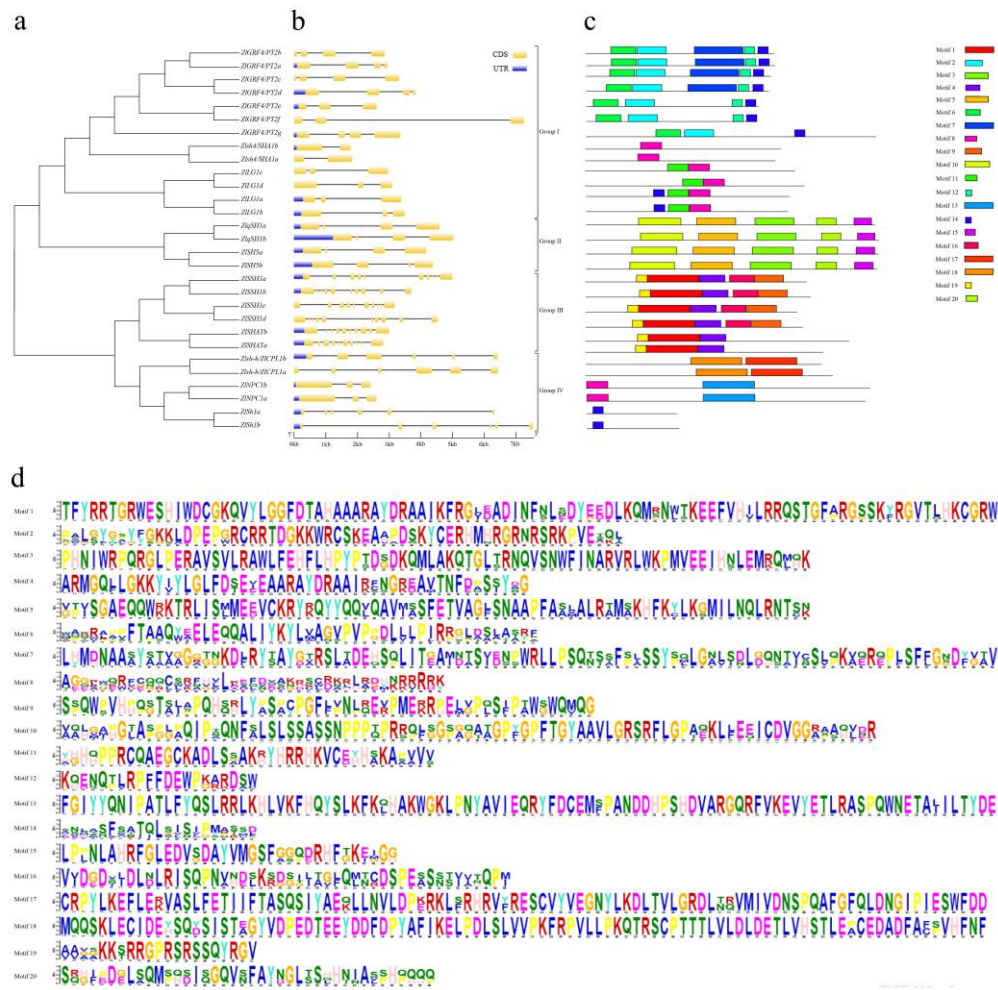
Supplementary Fig. 4. KEGG enrichment analyses for the positive selection genes in *Zizania latifolia*.



Supplementary Fig. 5. Collinearity analysis between genomes (a) and seed shattering-related genes (b) of *Zizania latifolia* and *Oryza sativa*. Chr01–Chr17, chromosomes 1–17 of *Z. latifolia*; LG1–LG12; linkage groups 1–12 of *O. sativa*. Candidate genes for seed shattering in *O. sativa* and *Z. latifolia* are shown in Supplementary Data 4. The grey lines represent the collinearity between the *O. sativa* and *Z. latifolia* genome, and the red lines represent the collinearity of seed shattering-related genes between *Z. latifolia* and *O. sativa*.



Supplementary Fig. 6. A phylogenetic tree of seed shattering genes in *Zizania latifolia* and *Oryza sativa*.



Supplementary Fig. 7. Analysis on motif of seed shattering genes in *Zizania latifolia*. **a**, phylogenetic tree; **b**, gene structure analysis; **c**, motif distribution analysis; **d**, motif sequence analysis.

qSH1 1 NSSAAGGGG- YCCGCGGGAHHHHHH- GHAGHL L LHHHPQHVAAGA VAAAAAAGGQVYHVPOHSRREKLRFPPDAQDSP
Zl qSH1a 1 NSSAAGGGG- YGAGFGDGAHHHHHHHGHAGHL L LHHHPQHLG----- GQVYHVPOHSRREKLRFPPDAADSP
Zl qSH1b 1 NSSAAGGGG- YGAGFGDGAHHHHHHHGHAGHL L LHHHPQHVAS----- AAAAGGQVYHVPOHSRREKLRFPPDAADSP
consensus 1 NSSAa GgG YGa Gp g GAELh HhH h g h AGHL L LHHHP q Hv a a a a g GQVYHVPOHSRREKLRFPPDAa DSp

qSH1 79 PPHGHGHGAPQOOQHGSWPPPAFYSSASSSSSYSPHSPTLAQAQLVAHGLAPPLPQIPTONFSLSLSSASSNPPPPQ
Zl qSH1a 68 PPH--- GHVPQOOQHGWPPPPSYSSASSSSSYSPHSPTLAQAQLVAHGLAPPLPQIPTONFSLSLSSASSNPPPP-
Zl qSH1b 72 PPH--- GHLPQOOQHGWPPPPAFYSSASSSSSYSPHSPTLVQAQLVAHGLAPPLPQIPTONFSLSLSSASSNPPPP-
consensus 81 PPH GHVPQq QOOHGV WPPPAFYSSASSSSSYSPHSPTLa QAQLvAHGLAPPLPQIP TQNFSLSLSSASSNPP

qSH1 159 AQPRRQLGGLAQATGPF GPF TGYAAVLGRS RFLGPAEKLFEELCDVGGAAASHVDRLT SDEGLLDADPMDCVHDVVDHDL
Zl qSH1a 142 -QPRRLS GPAQATGPF GPF TGYAAVLGRS RFLGPAEKLFEELCDVGGAAAHVDRLT CDALLDADPMDGADP-- LDHIV
Zl qSH1b 146 -QPRRLS GPAQATGPF GPF TGYAAVLGRS RFLGPEKLFEEELCDVGGAAAHVDRLT SDALLDADPMDGADT-- VDHEF
consensus 161 QPRRqLs GpAQa TGPFGPF TGYAAVLGRS RFLGP a EKLFEeL CDVGGAAa HVdRt i s d d a LLdADPMdGa D vDHdI

qSH1 239 GGADRAAAAGPISGAEQQWKKTRLISMEEVCKRYROYQOQVAVNASFETVAGFSNAAPFAALALRAKHF KCLKSM
Zl qSH1a 219 GGADRAAELGPISGAEQQWKKTRLISMEEVCKRYROYQOQVAVNASFETVAGFSNAAPFAALALRAKHF KCLKSM
Zl qSH1b 223 GGADRAAVHAGPISGAEQQWKKTRLISMEEVCKRYROYQOQVAVNASFETVAGFSNAAPFAALALRAKHF KCLKSM
consensus 241 GGADRAA daGPISGAEQQWKKT LISMEEVCKRYROYQOQVAVNASFETVAGFSNAAPFAALALRAKHF KcLKS M

qSH1 319 ILNQLRNTSNKAVKDC LNKELAVFGLAGGS G- - - GAGLQRANSASAFGOPHNIWRPQGLPERAVSVLRAWLFEHFL
Zl qSH1a 299 ILNQLRNTCNKPAKE- - NNKDLAVFGLAGGS GGAGAGLQRGNSASAFGOPHNIWRPQGLPERAVSVLRAWLFEHFL
Zl qSH1b 303 ILNQLRNTSNKAAKEGL NNKDLAVFGLAGGSAG- - AGAVLIRGNASAFGOPHNIWRPQGLPERAVSVLRAWLFEHFL
consensus 321 ILNQLRNTs NK AaKc g nNKdI AVFGLAGGS g G aGAg Lq RgNs ASAf g QPHNIWRPQGLPERAVSVLRAWLFEHFL

qSH1 395 HPYPTDGDQVLAKOTGLTRNOVSNWFINARVRLWKPVVEEIHNLQMRQMKHS VLDKGQHGmHHG----- AQHSSQCS
Zl qSH1a 377 HPYPTDGDQVLAKOTGLTRNOVSNWFINARVRLWKPVVEEIHNLQMRQMKHIVLDKGQHGmHHG----- AQHSSQCS
Zl qSH1b 381 HPYPTDGDQVLAKOTGLTRNOVSNWFINARVRLWKPVVEEIHNLQMRQMKHS VLDKGQHGmHHQAQHS AQHSSQCS
consensus 401 HPYPT DGDQVLAKQI GLTRNOVSNWF I NARVRLWKPVVEE I HNLEMRQMKHS VLDKGQHGmHHq AQHSSQCS

qSH1 469 GNPSMPSDHPGQSSSI TRNHNTAASQGF PDELSQMSQSI QGOVNFAYNGLSSQHNIASPHHQHOQVGVGLG--- GGNG
Zl qSH1a 451 GKPSDPSDHPGQSSSI TRNHNTAASQGF PDELSQMSQSI QGOVNFAYNGLSSQHNIASPHHQHOQVTSVGLGSGGAGNG
Zl qSH1b 461 GKSSDPSDHPGQSSSI TRNHNTAASQGF PDELSQMSQSI QGOVNFAYNGLSSQHNIASPHHQHOQVAVGVG--- Gg
consensus 481 GkpsdPSDHP gQSSSI TRNHNT aASQGF PDELSQMSQSI QGOVnFAYnGLs Sq HNI ASPhhQHQQVg gVGi G ggnG

qSH1 546 GVSLLTLGLHQNRRVCI AEPLPA- ALPANLAHRF GLEEVSDAYVMSF GGQDRHF GKEI GGHL LHDfVG
Zl qSH1a 530 GVSLLTLGLHQNRRVCI AEPLPA- ALPANLAHRF GLEEVSDAYVMSF GGQDRHF GKEI GGHL LHDfVG
Zl qSH1b 536 GVSLLTLGLHQNRRVCI AEPLPA- ALPANLAHRF GLEEVSDAYVMSF GGQDRHF GKEI GGHL LHDfVG
consensus 561 GVSLLTLGLHQNRRVCI AEPLPA- ALPANLAHRF GLEEVSDAYVMSF GGQDRHF GKEI GGHL LHDfVG

Supplementary Fig. 8. The protein sequence alignment of qSH1 and its homologous sequences in *Z. latifolia* and *Oryza* species.

