

A

JK974084_Rp_	1	-----	MTGKAKPK	-----	KHTAKEIQAKVDAAT	TNRGGG	-----	
ABU95041.1_Ta_	1	-----	MTGKAKPK	-----	KHTAKEIAAKIDAAT	TNRGGG	-----	
NP_566182.1_At_	1	-----	MTGKAKPK	-----	KHTAKEIQAKIDAAL	TNRGGG	-----	
NP_197151_At_	1	-----	MTGKAKPK	-----	KHTAKEIQAKADAAL	TNRGGG	-----	
AA92680.1_Pa_	1	-----	MTGKAKPK	-----	KHTAKEIAAKLMRQRQVRGGG	-----	-----	
Y18788_Mt_	1	MAMEALNSPTTATPFPFEEPN---	LSYLETPWTK	CKRSKRSRMDQSSC	TEBEYLALCLIM	ARS	GNNDKK-----SDSVATPLTT	
AF539746_Ca_	1	MALEALNSPTGTPPPPFQFESDGGQLRYIEN-WRK	CKRSKRSRSMHQPT	TEBEYLALCLIM	ARS	GGSVNHQ	SLPPPAPVMKHLHAPSSSSAABEEK	
			zinc finger II		zinc finger I			
JK974084_Rp_	29	---KAGLADR---	TSSEKGGHAK	-----	YVCPHCKIT	TAPDVK	SMQIHHDARHP	-----
ABU95041.1_Ta_	29	---KAGQKDR---	LGQEKGGHAK	-----	MECPLCRV	TAPDVK	SMQIHHEARHP	-----
NP_566182.1_At_	29	---KAGLADR---	TKKEKGGHAK	-----	YECPHCKIT	TAPGLK	TMQIHHEAKHP	-----
NP_197151_At_	29	---KAGLADR---	TKKEKGGHAK	-----	YECPHCKIT	WFDLK	TMQIHHEAKHP	-----
AA92680.1_Pa_	29	---KAGLADR---	LGQEKGGHAK	-----	YECPHCKIT	TAPDIK	TMQIHHEAKHP	-----
Y18788_Mt_	82	LSHKCSWCNKAFSSYQALGGH	ASHRK	AVMSATTAEDQITTTSSAVTSSAS-NGK	--NKTECSI	CHKSF	TGQALGGHKRCHYEG	SVGAG-AGAGS
AF539746_Ca_	100	MVYKCSWCGKGFSGYQALGGH	ASHRK	LVPGGDDQSTTSTTTNATGTTT	SVNGNMRSGRTECSI	CHKCF	TGQALGGHKRCHYD	GGIGNGNANSV
			C C	H H	C C	H H		
JK974084_Rp_	71	-----	KIPFEE	SKVTNLH	-----	ASSVP	ESSKPRPG	WRGSLKK---
ABU95041.1_Ta_	71	-----	KLFPD	PEKINLHG	STAAAPAAAAAV	STSKPKPG	WRGSLKKTG	DEQK
NP_566182.1_At_	71	-----	NIIYE	EESKLVNLH	AV-----	LAPVAES	KPKPG	IRGSLK
NP_197151_At_	71	-----	KLTYE	EPR--NL	HEA-----	LAAP	ESSKPKPG	IRGSLKK
AA92680.1_Pa_	71	-----	KIPYE	ESKIINKH	-----	ATVQ	ADPNKPRPG	IRGSLKK
Y18788_Mt_	177	---VTASEGVGLSHSHRDFDLNLP	AFPDP	SKKF-FVD	VEVFSPLPAAK	KPCLF	KL	EIPSHY--
AF539746_Ca_	200	SVGVTSS	SEGVG-STVSHRDFDLNIPAL	PEFWL	GFSGEDEVESPHPAKKS	SRLCL	LP	PKYELFQH---