ZI GRF4/PT2e 1
 ZI GRF4/PT2f 1
 ZI GRF4/PT2c 1
 ZI GRF4/PT2d 1
 ZI GRF4/PT2a 1
 ZI GRF4/PT2b 1
 Os GRF4/PT2 1
 ZI GRF4/PT2g 1
 consensus 1

ZI GRF4/PT2e 1 - M L S - - - - - S S P S A A G C G C Y Q C - - - - - P Q R A A - - - - - V F T A A Q M A E L
 ZI GRF4/PT2f 1 - M L S - - - - - S S P S A A G C G C Y Q C H O H O Q Q Q Q R A T - - - - - V F T A A Q M A E L
 ZI GRF4/PT2c 6 A S L S - P A A D H R S - - - - - S I F H F C R S S P L S A - V G E L A Q Q - H N M G G R W A A A G - - - - - A A P R P P P T A A Q Y E E L
 ZI GRF4/PT2d 6 A S L S - P A A D H R S - - - - - S I F H F C R S S P L S A - V G E L A Q Q H N M G G R W A A S - - - - - F T A A Q Y E E L
 ZI GRF4/PT2a 6 A S L S - P A A D H R S - P A V A T A S L L P F C R S S P L S A G G G V G L G D D A P M S G R W P - - - - - A R L P P P T A A Q Y E E L
 ZI GRF4/PT2b 6 A S L S - P A A D H R S - P A A T A S L L P F C R S S P L S A G G G V G L G E D A P M S G R W P - - - - - V R L P P P T A A Q Y E E L
 Os GRF4/PT2 6 A S L S - P A A D H R S - P A A T A S L L P F C R S S P L S A G G G V G L G E D A P M T A R W P P A - - - - - A A A R L P P P T A A Q Y E E L
 ZI GRF4/PT2g 81 L L G A S C S P G H G R L Q M L S F S S A S S S P S A A A A A A A A G A V P L Y Y G T P A S C S G L S S V S L S S M Q G A N A R V R O P P T P S Q M E L
 consensus 81 a s l s a a d h r s s i p f c r s s p l s a g m g g a q a r w r p p l a a q y e e l m a m p l

ZI GRF4/PT2e 32 E Q Q A L I Y K Y L V A G V P V P D L L P I R P H P A A A A T Y P F A N P A A P F Y H H H H P S L S Y Y A Y Y G K K L D P E P W R C R R T D G K K W R C
 ZI GRF4/PT2f 38 E Q Q A L I Y K Y L V A G V P V P D L L P I R P H S A A A A T Y S F A N P T A A P F Y H H H H P S L S Y Y A Y Y G K K L D P E P W R C R R T D G K K W R C
 ZI GRF4/PT2c 66 E Q Q A L I Y K Y L V A G V P V P D L L P I R R G - - - - - L D S L A S R F Y H - - - - - L S V L G Y C S Y F G K K L D P E P G R C R R T D G K K W R C
 ZI GRF4/PT2d 59 E Q Q A L I Y K Y L V A G V P V P D L L P I R R G - - - - - L D S L S R F Y H - - - - - H P A L G Y C P Y F G K K L D P E P G R C R R T D G K K W R C
 ZI GRF4/PT2a 69 E Q Q A L I Y K Y L V A G V P V P D L L P I R R G - - - - - L D S L A A R F Y H - - - - - H P A L G Y C P Y F G K K L D P E P G R C R R T D G K K W R C
 ZI GRF4/PT2b 69 E Q Q A L I Y K Y L V A G V P V P D L L P I R R G - - - - - L D S L A A R F Y H - - - - - H P A L G Y C P Y F G K K L D P E P G R C R R T D G K K W R C
 Os GRF4/PT2 74 E Q Q A L I Y K Y L V A G V P V P D L L P I R R G - - - - - L D S L A A R F Y N - - - - - H P A L G Y C P Y F G K K L D P E P G R C R R T D G K K W R C
 ZI GRF4/PT2g 161 E Q Q A L I Y K Y L V A G V P V P D L L P I R R S - - - - - L T S P Y S P A A Y F G S N T L G W G S F Q L G Y S G S A D P E P G R C R R T D G K K W R C
 consensus 161 E q Q A L I Y K Y L V A g v P V P d L L I P I R r g l d s l a a r f y h h p a l g y g y f g k k l D P e P g R C R R T D G K K W R C

ZI GRF4/PT2e 112 S K E A A P D S K Y C E R H M I R G N R S R K P V E S K T S A A A P C S Q P Q L S T V T T A T Q D A - - - - - D A P M P S L A V G A C A K T H G - - -
 ZI GRF4/PT2f 118 S K E A A P D S K Y C E R H M I R G N R S R K P V E S K T A A P A A C S Q P Q L S T V T T A T H D A - - - - - D T P I P S L A V G - - - V K T H G - - -
 ZI GRF4/PT2c 134 S K E A A P D S K Y C E R H M I R G N R S R K P V E A Q L V A P H S C - P P A T S P - - - T A A T A - - - - - I Q N H S L Y P A I A Y C G G G G G G G - - G S
 ZI GRF4/PT2d 127 S K E A A P D S K Y C E R H M I R G N R S R K P V E A Q L V A P H S C O P P A T A P A A A A A A A A - - - - - F P N H S L Y P A I A N C G G G G G R - - - G S
 ZI GRF4/PT2a 137 S K E A A P D S K Y C E R H M I R G N R S R K P V E I Q L S P O S Q P P A - A A A A A A P L A A A A N G S F Q N H S L Y P A I A G S N G G G G - - - R N
 ZI GRF4/PT2b 137 S K E A A P D S K Y C E R H M I R G N R S R K P V E I Q L V A Q S Q P P - - A T A A A A P L A A A S N D N S F Q N H S L Y P A I A G S N G A G G G - - - R N
 Os GRF4/PT2 142 S K E A A P D S K Y C E R H M I R G N R S R K P V E I Q L V A Q S Q P P S S V V G S A A A P L A A A S N G S S F Q N H S L Y P A I A G S N G G G G - - - R N
 ZI GRF4/PT2g 234 S R D A V A D Q K Y C E R H M I R G R H R S R K H V E G A G H A A K A I P A A A A A A A S A T Q P S A P A A H G G G A A A G L V I N H Q Q Q Q V N Y P A S
 consensus 241 S k e A a p D s K Y C E R H M h R G R n R S R K p V e s q l v a q p a l a a a a a a f q n h s l y p a i a g n g g g g

ZI GRF4/PT2e 180 - - L S L S C A G S - S Q H I M D A S S Y G S K Y S I G A K A D V G E L S F F S G A S G N T R G - - - - - F T I D S P - - - - - T D S S W H L S P S V P P Y P I
 ZI GRF4/PT2f 184 - - L S L C G A G S - S Q H I M D A S S Y G S K Y S I G A K A D V G E L S F F S G A S G N T R G - - - - - F T I D S P - - - - - T D S S W H L S P S V P P Y P I
 ZI GRF4/PT2c 204 A P G S F G L G T S - T Q L H M D N A A S Y S T A A G G G N K D F R Y S A Y G A R S I I D E H S P L I T G A M D T S - - - - - I D N P W R L L P S Q T S H F S I
 ZI GRF4/PT2d 199 A L G S F G L G T N - T Q L H M D N A A S Y S T A A G A G N K D F R Y S A Y D V R S I I D E H S Q L T I G A M D T S - - - - - I D S S W R L L P S Q T I A F S A
 ZI GRF4/PT2a 213 M P S S F G S A L G S S Q L H M D N A A A Y A T V G G R T G K D L R Y T A Y G T R S L A D E S Q L T I E A I N A S - - - - - V E N P W R L L P S Q N S S F P L
 ZI GRF4/PT2b 212 M P S S F G S A L G S S Q L H M D N A A A A V G G T G K D L R Y T A Y G T R S L A D E S Q L T I E A I N T S - - - - - V E N P W R L L P S Q N S S F P L
 Os GRF4/PT2 219 M P S S F G S A L G - S Q L H M D N A A P Y A A V G G T G K D L R Y T A Y G T R S L A D E S Q L T I E A I N T S - - - - - I E N P W R L L P S Q N S S F P L
 ZI GRF4/PT2g 314 T A D P C S L O Y S R E L V N K H N E S E H V Q D S D S L M L T S M S T R N I G N L P F S K Q H N P F E V A S R P D F G V V S P D S L M S S P H S S L E N
 consensus 321 p s f g a s s q l h m d n a a y s a g g a k d v r y s a y g t r s l g d e r s q l i t a i d t s i d p w r l l p s q s t f p l

ZI GRF4/PT2e 248 S K P R D S G L L P G A Y S Y S H I E P S Q E L G R V I T I A S - - - - - L S Q E C E R R S F G S C G A A C G L P G N M K H E N C P L R P F F D E W P G R K
 ZI GRF4/PT2f 252 S K P R D S G L L P G A Y S Y S H I E T S Q E L G V M T I A S - - - - - L S Q E C E R H S F G W C - - - - - V G G L L G N V K H E N C P L R P F F D E W P G R K
 ZI GRF4/PT2c 278 S - - - - - S Y P L M G N S D L D H N T I C S - - - - - V P R A E R E P L S F F G G Y V T V - - - - - N E S T I K O E N O T L R P F F D E W P K A R
 ZI GRF4/PT2d 273 S - - - - - S Y P L G N L I N L D N A I C S - - - - - V P R A E R E P L S F F G G Y V T V - - - - - V D S I K O E N O T L R P F F D E W P K A R
 ZI GRF4/PT2a 288 S - - - - - S Y S Q L G A L S D L G O N T I V S - - - - - L K K V O R Q P L S F F G G N D E C A - - - - - V D S A K O E N O T L R P F F D E W P K A R
 ZI GRF4/PT2b 287 S - - - - - S Y S Q L G A L S D L G O N T I V S - - - - - L K K V O R Q P L S F F G G N D E C A - - - - - V D S A K O E N O T L R P F F D E W P K A R
 Os GRF4/PT2 293 S - - - - - S Y S Q L G A L S D L G O N T I P S S - - - - - L S K V O R Q P L S F F G G N D Y A A - - - - - V D S V K O E N O T L R P F F D E W P K A R
 ZI GRF4/PT2g 394 V N L L T S Q S L N E H Q S I S V S L Q H F V D W P R T P A Q G P L A W P D A E D M Q A O R I Q L S I S A P M A S S D L T S A S T S P I H E R L N L S P L K L S R
 consensus 401 s s y s l g l s d l g n t i s l k q r p l s f f g d y g g v d s m k e n q t l k p f f d e w p k a r

ZI GRF4/PT2e 320 D S - - - - - W S E M D D R S N - - - - - Q T S F S T I T O L S I S I P M
 ZI GRF4/PT2f 322 D S - - - - - W S E M D D R S N - - - - - Q T S F S T I T O L S I S I P M
 ZI GRF4/PT2c 337 D S - - - - - W P E L A D D N S S - - - - - L A S F S A T O L S I S I P M
 ZI GRF4/PT2d 333 D S C - - - - - W P E I A D D N S S S - - - - - L A S F S A T O L S I S I P M
 ZI GRF4/PT2a 347 D S - - - - - W S D L A D E N A N - - - - - L S S F S A T O L S I S I P M
 ZI GRF4/PT2b 346 P A - - - - - W S D L A D E N A S - - - - - L S S F S A T O L S I S I P M
 Os GRF4/PT2 352 D S - - - - - W S D L A D E N A N - - - - - L S S F S A T O L S I S I P M
 ZI GRF4/PT2g 474 I Y S P V G L G I A A N R D E I N Q G E A N W I H M F R L S S V G G P L G E V L T K N N N L E A R N Y L S S S L N L L T D G W D S S G F E S S P V G V L Q K T
 consensus 481 d s w s e l a d d n s n l s s f s a t q l s i s i p m

ZI GRF4/PT2e 347 P R C G S P I G D - - - - -
 ZI GRF4/PT2f 349 P R C D - - - - -
 ZI GRF4/PT2c 363 A I S D F S T S S S R S P N C I Y S C - - - - -
 ZI GRF4/PT2d 362 A I S D F S T S S S R S P N D E - - - - -
 ZI GRF4/PT2a 374 A S S D F S A A S S R S T N C D - - - - -
 ZI GRF4/PT2b 373 A S S D F S A A S S R S T N C D - - - - -
 Os GRF4/PT2 379 A S S D F S A A S S R S T N C D - - - - -
 ZI GRF4/PT2g 554 T F G S V S S T G S S P R L E N H S V Y D G I S N L R D D L G S I V V N H P S I R L V
 consensus 561 a t s d f s g s s r s n g e

Supplementary Fig. 9. The protein sequence alignment of OsGRF4/PT2 and its homologous sequences in *Z. latifolia* and *Oryza* species.

```

Os Sh1      1  MS AQLVP APEHVCYVHCNFCNTI FAVSVP SNSMLNI VT VRCGHCTSLLSVNLRGLVQALP AEDHLDNLKMHNMSFRENY
Ob SH3      1  MS AQLVP APEHVCYVHCNFCNTI LAVSVP SNSMLNI VT VRCGHCTSLLSVNLRGLVQALP AEDHLDNLKMHNMSFRENY
Zl Sh1a     1  MS AQLVP APEHVCYVHCNFCNTI LAVSVP SNSMLNI VT VRCGHCTSLLSVNLRGLVQSLPVEDHLDNLKMHNMSFRENY
Zl Sh1b     1  MS AQLVP AAEHVCYVHCNFCNTI LAVSVP SNSMLNI VT VRCGHCTSLLSVNLRGLVQSLPVEDHLDNLKMHNMSFRENY
consensus  1  MS AQLVP APEHVCYVHCNFCNTI FAVSVP SNSMLNI VT VRCGHCTSLLSVNLRGLVQ TP EDHL QdNLKMHNmSFr ENy

Os Sh1      81  SEYGSSSR YGRVPMF S KNDTEHMLHVRPPEKRQRVPSAYNRFTKEEI RRI KANNPDI SHREAFSTAAKNWAHFPNI HFGL
Ob SH3      81  SEYGSSSR YGRVPMF S KNDTEHMLHVRPPEKRQRVPSAYNRFTKEEI RRI KANNPDI SHREAFSTAAKNWAHFPNI HFGL
Zl Sh1a     81  SEYGSSSR YGRVPMF S KYDPDHMLHVRPPEKRQRVPSAYNRFTKEEI RRI KANNPDI SHREAFSTAAKNWAHFPNI HFGL
Zl Sh1b     81  SEYGSSSR YGRVPMF S KNDPDHMLHVRPPEKRQRVPSAYNRFTKEEI RRI KANNPDI SHREAFSTAAKNWAHFPNI HFGL
consensus  81  SEYGSSSR YGRVPMf s kndtehmL hvr Ppekrqr vpsaynr f i kEei Rr i kaNnpdi shreafst Aaknwhf pni hf g

Os Sh1      161  LGSHESSKKLDEAI GAPSPQKVQRLY-----
Ob SH3      161  LGSHESSKKLDEAI GAPSPQKVQRLY-----
Zl Sh1a     161  LGSHESSKKLDEAI GAPDPQKVQRLY-----
Zl Sh1b     161  GSNESSRKLDEAELRVLKKF KDS TERGLI KS
consensus  161  lgsheSSKkldeai gapspqkvqrly

```

Supplementary Fig. 10. The protein sequence alignment of OsSh1 and its homologous sequences in *Z. latifolia* and *Oryza* species.

Zl NPC1a	1	MAAGEAGERCGRRLLVAVLLALVVS GHCLD AHHRGMKRRRRKHET HSP I KT VVVVVVMENRSFDHI LGWLS RSRPDI D
Zl NPC1b	1	MAAGEARERRGGRLLVAVLLALVVS GHCLD AHHRGMKRRRRKHET HSP I KT VVVVVVMENRSFDHI LGWLS RTRPDI D
Os NPC1	1	MAAGGGRERRGGRLLVAVLLALVVS GHCLD AHHRGMKRRRRKHET HSP I KT VVVVVVMENRSFDHI LGWLS RTRPDI D
consensus	1	Ma a Ge a r ERr GG RLLVa Vt LLa L VVS GHCLd a hHRGmKRRRRKHET HSP I KT VVVVVVMENRSFDHI LGWLS Rr RPDI D
Zl NPC1a	79	GLKGTESNHLNASDPSSPEIFVtDEAGYVDSDP GHGFEDI REQI FGSADTS AVPA PMS GF AQNARGMGLGMPQNVMSGF K
Zl NPC1b	79	GLKGTESNHLNASDPSSPEIFVtDEAGYVDSDP GHGFEDI REQI FGSADTS AVPA PMS GF AQNARGMGLGMPQNVMSGF K
Os NPC1	81	GLNGTQSNRLNASDPSSPEIFVtDEAGYVDSDP GHGFEDI REQI FGSADTS AVPA PMS GF AQNARGMGLGMPQNVMSGF K
consensus	81	GLkGTESNHLNASDPSSPEIFVt DEAGYVDSDP GHGFEDI REQI FGSa DTS AVPa PMS GF AQNARGMGLGMPQNVMSGF K
Zl NPC1a	159	PDSVPVYIALADEFAVFDRWFASVPTSTQPNRLVYHSATSHGLTFNARKDLIHGFPQKTI FDSLEENGLSFGI YYQNI PA
Zl NPC1b	159	PDSVPVYIALADEFAVFDRWFASVPTSTQPNRLVYHSATSHGLTFNARKDLIHGFPQKTI FDSLEENGLSFGI YYQNI PA
Os NPC1	161	PESVPVYIALADEFAVFDRWFASVPTSTQPNRLVYHSATSHGLTFNARKDLIHGFPQKTI FDSLEENGLSFGI YYQNI PA
consensus	161	PdSVpVYt ALADEFAVf DRWFASVPTSTQPNRLVYHSATSHGLTFNARKDLIHGFPQKTI FDSLe ENGLSFGI YYQNI PA
Zl NPC1a	239	TLFYQSLRRLKHLVKFHQYSLKFKLHAKWGKLPNYAVIEQR YFDCEMFPANDDHPSHDVARGQRFVKEVYETLRASPQWN
Zl NPC1b	239	TLFYQSLRRLKHLVKFHQYSLKFKLHAKWGKLPNYAVIEQR YFDCEMFPANDDHPSHDVARGQRFVKEVYETLRASPQWN
Os NPC1	241	TLFYQSLRRLKHLVKFHQYSLKFKLHAKWGKLPNYAVIEQR YFDCEMFPANDDHPSHDVARGQRFVKEVYETLRASPQWN
consensus	241	TLFYQSLRRLKHLVKFHQYSLKFKI HAKWGKLPNYAVIEQR YFDCEMFPANDDHPSHDVARGQRFVKEVYETLRASPQWN
Zl NPC1a	319	ETALITLYDEHGGFYDHVPTPVVGVPQPDGI VGPDPYF KFDRLGVRVPTFLI SPWIEKR-----T VI HAPNGPQ
Zl NPC1b	319	ETALITLYDEHGGFYDHVPTPVVGVPQPDGI VGPDPYF KFDRLGVRVPTFLI SPWIEKR-----T VI HAPNGPQ
Os NPC1	321	ETALITLYDEHGGFYDHVPTPVVGVPQPDGI VGPDPYF KFDRLGVRVPTFLI SPWIEKR-----T VI HAPNGPQ
consensus	321	ETALITLYDEHGGFYDHVPTPVvGVPQPDGI VGPDPy YF KFDRLGVRVPTf FLI SPWIEKR-----t VI HaPNGPQ
Zl NPC1a	389	DTSQYEHSSTPATVKKLFNLHSNFLT KRDAWAGTFENYFKI RKTPrTDCPEKLPEVTKSLRPF GPKEKDSLSSEFQVELIQ
Zl NPC1b	399	DTSQYEHSSTPATVKKLFNLHSNFLT KRDAWAGTFENYFKI RKTPrSDCPDKLPEVTKSLRPF GPKEKDSLSSEFQVELIQ
Os NPC1	391	DTSQYEHSSTPATVKKLFNLHSNFLT KRDAWAGTFENYFKI RKTPrTDCPEKLPEVTKSLRPF GPKEKDSLSSEFQVELIQ
consensus	401	Dt SQYEHSSTPATVKKLFNLHSNFLT KRDAWAGTFENYFKI RKTPrT DCPe KLPEVTKSLr PF GPKEKDSLSSEFQVELIQ
Zl NPC1a	469	LASQLNGDHVLTYPDI GRTMTVGEANRYAEDAVARFLEAGRI ALRAGANESALVTMRPALTS RASMSGLSSEL
Zl NPC1b	479	LASQLNGDHVLTYPDI GRTMTVGEANRYAEDAVARFLEAGRI ALRAGANESALVTMRPALTS RASMSGLSSEL
Os NPC1	471	LASQLNGDHVLTYPDI GRTMTVGEANRYAEDAVARFLEAGRI ALRAGANESALVTMRPALTS RASMSGLSSEL
consensus	481	LASQLNGDHVLTYPDI GRTMTVGEANRYAEDAVARFLEAGRI ALRAGANESALVTMRPALTS RAs mSs dLSSEL