B

			10	20	30	40	50	60	70	80	90	100
XP_002281411.1_Vv_	1	MAGLDLGTASRYI-HQLHRSDLHLQRPQ	-----	-----	-----	-----	DSDEDNNTN--RGG	AQYSGDHQDD-V	-----	AHHCLELIVS	-----	ANAGPCDI
NP_193515.1_At_	1	MAGLDLGTAFRYVNHQLHRPDLHLRHNS	-----	-----	-----	-----	SSD-DVTPG--AC	MGHFTVDD	-----	NHQCCLDLAS	GGGSGSSGGGGHG	SDV
XP_003524711.1_Gm_	1	MAGLDLGSASRFV-QMLHLPDLHLQNYQQP	-----	-----	-----	RHKRD	SBEQETPPN--P	TALAPFDND	-----	KSQCLELAS	-----	GPCDI
XP_003608618.1_Mt_	1	MQNIHR-----QNLHLQQQHHLQNHHD	-----	-----	-----	-----	SEEQESNRASV	GGAPFSSNEED	-----	RSQCLELGS	-----	AAGPGDV
JK974195_Rp_	1	MAGLDLGSASRFV-QMLHRPDLHLQHHHD	-----	-----	-----	-----	SEDQQQEPN--R	AGSLFSS	-----	RSQCLELGS	-----	AAGPGDV
XP_002320066.1_Pt_	1	MDPWSAHGRPLPP--PFITRDFHLHQFQ	-----	-----	-----	-----	HHQQQN	SEDEQSGNGD	-----	LNRRGKREH	-----	SSGSEET
XP_002509637_Rc_	1	MDPWTAHGHSLLPP--PFITRDFQLHHHQQQQF	SHHHQQQN	SEDEQSGSSSGA	SLNKSQKRER	DEGS	-----	EKELIP	-----	-----	-----	SGSSEI

			110	120	130	140	150	160	170	180	190	200
XP_002281411.1_Vv_	68	WARRRPRGRPPGSKNKPKPPVITRESANTLRAHILEVGN	CDWFDCVATYARRR	QRCICVLSGSGT	VTVNVS	LRQPA	AAAGAVTLHGRFEILSLSGSFLPP					
NP_193515.1_At_	80	WGRRRPRGRPPGSKNKPKPPVITRESANTLRAHILEVING	CDWFDCVATYARRR	QRCICVLSGSGT	VTVNVS	LRQPA	AAAGAVTLHGRFEILSLSGSFLPP					
XP_003524711.1_Gm_	72	WGRRRPRGRPPGSKNKPKPPVITRESANTLRAHILEVGS	SDWFDCVATYARRR	QRCICVLSGSGT	VTVNVS	LRQPA	AAAGAVTLHGRFEILSLSGSFLPP					
XP_003608618.1_Mt_	63	WGRRRPRGRPPGSKNKPKPPVITRESANTLRAHILEVAG	SDWFDCVATYARRR	QRCICVLSGSGT	VTVNVS	LRQPA	AAAGAVTLHGRFEILSLSGSFLPP					
JK974195_Rp_	68	WGRRRPRGRPPGSKNKPKPPVITRESANTLRAHILEVGS	ADWFDCVATYARRR	QRCICVLSGSGT	VTVNVS	LRQPA	AAAGAVTLHGRFEILSLSGSFLPP					
XP_002320066.1_Pt_	78	SARRRPRGRPPGSKNKPKPPITITRDSANALRSHVMEIAT	SDIMESVSTFARRR	QRCICVLSGSGT	VTVNVS	LRQPA	AAAGAVTLHGRFEILSLSGSFLPP					
XP_002509637_Rc_	81	TARRRPRGRPPGSKNKPKPPITITRDSANALRTHLMEVAD	CDIVESVATFARRR	QRCICVLSGSGT	VTVNVS	LRQPA	AAAGAVTLHGRFEILSLSGSFLPP					

AT-hook motifs

			210	220	230	240	250	260	270	280	290	300
XP_002281411.1_Vv_	168	PAPPGATSLTIIFLAGGQGGVVGGSVVGELTAAGPVIVIAASF	INVAAYERLPLEEBE	PLQMQQVVPQ	SSGSG	-----	GGINNPF	PDPSAGLPFFML				
NP_193515.1_At_	180	PAPPGATSLTIIFLAGGQGGVVGGSVVGELTAAGPVIVIAASF	INVAAYERLPLEEBE	-----	QQQLG	SSN	-----	NLFPEVAAGGGGLPFFML				
XP_003524711.1_Gm_	172	PAPPGATSLTIIFLAGGQGGVVGGSVVGELTAAGPVIVIAASF	INVAAYERLPLEEBE	QQQLIQSPATTSSQ	-----	-----	-----	MNNNNPFDPSSGLPFFML				
XP_003608618.1_Mt_	163	PAPPGATSLTIIFLAGGQGGVVGGSVVGELTAAGPVIVIAASF	INVAAYERLPLEEBE	QLMQAEAG	SSGSGSGG	-----	-----	MNNNNPFDPSSGLPFFML				
JK974195_Rp_	168	PAPPAASGLTIIFLAGGQGGVVGGSVVGELTASGPVIVIAASF	INVAAYERLPLEEBE	TAMTAPGL	SPGLVSPQSQPQQ	QQEQQFVDPNTN	-----	PHGV				
XP_002320066.1_Pt_	177	PAPPAASGLTIIFLAGGQGGVVGGSVVGELTASGPVIVIAASF	INVAAYERLPLEEBE	IESSTPML	SSPLSP	INNIGQQQ	-----	QQQQQLMQDPKTSLQQGL				
XP_002509637_Rc_	180	PAPPAATGLTIIFLAGGQGGVVGGSVVGELTASGPVIVIAASF	INVAAYERLPLEEBE	DAQLPMQ	SAIS	-----	-----	VGQQQHQQQQQLGEANAQLQQGL				