Supplementary Fig. 11. The protein sequence alignment of OsNPC1 and its homologous sequences in *Z. latifolia* and *Oryza* species.

```

sh4/ SHA1      1 MSGSADPSPSASTAGAAVSPALALRAHGHGHHTATPPSCATGPAPPPSPASGSAPRDYRKGNWILHETLI LI TANR
GL4_Wi411     1 MGSAADPSPSASTAGAAASPLALLRAHGHGHHTPPSAATGPAPPPSPASGSAPRDYRKGNWILHETLI LI TAKRLDD
GL4_I RGC102305 1 MSGSADPSPSASTAGAAASPLALLRAHGHGHHTPPSAATGPAPPPSPASGSAPRDYRKGNWILHETLI LI TAKRLDD
Zl sh4/ SHA1a  1 MSSSDPSPSATAASPLALI RANPHOHPHLLTPSPATGPAPPPSPASVSAPRDYRKGNWILHETLI LI TAKRLDDDRAG
Zl sh4/ SHA1b  1 MSSSDPSPSASSAASPLALI RANPHOHPHLLTPSPATGPAPPPSPASVSAPRDYRKGNWILHETLI LI TAKRLDDDRR
consensus     1 MSgSSa ps sattaaaaa al rh h tp ta sgp pppps p s r yr l tl l r

sh4/ SHA1      81 LDDRRAGVCGAAAGGGCAGSPPTPRS AEQRWKVVENYCWKNGCLRSQNCNDKWDNLLRDYKKVRDYESRVAATAATGG
GL4_Wi411     81 DRRAGVCGGAAAGGGGAGSPPTPRS AEQRWKVVENYCWKNGCLRSQNCNDKWDNLLRDYKKVRDYESRVAATAATGGAA
GL4_I RGC102305 81 DRRAGVCGGAAAGGGGAGSPPTPRS AEQRWKVVENYCWKNGCLRSQNCNDKWDNLLRDYKKVRDYESRVAATAATGGAA
Zl sh4/ SHA1a  81 GGGGVSAGVAAAGSPPTPRS AEQRWKVVENYCWKNGCLRSQNCNDKWDNLLRDYKKVRDYESRVAATAATGGAAASAAAP
Zl sh4/ SHA1b  81 AGGGSACAGACAGAAAGSPPTPRS AEQRWKVVENYCWKNGCLRSQNCNDKWDNLLRDYKKVRDYESRVAATAATGGAA
consensus     81 agvagggaaggaggagspptp r rw nyc k nq qdn d r k k rd saaAA tggaa

sh4/ SHA1      161 AAANSAPLPSYWTMERHERKDCNLP TNLAPEVYDALSEVLSRRAARRGGATI APTPPPPPLALPLPPPPPPSPKPLVA
GL4_Wi411     161 AAANSAPLPSYWTMERHERKDCNLP TNLAPEVYDALSEVLSRRAARRGGATI APTPPPPPLALPLPPPPPPSPKPLVAQ
GL4_I RGC102305 161 AAANSAPLPSYWTMERHERKDCNLP TNLAPEVYDALSEVLSRRAARRGGATI APTPPPPPLALPLPPPPPPSPKPLVAQQH
Zl sh4/ SHA1a  161 PHAALPSYWTMERHERKDCNLP TNLAPEVHDALSEVLSRRAARRGGATI APTSLAPLALPLPPPPPPSPKPLT AQQQHH
Zl sh4/ SHA1b  161 AGASTGAAPLAILLPSYWTMERHERKDCNLP TNLAPEVYDALSDVLSRRAARRGGVTI APTPPPPPLALPLPPPPPPSP
consensus     161 aaas p p tm h hk ev g gg pppppp lpppppppp p

sh4/ SHA1      241 QQQHHGHHLHPPPPOPPPSLQLPPAVVAPPASVS AEEMSGSSES GEEEGSGGEPEAKRRRLSRLGSSVVR SATV
GL4_Wi411     241 QHHHGHHLHPPPPOPPPSLQLPPAVVAPPASVS AEEMSGSSES GEEEGSGGEPEAKRRRLSRLGSSVVR SATV
GL4_I RGC102305 241 HHHHGHHLHPPPPOPPPSLQLPPAVVAPPASVS AEEMSGSSES EEDNGSGGPEAKRRRLNRLGSSVVR SATVVARTLVACEEKRERHREL
Zl sh4/ SHA1a  241 HHHHGHHLHPPPPOPPPSLQLPPAVVAPPASVS AEEMSGSSES EEDNGSGGPEAKRRRLNRLGSSVVR SATVVARTLVACEEKRERHREL
Zl sh4/ SHA1b  241 PKPLITQQHHHHPHPLPPAVAPPASVS AEEMSGSSES GEEEGSGGPEAKRRRLSRLGSSVVR SATVVARTLVACE
consensus     241 hhh h hhhpp ppp p s pa v e e e e e g g e s k r r r l s r l g s s v v r s a t v v a r t l v a c e

sh4/ SHA1      321 VARTLVACEEKRERHRELLEERRLLEEERTEVRRQGFAGLI AAVNSLSSAI HALVSDHRS GDSSGR
GL4_Wi411     321 ARTLVACEEKRERHRELLEERRLLEEERTEVRRQGFAGLI AAVNSLSSAI HALVSDHRS GDSSGR-
GL4_I RGC102305 321 LQLEERRRLLEEERTEVRRQGFAGLVSAVNSLSSAI HALVSDHRS GDSSR-
Zl sh4/ SHA1a  321 LQLEERRRLLEEERTEVRRQGFAGLVSAVNSLSSAI HALVSDHRS GDSSR-
Zl sh4/ SHA1b  321 EKRERHRELLEERRLLEEERTEVRRQGFAGLSAIVNSLSSAI HALVSDHRS GDSSR-
consensus     321 e r h e vr a

```

Supplementary Fig. 12. The protein sequence alignment of sh4/SHA1 and its homologous sequences in *Z. latifolia* and *Oryza* species.

```

SHAT      1  MWDLNDSPAAEa-APPP-LSPSADDSGASSSSAAAVVEI PDDADDL--SAAVV---VVTROFFPPAPGGGGDPAPGNA
Z1 SHATa  1  MWDLNDSPAAEg-APPPQLSPSADDSGASSSSAAAVVEI PDDADDSSATAAGa---VVTROFFPPAPb---GEPAPGNV
Z1 SHATb  1  MWDLNDSPAAEgQPPPPALSPSGDDSGASSSSAAAVVEI PDDADDSSVAAADADADA VVTROFFPPAPb---GEPAPGNT
consensus 1  MWdLNDSPAAEg aPPP LSPSaDDSGASSSSAAAVVEI PDDADDSS sAA a VVTROFFPPA P GePAPGN

SHAT      73  RAGWLRLAGAAPPMAATG---PAASAAVS KKSRRGPRSRSSQYRGVTFYRRITGRWESH WDCGKQVYLGGFDTAHAAARA
Z1 SHATa  73  RAGWLRLAGTAPPAAAAGGAAAAAAAAS KKSRRGPRSRSSQYRGVTFYRRITGRWESH WDCGKQVYLGGFDTAHAAARA
Z1 SHATb  78  RAGWLRLARAAPPAPAAG-VAAAAAAAAS KKSRRGPRSRSSQYRGVTFYRRITGRWESH WDCGKQVYLGGFDTAHAAARA
consensus 81  RAGWLRLAg aAPPa aAa G a aAAaAAa SKSRRGPRSRSSQYRGVTFYRRITGRWESH WDCGKQVYLGGFDTAHAAARA

SHAT      150 YDRAAI KFRGVEADI NFSLEDYEDDLKQMSNLtKEEFVHVLRRQSTGFPRGSSKYRGVTLHKCGRWEARMGQFLGKKYVY
Z1 SHATa  153 YDRAAI KFRGVEADI NFSLEDYEDDLKQMSNLtKEEFVHVLRRQSTGFPRGSSKYRGVTLHKCGRWEARMGQFLGKKYVY
Z1 SHATb  157 YDRAAI KFRGVEADI NFSLEDYEDDLKQMSNLtKEEFVHVLRRQSTGFPRGSSKYRGVTLHKCGRWEARMGQFLGKKYVY
consensus 161 YDRAAI KFRGVEADI NFSLEDYEDDLKQMSNLtKEEFVHVLRRQSTGFPRGSSKYRGVTLHKCGRWEARMGQFLGKKYVY

SHAT      230 LGLFDTEEEAARAYDRAAI KCNGKDAVINFDPSIYAGETEPMAAATGDAAEHNLDSLGS SAGSKRGNDGGGDDEITGG
Z1 SHATa  233 LGLFDTEEEAARAYDRAAI KCNGKDAVINFDPSIYAGETEPMAAATGDG-EHNLDSLGS SAGSKRGSLGGGDDESAG-
Z1 SHATb  237 LGLFDTEEEAARAYDRAAI KCNGKDAVINFDPSIYAGETEPMAAATGDG-EHNLDSLGS SAGSKRGSLGGGDDESAG-
consensus 241 LGLFDTEEEAARAYDRAAI KCnGKDAVINFDpSIYAGETeP AaATGDg EHNLDSLGS SAGSKRGsIdGGGDDESaG

SHAT      310 GGGGAGSDQRVPAF DL DWQTAAARSTKAKFDQNSNHPQMPPVLQVTHLPFSPRHHQFLSN---GDPG-----
Z1 SHATa  311 -GSAAGSDQRVPAF DL DWQTAAARSTKAKFDQNSKHPQMPPVLQVSHLPFSPRHHQFVSN---GDPG-----
Z1 SHATb  315 -GAAAGSDERVPAF DL DWQTAAARSTKAKFDQNSKHPQMPPVLQVSHLPFSPRHHQVVRQWRS GDSICRPVSDRRRRG
consensus 321 GgaAGSDqRV MAFDL DWQTAAARSTKAKFDQNS kHPQMPPVLQVs HLPFSPR HHQFvsn GdpG

SHAT      376 -----TAGGLSLTTG-----AGMAG--HWPPQ- QOQGWGNAGGMSWPPEPH-----PPPPPT
Z1 SHATa  374 -----TAGGLPLTTGG-----VGGAGGGHWPPQPLHGWGSGGGMNWPMPEH-----PPPPAN
Z1 SHATb  392 RRGRRALASSADASAAATRLGQRRRRHELADAAAP AANQRRRRRRNRYS SCSTITITLTHCNASLKLAAEWWPPDPPT
consensus 401 tAgglslttG aggagg hwppq q qgwgsaggmswp pph PpPpPt

SHAT      420 NAAAAATATAAAASSRFPPYIATQASTWLQR-NGEHSLTRPT
Z1 SHATa  422 DAAATATATAAAASSRFPPYIATQASSWLQR-NGYHSLTRPT
Z1 SHATb  472 HLDDDHQPCASVVSPPWKRPITLRSTERRCVSLTQGNYSIATAI
consensus 481 aaa atatAaaaSsrfpPyIatqastwlqk nGfHSItrpI

```

Supplementary Fig. 13. The protein sequence alignment of SHAT1 and its homologous sequences in *Z. latifolia* and *Oryza* species.