			310	320
XP_002281411.1_Vv_	259	PLNMP-NCQLPVDGWAEN	-----	SGARPPF
NP_193515.1_At_	268	PMNMQPNVQLPVEGWPEIN	-----	SGGRGPF
XP_003524711.1_Gm_	265	PLNMQ-NVQLP	-----	PF
XP_003608618.1_Mt_	258	PLNMQNVHQLPVDGWAENNNNSASRQPF	-----	-----
JK974195_Rp_	266	PNLLMSCQLPAEGYVCG	-----	SARPPF
XP_002320066.1_Pt_	277	PNLLNSVQLPAEAYWTGG	-----	RPPY
XP_002509637_Rc_	276	QPNLLNSIQLPAEAYWATG	-----	RPPY

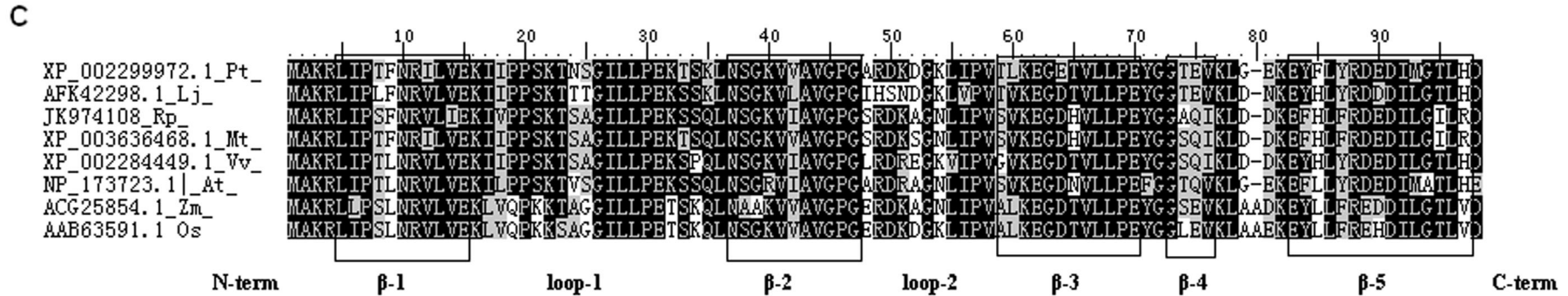


Figure S1 Clustal W multiple alignment of the amino acid sequences. Identical amino acid sequences are highlighted in black and sequence similarities in grey and conserved domains are boxed-in and conserved amino acid residues are marked in the text. (A) Deduced amino acid sequence of predicted C2H2-type zinc finger-containing protein (JK974084-Rp) is compared with other C2H2 type zinc finger plant proteins (sequences of *Arabidopsis thaliana* (NP_566182.1, NP_197151), *Triticum aestivum* (ABU95041.1), *Picea abies* (AAX92680.1), *Medicago truncatula* (EMBL accession no. Y18788) and *Capsicum annuum* (EMBL accession no. AF539746)). C2H2 conserved domains are boxed-in. JK974084 has only one C2H2 domain, which is different from *Medicago truncatula* (Y18788) with two C₂H₂ domains. (B) Multiple alignment of the predicted DNA-binding protein ESCAROLA-like (JK974195-Rp) amino acid sequence with other plant proteins. Sequences of *Glycine max* (XP_003524711.1) , *Vitis vinifera* (XP_002281411.1), *Medicago truncatula* (XP_003608618.1), *Arabidopsis thaliana* (NP_193515.1), *Catharanthus roseus* (ABL63118.1),

Populus trichocarpa (XP_002320066.1) and *Ricinus communis* (XP_002509637) are compared. Conserved domains of AT-hook motifs are boxed-in, which strongly suggests a DNA-binding function for the proteins as a whole. (C) JK974108-Rp, which encode a 10 kDa chaperonin compared with other plant cpn10 amino acid sequences. Sequences of *Medicago truncatula* (XP_003636468.1), *Vitis vinifera* (XP_002284449.1), *Arabidopsis thaliana* (NP_173723.1), *Populus trichocarpa* (XP_002299972.1), *Zea mays* (ACG25854.1), *Oryza sativa indica* (AAB63591.1) and *Lotus japonicus* (AFK42298.1) are compared. Predicted β strands are boxed-in.