ZI LG1a 1 --- MNVP --- SANSCEFGYAQ- -AAAAHINOTPPPPALLPT MDQDGGGG- IQRDHH- - - - - LGYNLEPSSLALLPP
 ZI LG1b 1 --- MNVP --- AANSCEFGYAQAAAAAHNPTOPPPALFPI MDHDGGGGGIHRDQO- - - - - LGYNLEPSSLALLPP
 Os LG1 1 --- MNVPSAAAASSCDDFGYNA- - - - - TP PPPPSILPT MDQDGGGGSIQRDHHQHNNHQQLGYNLEPSSLALLPP
 ZI LG1c 1 MMSGRVNAAVTAAG- - - DDFPEAP- - - - - MHQPPQGYVGF EHGAAAQRAQVQHIIHHH- - - - - MYDS- LDF AAAMQF
 ZI LG1d 1 MMSRLNTGAMASAVDVVDFGYAP- - - - - MQPYVGFDPAGMGAAAGERP- - - LFCQYHG- - - - - LYDGLDFASAAAF
 consensus 1 mmNvp aaa scddfgya hn pppallpimd dgggg iqrdhh lgynepslallpp

 ZI LG1a 64 SN- -AAAAHHASVAHGS PHDILLQFYPTS- - - - - HYITTA- - - - - GNPYSHFT- - - - - AGSTFCSSYYQPPQATPEYYFPTLVSS
 ZI LG1b 67 SN- -AAAAHHS- - - - - HGGPHDILLQFYPTS- - - - - SHYITG- - - - - GNPYSHFT- - - - - AGSTFCPPYYQPPQATPEYYFPTLVSS
 Os LG1 69 SNAAAAAAHHATI AHAS PHDILLQFYPTS- - - - - HYIAAGGAGGGGNPYSHFTAAAAAGSTFCSSYYQPPQDAPPEYYFPTLVSS
 ZI LG1c 65 - - - QHASQ- - - - - - - - - - - LIALPNLP- - - - - NAPPPMPMPMLQMP- - - - - MPMPGDVYP- - - - - ALGMVKREGG
 ZI LG1d 67 PSFQEAALGLPGSAAAPGNNLLQQLAPP- - - - - ATPSSILQMPMMVTLPLPATADVYFDSGAGGLVKRED-
 consensus 81 sn aAaahh ahg phdlLqfypts hyl a gnPyshft agstfq yyqqpPq tpeyyfptlvss

 ZI LG1a 130 AEENMASFAATQLGLNLGYRTYFPPr- - - - - GGYTYGHHPPRCAEGCKADLSSAKR
 ZI LG1b 131 AEENMASFAATQLGLNLGYRTYFPPr- - - - - GGYTYGHHPPRCAEGCKADLSSAKR
 Os LG1 148 AEENMASFAATQLGLNLGYRTYFPPr- - - - - GGYTYGHHPPRCAEGCKADLSSAKR
 ZI LG1c 112 GGGQEGSAAGRI GLNLGRRTYFSIPGDM LAVDRLLMRSRLG- - - - - GVFGGLFG GAHH- - - - - QPPRCQAECKADLSSAKR
 ZI LG1d 133 VPLDVGGGPGRILGLNLGRRTYFSIPADVLAVDRLLRSRLGGLGMGMVGLGLGL GAHHHHHPPRCQAECKADLSSAKR
 consensus 161 aeenmasfaatql GLNLGYRTYFPPr GgytyghhPPRCQAECKADLSSAKR

 ZI LG1a 182 YHRRHKVCEHHS KAPVVVT AGGLHQRF CQQCSRFHL LDEF DDAKKS CRKRLADHNRRRRRS KPS DGEHAADKRAQANKA
 ZI LG1b 183 YHRRHKVCEHHS KAPVVVT AGGLHQRF CQQCSRFHL LDEF DDAKKS CRKRLADHNRRRRRS KPS DGEHAADKRAQANKA
 Os LG1 200 YHRRHKVCEHHS KAPVVVT AGGLHQRF CQQCSRFHL LDEF DDAKKS CRKRLADHNRRRRRS KPS DGEHSGEKKRAQANKS
 ZI LG1c 184 YHRRHKVCEYHAKASIVSTGG- KQQRFCQQCSRFHVL IEF DDAKKS CRKRLADHNRRRRRKPATSTAAATVAGKDAAPSAAP
 ZI LG1d 213 YHRRHKVCDYHAKAAAVVAGC- KQQRFCQQCSRFHVL IEF DDAKKS CRKRLADHNRRRRRKAG- - - GAQAKDLPPPLKKA
 consensus 241 YHRRHKVCEhHs KApvVvt aGglhQRFCQQCSRFHl LdeFdaKkS CRKRLadHNRRRRRS kpsdgeh aeKrraQanka

 ZI LG1a 262 ATAKDKAGSSSKNAGI G- DGLE TQL LGSAL LMSKDE- - - - - DETMDLGDVVKEAVDPKKGASVQQO- - - - - AHHGLHHQ
 ZI LG1b 263 ATAKDKAGSSSKNAGI GGDGME TQL LGSAL LMSKDE- - - - - DETMDLGDVVKEAVDPKKGASVQQO- - - - - AHHGLHHQ
 Os LG1 280 ATAKDKAGSSSKNAGI G- DGEETQL LGSAL LMSKDE- - - - - DETMDLGDVVKEAVDPKKGASVQQOQQQAHHGLHHQ
 ZI LG1c 263 GKKNATCYTGDSKNVMSAAKSPISSTSVI SCLAEQTKQ- - - - - AAAPPTTLTLGAPPP- ESSPOLGSPMLHVHPSG- -
 ZI LG1d 289 ADS SI ASSYTS EHK- - - AAAS TITAS CVSC IQLADNGQTGGI AAEAAAPTTLTLGAPPPREEDVGLGTVLVMQQRVQED
 consensus 321 a kdkagsssknagig dg etqlgsa lmskde deam lgdvvkeAvdPkgkasmqqq ahhglh q

 ZI LG1a 330 S- - - QOHGFPFPSSSGS CLFPQSCAAVSS TDTITSNI AQVQEPS- - - - - S F H Q H H H Q H S N I L Q L G Q A M F D L D F
 ZI LG1b 331 S- - - QOHGFPFPSSSGS CLFPQSCAAVSS TDTITSNI AQVQEPS- - - - - S F H Q H H H Q H S N I L Q L G Q A M F D L D F
 Os LG1 350 SH- - - QOHGFPFPSSSGS CLFPQSCAAVSS TDTITSNI AQVQEPS- - - - - S F H Q H H H Q H S N I L Q L G Q A M F D L D F
 ZI LG1c 335 - - - AHHHHHHQOEHQLSLSLMT GGGGSSNN- - - NI LSCSS VCS SALPS TATNGEVSDNNNDNS HSNGGNS VHLFEVDIM
 ZI LG1d 366 DTHDHDHHLHMSLSLAQPHQQRHS SGASNNNDNI LSCSSAS- - - - - EQQNS SCNG- - - AVHFEVDIM
 consensus 401 s qhgfppfsssgsc lfpQsqaavss tdt sNlaqvqeps lfhq q snilqlgqamfdlDfD

 ZI LG1a 396 H
 ZI LG1b 393 H
 Os LG1 416 H
 ZI LG1c 410 -
 ZI LG1d 428 -
 consensus 481 h

Supplementary Fig. 14. The protein sequence alignment of OsLG1 and its homologous sequences in *Z. latifolia* and *Oryza* species.

```

Zl SH5a      1  MSSSAGGCVGGYGGGAEQH-----QQLLLGGAGQLYHVPQHSRREKLRFPDDPTTDSPL--PGSWLPPLPPFFYSYASS
Zl SH5b      1  MSSAAGC--GGYGGGAEQHQL-----QQLLLGQAAGQLYHVPQHSRREKLRFPDDPADDSSPL--LGSWLPPLPPFFYSYASS
SH5          1  MSSAAGGC--GGYGGGGGHHQHQHLLLLGQAAGQLYHVPQHSRREKLRFPDDPABSPPPPPGGSWLPPLPPFFYSYASS
consensus   1  MSSaAGGg--GGYGGGaeqHqh-----q q LLLGQaAGQLYHVPQHSRREKLRFPDDPa dSs Pp--pGSWLPPLPPFFYSYASS

Zl SH5a      72  SSSYSPHSPTLAHAQVQLVAHGMP--GTASGGAQIPSNQFALSLSSASSNPPTPRRQFGSGG---AAGPYGPFITYAA
Zl SH5b      72  SSSYSPHS---HAQAQLVAHGMP--GTASGGAQIPSNQFALSLSSASSNPPTPRRQFGSGG---AAGPYGPFITYAA
SH5          80  SSSYSPHSP--TLAHAQVVAHGMP--GAATSGGAQIPSNQFALSLSSASSNPPTPRRQFGSGGGGGAAGPYGPFITYAA
consensus   81  SSSYSPHSp--hAq aQLVAHGMP--g t aSGGAQIPs QNFALSLSSASSNPPTPRr QFGGs GG--AAGPYGPFITYAA

Zl SH5a      147 VLGRSRLFGPAQKLEELCDVGGRSQAQLDRGSDEGLLHVDAWDAAGSVDYEMDSDRAAAEAVTVSGAEQQWRKTRLISL
Zl SH5b      143 VLGRSRLFGPAQKLEELCDVGGRPQLNRGSDEGLLDLDADDAAGSVVDHMDACDRAATIEAVTVSGAEQQWRKTRLISL
SH5          158 VLGRSRLFGPAQKLEELCDVGGRPQAQLDRGSDEGLLDVDAWDAAGSVVDHEMDGSDRAAADAVTVSGAEQQWRKTRLISL
consensus   161 VLGRSRLFGPAQKLEELCDVGGRPaQLdRGSDEGLLdvDAMDAA GSVDheMDa sDRAa a eAVTVSGAEQQWRKTRLISL

Zl SH5a      227 MEDVCKRYKQYYQQLQAVISSFETVAGLSNAAPFASVALRTMSKHF KYLKGMI LNQLRNTSKGAAKDGLGKEDVAFGLM
Zl SH5b      223 MEDVCKRYKQYYQQLQAVISSFETVAGLSNAAPFASVALRTMSKHF KYLKGMI LNQLRNTSKGAAKDGLGKEDVDFGLM
SH5          238 MEDVCKRYKQYYQQLQAVISSFETVAGLSNAAPFASVALRTMSKHF KYLKGMI LNQLRNTSKGATKDGKEDTTNFGLM
consensus   241 MEDVCKRYkQYYQQLQAVi SSFETVAGLSNAAPFASmALRTMSKHF KYLKGml LNQLRNTs Kg Aa KDGLGKEDv nFGLM

Zl SH5a      307 GGGASLLRGNNVNSFSQPHNIWRPQRGLPERAVSVLRAWLFEHFLHPYPTDSDKQMLAKQTGLSRNQVSNWFI NARVRLW
Zl SH5b      303 GGGASLLRGNNVNSFSQPHNIWRPQRGLPERAVSVLRAWLFEHFLHPYPTDSDKQMLAKQTGLTRNQVSNWFI NARVRLW
SH5          318 GGGASLLRGNNVNSFSQPHNIWRPQRGLPERAVSVLRAWLFEHFLHPYPTDSDKQMLAKQTGLTRNQVSNWFI NARVRLW
consensus   321 GGGAs LLr GNNVNSFSQPHNIWRPQRGLPERAVSVLRAWLFEHFLHPYPTDSDKQMLAKQTGLt RNQVSNWFI NARVRLW

Zl SH5a      387 KPMVEEIHNLKMQKQNPSLDKNQLSMQHTQHSSDSSGKPSDPSNS--LQGGQSSMTRNHSVN--ASRHI EDGLSQMPHDI
Zl SH5b      383 KPMVEEIHSLKMQKQNPSLDKNQLSMQHTQHLSDDSSGKPSDPSNS--LQGGQSSSMARNHGVS--AASRHI DDGLAQMPHDI
SH5          398 KPMVEEIHNLKMQKQNPSLDKNQLSMQHTQHSSDSSGKPSDPSNS--LQGGQSSMTRNHSVN--ASRHI EDGLSQMPHDI
consensus   401 KPMVEEIHnLEMrQLKQNPSLDKNQLSMQHTQHsSDSSGKPSDPSNS--LQGGQsSSMtrNHsVs--ASRHI eDGLs QMPHDI

Zl SH5a      465 SGQVSFAYNGLTSHHNNALSHPQQOPDLTGAAGGAANAAGVSLTLGLHQNN--RAYIAEPLPAALPLNL AHRFGL EDVSDAY
Zl SH5b      463 SGQVSFAYNGLTSHHNTAVSHPQQOPDLTGTGSAANS GGVS LTLGLHQNN--RAYIAEPLPAALPLNL AHRFGL EDVSDAY
SH5          476 SGQVSFAYNGLAAHHSTAAAHHQOPDLTGTGGAANAAGVSLTLGLHQNN--RAYIAEPLPAALPLNL AHRFGL EDVSDAY
consensus   481 SGQVSFAYNGLt sHHt i Al sHp qQ PdLI Gt Gg AANa gGVSLTLGLHQNN RAYIAEPLPAALPLNL AHRFGL EDVSDAY

Zl SH5a      544 VMGSFGGQDRHFTKETGGHLLHDFVG
Zl SH5b      541 VMGSFGGQDRHFTKETGGHLLHDFVG
SH5          555 VMSFSGGQDRHFTKETGGHLLHDFVG
consensus   561 VMgSF GgQDRHFTKETGGHLLHDFVG

```

Supplementary Fig. 15. The protein sequence alignment of SH5 and its homologous sequences in *Z. latifolia* and *Oryza* species.

```

Zl sh-h/ Zl CPL1a      1  MQTRKKKGAPRNASGDHANIKTSRQPRRAIQVAALKKVNDLITSSVKKR- SVGAPSKKNRASKGRRKLT SASDADLAENE
Zl sh-h/ Zl CPL1b      1  MQTRKKKGAPKNASGDHANIKTSRQPRRAIQVAALKKVNDLITSSAKKKKSAGAPSKKNRASKGARRST SASDADLAENE
sh-h/ Os CPL1          1  MQTRKKKGATKNATGDPASTIKTSRQPRRAQAAASEKKVNDLITSSAKKKKS VGAPSKKNRASKGRRKLT SACDAANSENE
consensus              1  MQTRKKKGApKNAsGDhAntKTSRQPRRAiQvAAiEKKVNDLITSSaKKKKSvGAPSKKNRASKGaRK LTSAAsDaadtENE

Zl sh-h/ Zl CPL1a      80  VSQVVSICISIHQKSHHYNVDCILCPSTISSAFHHQS ECGESNFAKLGLEHKGDITLGLADSLTEGKTNAQGRKEATTIRFC
Zl sh-h/ Zl CPL1b      81  VS-----QHSDNDVDGRFCSTISSVFEHHQNECGESNFAKLGLEHKGDITLGLADSLTEGKTHAQGRKEATTIRSC
sh-h/ Os CPL1          81  VSQVVSIG-IHDKKSPDNVDGRPCSTISSPAPHILQK ECGESNFAKLGLEHKGDITLGSVSSVEIRTTTHAQGRKEVITIRSE
consensus              81  VSqvvs- p qhS dNVdGrpC SIFSSafHhQ ECGeSNfAK gLehKgdtLGla StEegkThAQGRKEaTTiRs g

Zl sh-h/ Zl CPL1a      160  SNQNNETLSHAVKTCVGS SDQHTLVOSTGNTIIVEEDEFSELGNLSSEVSAIYLAMQOSKLECI DEHSQDSI STEGYVD
Zl sh-h/ Zl CPL1b      150  SNQNNETLSOSTSOAVGTS DQHTLVOSTGNTIIVEEDEFSELGNLSSEVSAIYLAMQOSKLECI DEHSQDSI STEGYVD
sh-h/ Os CPL1          159  S-----TSHAVKTCVG-SDHHTLVNOSTGNTIIVEEDEFSELGNLSSEVSAIYLAMQOSKLECI DEHSQDSI STEGYVD
consensus              161  Snqnne tShavktcVG SDqhtLVnQStg NTiVEEDEFSELGNLSSEVSAIYLAMQOSKLECI DEHSQdSI SteGYVD

Zl sh-h/ Zl CPL1a      240  PEDTEEYDDFDPYAFI KELPDLSLVVPKFRPVLLPKQTRSCPITTLVLDDLDELTVHSTLEPCEDADFACS VHFNFKEHPI
Zl sh-h/ Zl CPL1b      229  PEDTEEYDDFDPYAFI KELPDLSLVVPKFRPVLLPKQTRSCPITTLVLDDLDELTVHSTLEPCEDADFAPVHFNFKEHPI
sh-h/ Os CPL1          232  PEDTEEYDDFDPYAFI KELPDLSLVVPKFRPVLLPKQTRSCPITTLKHLCTQLLNHAR-----MLTSHHQF-----
consensus              241  PEDTEEYDDFDPYAFI KeLPDLSLVVPKFRPVLLPKQTRSCPITTLvldldetLvHstle cedadfa svHfNfkeh i

Zl sh-h/ Zl CPL1a      320  YVRCRPLYKEFLERVASLFEETITFTASQSIYAEKLLNVLDPEERKLSRHRVRES CVYVEGNYLKDLTVLGRDLNQVMVD
Zl sh-h/ Zl CPL1b      309  YVRCRPLYKEFLERVASLFEETITFTASQSIYAEKLLNVLDPEERKLSRHRVRES CVYVEGNYLKDLTVLGRDLTRVMVD
sh-h/ Os CPL1          298  -----TITSENTOVMSDAVPLSKSSWRGLPILCRQSFSLQK-----AFMQNSISMFLILK
consensus              321  yvrerpylkeflekvastfetiitftasqsiyae llnvldp rkl rhrvrescvyvegnylkdl tvlgrdlt vmlvd

Zl sh-h/ Zl CPL1a      400  NSPOAFGFQLDNGIPIESWFDPRNDQELLKLLPFLESLVGVEDVRPYIASKFNLRKVATASSLPMDLQMGSTFAKENFR
Zl sh-h/ Zl CPL1b      389  NSPOAFGFQLDNGIPIESWFDPRNDQELLKLLPFLESLVGVEDVRPYIASKFNLRKVATASSLPMDLQMGSTFAKENFR
sh-h/ Os CPL1          349  ESCSVIVETV-----LPVFWRETT-----
consensus              401  nSpqafgfqldnglPiesWfdd ndqellkllpfleslvgedvrpyiaskfnlr kvatasslpmdlqm

Zl sh-h/ Zl CPL1a      480  V
Zl sh-h/ Zl CPL1b      -
sh-h/ Os CPL1          -
consensus              481

```

Supplementary Fig. 16. The protein sequence alignment of sh-h/OsCPL1 and its homologous sequences in *Z. latifolia* and *Oryza* species.

ZI SSH1a 1 MVLDLNVESPGGS AATSSSSAP --- EGGGGG --- FRFDLLGGSPDEEGCSPPVMT RQLFPSPAAVA AAVEADGSP
 ZI SSH1b 1 MVLDLNVESPGGS AATSSSSAP --- EGGGGGNGSGGVFRFDLLGGSPDEEGCSPPVMT RQLFPSPASAVAAVAEADGSP
 SSH1 1 MVLDLNVESPGGS AATSSSSAPPPPPDGGGGG --- FRFDLLGGSPDEEGCSPPVMT RQLFPSPSPAVVALAGDGSSTP
 ZI SSH1c 1 MVLDLNVASPG --- TSSSSVLN --- SGD --- VGLRLSLLA - SPDEDCS G - - - - - EHPFVASGITVIRQLLP
 ZI SSH1d 1 MVLDLNVASPGSGTSSSSVLN --- SGDCVEASVGRIGLLG - SPDEDCS G - - - - - EHPFVASGITVIRQLLP
 consensus 1 MvLDLNVeSPggsaatSSSS p e g g G g g FRfdLLggSPDEdgCS pvmtrqlf p s Pa vvaav s P

 ZI SSH1a 72 PP - - - - TPVGEASWPVRAADLGVAQS QRS P VGAAGGKKSRRGPRSRS SQYRGVTFYRRTGRWESH I WDCGKQVYLGGFD
 ZI SSH1b 77 PPA - - - - AAAGEGSWPVRAADLGVAQS QRS P VGAAGGKKSRRGPRSRS SQYRGVTFYRRTGRWESH I WDCGKQVYLGGFD
 SSH1 76 PLTMPMPAAAGEGPVPRRAADLGVAQS QRS P - - - AGGKKSRRGPRSRS SQYRGVTFYRRTGRWESH I WDCGKQVYLGGFD
 ZI SSH1c 57 PPPAPAP - - AAPAWQPRRAEDLG - AAQRVYV - - - AKKTRRGPRSRS SQYRGVTFYRRTGRWESH I WDCGKQVYLGGFD
 ZI SSH1d 65 PPPAPAPVAAAPAWQPRRAEDLG - LAQRVYV - - - AKKTRRGPRSRS SQYRGVTFYRRTGRWESH I WDCGKQVYLGGFD
 consensus 81 Pp p p aagea wPrRAaDLG a qRs p v aggKKSRRGPRSRS SQYRGVTFYRRTGRWESH I WDCGKQVYLGGFD

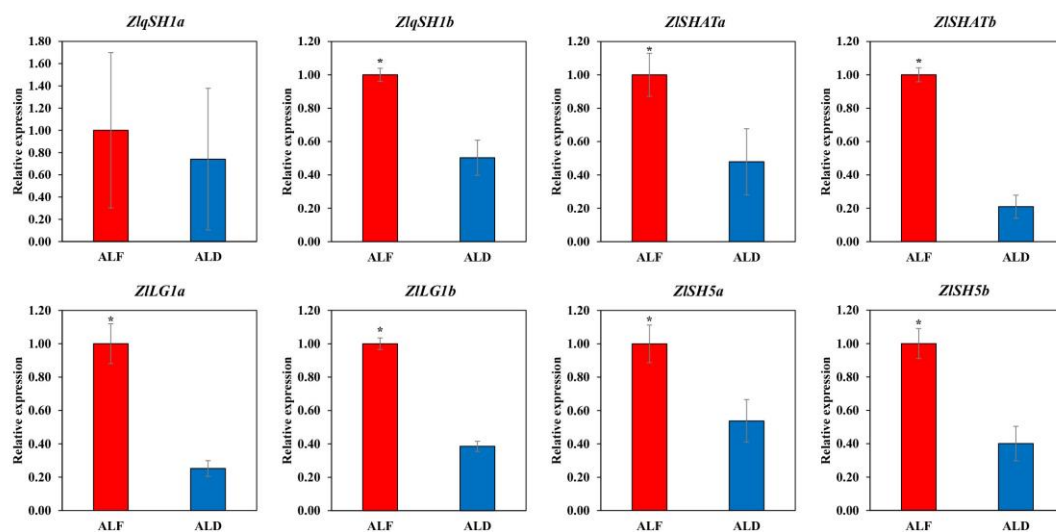
 ZI SSH1a 147 TAHAARAYDRAAI KFRGLDADI NFNLNDYEEDLKQMRNWTKEEFVHI LRRQSTGFARGSS KYRGVTLHKCGRWEARMGQ
 ZI SSH1b 153 TAHAARAYDRAAI KFRGLDADI NFNLNDYEEDLKQMRNWTKEEFVHI LRRQSTGFARGSS KYRGVTLHKCGRWEARMGQ
 SSH1 153 TAHAARAYDRAAI KFRGLDADI NFNLNDYEEDLKQMRNWTKEEFVHI LRRQSTGFARGSS KYRGVTLHKCGRWEARMGQ
 ZI SSH1c 130 TAHAARAYDRAAI KFRGLDADI NFNLNDYEEDLKQMRNWTKEEFVHI LRRQSTGFARGSS KYRGVTLHKCGRWEARMGQ
 ZI SSH1d 139 TAHAARAYDRAAI KFRGLDADI NFNLNDYEEDLKQMRNWTKEEFVHI LRRQSTGFARGSS KYRGVTLHKCGRWEARMGQ
 consensus 161 TAHAARAYDRAAI KFRGL dADI NFNLnDyEeDLKQMr NWTKEEFVHI LRRQSTGFARGs SKyRGVTI HKCGRWEARMGQ

 ZI SSH1a 227 LLGKKYI YLGLFDSEI EAARAYDRAAI RfNGREAVTNFDPSSYDGDVLPESkNE - VVDGDI I DLNLRISQPNVNDLKS DG
 ZI SSH1b 233 LLGKKYI YLGLFDSEI EAARAYDRAAI RfNGREAVTNFDPSSYDGDVLPET HNEAVVDGDI I DLNLRISQPNVNDLKS DG
 SSH1 233 LLGKKYI YLGLFDSEI EAARAYDRAAI RfNGREAVTNFDPSSYDGDVLPET DNE - VVDGDI I DLNLRISQPNVNDLKS DG
 ZI SSH1c 210 LLGKKYI YLGLFDSEI EAARAYDRAAI RfNGREAVTNFDPSSYDGDVLPET DNE - VVDGDI I DLNLRISQPNVNDLKS DG
 ZI SSH1d 219 LLGKKYI YLGLFDSEI EAARAYDRAAI RfNGREAVTNFDPSSYDGDVLPESkNE - VVDGDI I DLNLRISQPNVNDLKS DG
 consensus 241 LLGKKYI YLGLFDSEI EAARAYDRAAI RfNGREAVTNFdpSSYdGDvLPes kNE VvDgDi i DLNLRISQPNvhd1 Ks dg

 ZI SSH1a 306 ILTGFQNCDSPEASSSVVTQPI SP - - QWPVHPQGTST APQHSHLYASPCPGFFVNLREVPMEKRP ELGPQSFPT - WS WQ
 ZI SSH1b 313 ILTGFQNCDSPEASSSVVTQPI SP - - QWSVHPQGTST APQHSHLYASPCPGFFVNLREVPMEKRP ELGPQSFPT - WS WQ
 SSH1 312 ILTGFQNCDSPEASSSVVTQPI SP - - QWPVHPQGTSMs - QHPHLYASPCPGFFVNLREVPMEKRP ELGPQSFPT - WS WQ
 ZI SSH1c 290 TIAVLQITCDSPSSNTI ANQPMGSSS QWPVHQPSTALPPQHRLYPSACP GFLPNQERP MERRPELVPLSTPA - WG WQ
 ZI SSH1d 299 GIACLQITCDSPSSNTI VHQPMSGS QWPVHQPSTALPPQHRLYPSACP GFLPNQERP MERRPELVPLSTPA - WG WQ
 consensus 321 I t g f Q i n c D S P E a S s s v v t Q P i s p Q w P h p q g T s i p Q H h L Y a S p c P G f f v N L r E v P M e K r P e l g P q S f P t W s W Q

 ZI SSH1a 383 VQG - SPLPLPTAASSGFSTGTGADAPRVPISRPHPFP - HHQFYFPPTA -
 ZI SSH1b 390 VQG - SPLPLPTAASSGFSTGTGADAA RVPISRPHPFPCHHHQFYHPPTA -
 SSH1 389 VQG - SPLPLPTAASSGFSTGTGADAARS PISRPHPFP - HHQFYFPPTA -
 ZI SSH1c 369 VQGSHPPLTHAAASGF SAGAGAGASRRP - - - PP - FPD - - HPPFYFPPTA -
 ZI SSH1d 378 VQGSHPPLTHAAASGF SAGAGAGASRRP - - - PPFPPD - - HPPFYLPPTGS
 consensus 401 mQG SpLPLPtAASSGFStgtgAdA R PtsrPhpFPg hHqFYfpPTa

Supplementary Fig. 17. The protein sequence alignment of SSH1 and its homologous sequences in *Z. latifolia* and *Oryza* species.



Supplementary Fig. 18. Gene expression analyses of the selected seed shattering genes in *Zizania latifolia* as measured by qRT-PCR. Error bars indicate the standard error of the mean of the three independent experiments. Asterisks (*) indicate values that differ significantly according to Tukey's test ($P < 0.05$).

Supplementary Tables

Supplementary Table 1. Sequencing data statistics of Oxford Nanopore technology (ONT) and Illumina sequencing for Chinese wild rice Huai'an

Clean Reads	ONT sequencing		Illumina sequencing
	Raw reads	Clean reads	Clean reads
Size of library (bp)	20	20	350
Number of reads	5,009,027	3,970,614	228,677,035
Average length of reads (bp)	13,067	15,503	150
Total bases (bp)	65,456,644,980	61,559,056,417	68,499,340,900
Mean quality	7.89	8.37	Q30 > 92.19%
Sequencing depth	119.58×	112.46×	113.01×

Supplementary Table 2. Statistics of 17 Chinese wild rice assembled chromosomes

Group	Cluster number	Cluster length (bp)	Order number	Order length (bp)
Chr01	18	50,101,234	13	49,612,983
Chr02	29	48,046,012	19	47,112,129
Chr03	21	46,583,488	10	45,588,614
Chr04	12	45,369,969	8	44,977,458
Chr05	18	33,710,764	10	32,992,499
Chr06	12	33,104,032	9	32,889,764
Chr07	21	33,547,554	15	32,783,925
Chr08	12	32,655,998	8	32,287,380
Chr09	17	28,790,088	9	28,249,098
Chr10	25	29,018,640	14	27,421,652
Chr11	11	27,637,196	7	27,423,664
Chr12	22	26,660,827	14	25,649,991
Chr13	12	23,914,037	7	23,650,067
Chr14	31	24,459,573	21	23,072,643
Chr15	10	22,050,986	4	21,753,664
Chr16	16	22,318,559	10	21,397,443
Chr17	13	17,392,854	7	17,007,237
Total (ratio %)	300(90.36)	545,361,811(99.6 3)	185(61.67)	533,870,211(97. 89)

Supplementary Table 3. Evaluation of genome assembly completeness with 1,375 BUSCO¹ groups

Type	Genome	
	Number	Percent (%)
Complete BUSCOs	1,577	97.71
Complete and single-copy BUSCOs	1249	77.39
Complete and duplicated BUSCOs	328	20.32
Fragmented BUSCOs	9	0.56
Missing BUSCOs	28	1.73
Total Lineage BUSCOs	1,614	100

¹BUSCO, Benchmarking Universal Single-Copy Orthologue

Supplementary Table 4. Comparison of BUSCO groups of Next Generation Sequencing (NGS) and Oxford Nanopore technology (ONT) of the Chinese wild rice genomes

Type	Complete BUSCOs¹	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total lineage BUSCOs
NGS	4426 (90.40%)	3208 (65.52%)	1218 (24.88%)	96 (1.96%)	374 (7.64%)	4896
ONT	4634 (94.65%)	3130 (63.93%)	1504 (30.72%)	37 (0.76%)	225 (4.60%)	4896

¹BUSCO, Benchmarking Universal Single-Copy Orthologue

Supplementary Table 5. Comparison of long terminal repeat (LTR) assembly index (LAI) of Next Generation Sequencing (NGS) and Oxford Nanopore technology (ONT) of the Chinese wild rice genome

Type	From	To	Intact	Total	Raw LAI	LAI
NGS	1	603,989,347	0.0022	0.1782	1.25	6.88
ONT	1	547,397,560	0.0299	0.3773	7.94	13.57

Supplementary Table 6. Statistics of repetitive sequences in the Chinese wild rice genome

Type	Number	Length (bp)	Rate (%)
ClassI	244,890	206,264,154	37.68
ClassI/LTR/Copia	120,603	124,698,688	22.78
ClassI/LTR/Gypsy	89,814	68,441,219	12.50
ClassI/LTR/TRIM	2,276	416,227	0.08
ClassI/LTR/Unknown	30,558	12,172,708	2.22
ClassI/LINE	1,354	490,060	0.09
ClassI/SINE	284	44,904	0.01
ClassI/Unknown	1	348	0
ClassII	263,498	83,298,065	15.21
ClassII/TIR/CACTA	41,332	19,593,368	3.58
ClassII/TIR/Mutator	35,375	10,591,002	1.93
ClassII/TIR/PIF_Harbinger	14,615	3,554,936	0.65
ClassII/TIR/Tc1_Mariner	19,714	4,591,368	0.84
ClassII/TIR/hAT	31,088	9,096,346	1.66
ClassII/TIR/polinton	13	6,986	0
ClassII/TIR/Unknown	2,208	857,661	0.16
ClassII/helitron	102,073	29,475,694	5.38
ClassII/repeat_region	17,080	5,530,704	1.01
Total	508,388	289,562,219	52.89

Supplementary Table 7. Statistics of gene family clustering in 11 species

Species	Number of genes	Number of genes in orthogroups	Number of unassigned genes	Percentage of genes in orthogroups (%)	Percentage of unassigned genes (%)	Number of orthogroups containing species	Percentage of orthogroups containing species (%)	Number of species-specific orthogroups	Number of genes in species-specific orthogroups	Percentage of genes in species-specific orthogroups (%)
<i>Arabidopsis thaliana</i>	27,379	15,356	12,023	56.1	43.9	9,637	25.2	3,056	7,926	28.9
<i>Brachypodium distachyon</i>	32,439	26,402	6,037	81.4	18.6	20,713	54.3	329	957	3.0
<i>Hordeum vulgare</i>	63,658	50,307	13,351	79.0	21.0	23,472	61.5	2,981	17,345	27.2
<i>Leersia perrieri</i>	29,017	24,065	4,952	82.9	17.1	19,855	52.0	212	586	2.0
<i>Oryza brachyantha</i>	32,010	22,966	9,044	71.7	28.3	19,528	51.2	311	733	2.3
<i>Oryza sativa</i>	42,189	31,919	10,270	75.7	24.3	22,797	59.7	828	3,278	7.8
<i>Sorghum bicolor</i>	34,129	28,190	5,939	82.6	17.4	21,634	56.7	419	1,206	3.5
<i>Setaria italica</i>	34,584	29,831	4,753	86.3	13.7	21,966	57.5	399	1,363	3.9
<i>Zizania latifolia</i>	38,852	33,924	4,928	87.3	12.7	20,342	53.3	310	788	2.0
<i>Zea mays</i>	39,422	33,356	6,066	84.6	15.4	21,092	55.3	1,405	4,661	11.8

Supplementary Table 8. Seed shattering related genes in *Oryza sativa*

Gene	Chromosome	Genomic locus	Coding protein	Reference
<i>qSH1</i>	1	<i>LOC_Os01g62920.1</i>	BEL1-type homeobox	[1]
<i>OsGRF4/PT2</i>	2	<i>LOC_Os02g47280.1</i>	Growth-regulating factor	[2]
<i>OsSh1</i>	3	<i>LOC_Os03g44710.1</i>	YABBY transcription factor	[3]
<i>OsNPC1</i>	3	<i>LOC_Os03g61130.1</i>	Phospholipase C1	[4]
<i>sh4/SHA1</i>	4	<i>LOC_Os04g57530.1</i>	Myb3 transcription factor	[5,6]
<i>SHAT</i>	4	<i>LOC_Os04g55560.2</i>	AP2 transcription factor	[7]
<i>OsLG1</i>	4	<i>LOC_Os04g56170.1</i>	SQUAMOSA promoter-binding protein	[8,9]
<i>SH5</i>	5	<i>LOC_Os05g38120.1</i>	BEL1-type homeobox	[10]
<i>sh-h/OsCPL1</i>	7	<i>LOC_Os07g10690.1</i>	CTD phosphatase	[11]
<i>SSH1</i>	7	<i>LOC_Os07g13170.1</i>	AP2 transcription factor	[12]

Supplementary Table 9. Contents of phytohormones (ng/g) between abscission layer formation (ALF) and abscission layer degradation (ALD) tissues in Chinese wild rice. Error bars indicate the standard error of the mean of the three independent experiments. Asterisks (*) indicate values that differ significantly according to Tukey's test ($P < 0.05$).

Phytohormones	Category	ALF	ALD
Abscisic acid	ABA	79.5667±17.8646	906.6667±169.9098*
ABA-glucosyl ester	ABA	17.6667±1.4572	139±29.0517*
1-Aminocyclopropanecarboxylic acid	ETH	109±5.2915	152±13.1149*
<i>cis</i> -Zeatin	CK	0.182±0.033	0.2757±0.0282*
Dihydrozeatin-7-Glucoside	CK	0.489±0.1037	4.2±2.5247
Dihydrozeatin	CK	0±0	0.2177±0.377
<i>trans</i> -Zeatin	CK	0.6427±0.0636	1.6793±0.6311
<i>trans</i> -Zeatin riboside	CK	0.52±0.164	1.5613±0.5313*
N6-isopentenyladenine	CK	0.1807±0.0431	0.3923±0.0335*
N6-isopentenyladenosine	CK	0.2737±0.0496	0.3583±0.1334
Gibberellin A3	GA	0.4413±0.3823	0.56±0.9699
Gibberellin A4	GA	1.2133±1.3146	0±0
Gibberellin A7	GA	0.273±0.1495	0.1196±0.0709
Gibberellin A9	GA	2.0833±0.3099*	0±0
Gibberellin A15	GA	0.2667±0.1334	0.3433±0.5861
Gibberellin A19	GA	5.0633±0.6573*	0.98±1.6974
Gibberellin A24	GA	1.3367±1.1853	0.28±0.485
Gibberellin A53	GA	3.0933±1.6407	2.13±0.3704
Indole-3-acetic acid	Auxin	3.94±1.0553	121.0333±36.3704*
1- <i>O</i> -indol-3-ylacetylglucose	Auxin	0±0	33.0333±4.9238*
Indole-3-carboxylic acid	Auxin	3.1333±0.5829	12.2767±3.6382*
Indole-3-carboxaldehyde	Auxin	9.7233±0.9063	12.5±5.5669
Methyl indole-3-acetate	Auxin	2.27±0.6528	6.9±2.0219*
3-Indolepropionic acid	Auxin	2.3333±0.5853	1.6467±0.5829
Dihydrojasmonic acid	JA	2.1333±0.2715	2.9467±1.0228
Jasmonic acid	JA	267±67.5796*	93.6±16.4037
Jasmonoyl-L-isoleucine	JA	134±4	129.9±37.5237
Methyl jasmonate	JA	0.722±0.2413*	0.1857±0.0427
<i>cis</i> (+)-12-oxophytodienoic acid	JA	136±17.0587	141±44.3058
Salicylic acid	SA	131.6667±23.8607	311±14.7309*
Salicylic acid 2- <i>O</i> -β-Glucoside	SA	2633.3333±303.5347	5656.6667±190.0877*

Supplementary Table 10. Expression level of genes in *Zizania latifolia* phytoalexane gene cluster. ALF, abscission layer formation; ALD, abscission layer degradation.

Gene Name	Zlat_ID	Chr	Start	End	ALD-1	ALD-2	ALD-3	ALF-1	ALF-2	ALF-3	LEA-1	LEA-2	LEA-3	Stem-1	Stem-2	Stem-3
CYP76M5	Zla08 G012 880.1	C hr 8	225 389 24	225 404 29	0. 12 1	0. 06 96	0. 45 92	0.2 49 7	0.1 90 3	0.1 26 1	0. 05 47	0. 13 36	0. 05 29	0	0	0
CYP76M5	Zla08 G012 900.1	C hr 8	225 486 33	225 501 44	0. 57 84	0. 37 15	0. 67 65	2.3 28 6	3.2 15 8	4.8 43 5	0. 17 51	0. 0 88	0. 04 88	0	0	0
KSL11	Zla08 G012 910.1	C hr 8	225 519 04	225 541 44	0. 15 76	0. 23 62	0. 42 11	0.9 58 2	1.3 38 8	0.8 59 8	0 0 8	0 11 86	0. 0 86	0	0	0
CPS2	Zla08 G012 920.1	C hr 8	225 812 85	225 859 82	0. 47 15	1. 29 34	0. 60 4	4.9 56 8	3.3 70 5	1.8 53 21	0. 19 42	0. 24 75	0. 43 75	0	0	0
CYP76M8	Zla08 G012 930.1	C hr 8	225 926 10	225 941 09	2. 03 31	5. 93 86	3. 65 73	13. 16 44	7.8 39 6	3.7 43 4	2. 62 55	1. 50 49	2. 05 05	0	0	0
KSL7	Zla08 G012 940.1	C hr 8	226 037 52	226 098 36	5. 09 57	8. 21 51	5. 75 42	6.8 15 4	5.0 72 4	3.0 92 32	2. 39 05	1. 85 66	3. 29 66	0. 04 8	0	0. 42
CYP71Z6	Zla08 G012 960.1	C hr 8	226 199 82	226 218 64	2. 65 5	3. 39 9	2. 07 07	33. 56 23	33. 54 02	48. 41 51	2. 24 31	1. 45 47	0. 79 32	0	0	0
CYP76M7	Zla10 G007 850.1	C hr 10	532 751 5	532 908 6	5. 93 54	7. 16 67	7. 05 53	11. 56 48	11. 80 23	7.1 57 7	0. 98 7	0. 62 97	0. 98 04	0	0. 19 83	0
KSL5	Zla10 G007 880.1	C hr 10	534 152 8	534 605 6	0. 60 15	0. 77 43	0. 66 23	2.1 8 8	2.1 04 7	1.7 30 5	0. 28 02	0. 08 65	0. 22 12	0. 40 69	0. 24 82	0. 03 52
KSL6	Zla10 G007 890.1	C hr 10	534 757 0	535 222 9	4. 28 96	3. 94 19	4. 35 89	7.7 36 8	9.1 07 3	7.6 73 1	3. 25 15	3. 90 11	4. 61 59	5. 57 05	4. 26 78	5. 05 95
CYP76M6	Zla10 G007 900.1	C hr 10	535 947 3	536 097 2	0. 08 4	0 0	0. 03 58	2.1 72 4	1.7 43 2	0.5 50 3	0 0 0	0 0 0	0 0 0	0	0	0

Supplementary Table 11. Primers used for qRT-PCR of selected seed shattering genes

Gene name	Primer	Sequence (5'- 3')
<i>Actin</i>	Forward primer	AGAGCAGAGGCATTCCAAGT
	Reverse primer	ACTAACCGGCCACGTGTATT
<i>ZlqSH1a</i>	Forward primer	AACACCACTGCTTCTCAGGG
	Reverse primer	ATGCAATGTTGTGCTGCGAG
<i>ZlqSH1b</i>	Forward primer	CAACAACGCTGCTTCTCAGG
	Reverse primer	CAATGTTGTGCAGCGAGGAC
<i>ZISHATa</i>	Forward primer	TCGACCAGAACTCGAAGCAT
	Reverse primer	CGAATTGGTGATGCCTGGGA
<i>ZISHATb</i>	Forward primer	ACCCACCTAGATGACGACCA
	Reverse primer	GTGACACACACCTCCTCTCG
<i>ZILG1a</i>	Forward primer	GAACGTACCATCTGCCAACT
	Reverse primer	CCGTCCTGATCCATGATTGGAA
<i>ZILG1b</i>	Forward primer	GGCATCCATAGGGATCAGCA
	Reverse primer	TGAGGTAGTGGGAGCTGGTA
<i>ZISH5a</i>	Forward primer	TGCTCCCTTTGCTTCCGTAG
	Reverse primer	TGCCACATCTTCTTTGCCGA
<i>ZISH5b</i>	Forward primer	CGAGCTCATTGGCACTGGTA
	Reverse primer	TCAGCAATGTAAGCCCGGTT

Supplementary References

- [1] Konishi, S. et al. An SNP caused loss of seed shattering during rice domestication. *Science* **312**, 1392–1396 (2006).
- [2] Sun, P. et al. *OsGRF4* controls grain shape, panicle length and seed shattering in rice. *J. Integr. Plant Biol.* **58**, 836–847 (2016).
- [3] Lin, Z. et al. Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.* **44**, 720–724 (2012).
- [4] Cao, H. et al. Non-specific phospholipase C1 affects silicon distribution and mechanical strength in stem nodes of rice. *Plant J.* **86**, 308–321 (2016).
- [5] Li, C. et al. Rice domestication by reducing shattering. *Science* **311**, 1936–1939 (2006).
- [6] Lin, Z. et al. Origin of seed shattering in rice (*Oryza sativa* L.). *Planta* **226**, 11–20 (2007).
- [7] Zhou, Y. et al. Genetic control of seed shattering in rice by the APETALA2 transcription factor SHATTERING ABORTION1. *Plant Cell* **24**, 1034–1048 (2012).
- [8] Ishii, T. et al. *OsLGI* regulates a closed panicle trait in domesticated rice. *Nat. Genet.* **45**, 462–465 (2013).
- [9] Lee, J. et al. Mutations in the rice liguleless gene result in a complete loss of the auricle, ligule, and laminar joint. *Plant Mol. Biol.* **65**, 487–499 (2007).
- [10] Yoon, J. et al. The BEL1-type homeobox gene *SH5* induces seed shattering by enhancing abscission-zone development and inhibiting lignin biosynthesis. *Plant J.* **79**, 717–728 (2014).
- [11] Ji, H. et al. Inactivation of the CTD phosphatase-like gene *OsCPL1* enhances the development of the abscission layer and seed shattering in rice. *Plant J.* **61**, 96–106 (2010).
- [12] Jiang, L. et al. The APETALA2-Like transcription factor SUPERNUMERA RY BRACT controls rice seed shattering and seed size. *Plant Cell* **31**, 17–36 (2019).