BIOCELL
2020 44(1): 117-126

Tech Science Press

Dephosphorylated mutations affect the protein-protein interactions of ERF in *Populus simonii x P. nigra*

Yao SUN; Yao LI; Xin SUN; Qiong WU; Lei WANG*

Institute of Advanced Technology, Heilongjiang Academy of Sciences, Harbin, China

Key words: Dephosphorylation, PsnERF, Y2H, Secondary structure, Homology modeling

Abstract: Phosphorylation is a common type of post-translational modification (PTM). It plays a vital role in many cellular processes. The reversible phosphorylation and dephosphorylation affect protein structures and protein-protein interactions. Previously, we obtained five proteins that interact with ethylene-responsive factor (ERF) from the cDNA library of *Populus simonii x Populus nigra*. To further investigate the effect of dephosphorylation of PsnERF on its protein binding ability, we generated different phosphorylation states of PsnERF and demonstrated their protein binding capacity by the yeast two-hybrid assay (Y2H). The secondary structures and 3D structures of PsnERF, ERFm, TrunERF, and *psnerf*$^{197/198/202a}$ were predicted by homology modeling. The Y2H assay indicated that the deletion of serine-rich regions does not affect the interactions, while dephosphorylated mutations blocked the interactions. Homology modeling results suggested that the protein-binding activity was affected by dephosphorylation, and the S197/S198/S202 residues of PsnERF may be the key phosphorylation sites influencing its binding ability.

Introduction

Post-translational modification (PTM) refers to the covalent modification following protein synthesis. It is an essential process occurring prior to the formation of the mature protein. Phosphorylation is the most common PTM (Khoury *et al*., 2011), which plays a critical role in a wide range of cellular processes including cell metabolism, proliferation, differentiation, apoptosis, signal transduction, and gene expression (Clifford *et al*., 2004; Thingholm *et al*., 2009; Binz *et al*., 2014; Duan *et al*., 2015; Linke *et al*., 2015). Phosphorylation and dephosphorylation are catalyzed respectively by kinases and phosphatases. In eukaryotes, kinases phosphorylate proteins via the addition of a phosphate group to a polar uncharged R group of serine (Ser, S), threonine (Thr, T), and tyrosine (Tyr, Y) residues; consequently, this portion is turned into a negatively charged molecule (Fischer and Krebs, 1955; Krupa *et al*., 2004). Protein dynamics contribute to the conformational change of protein structures, and as a result, many enzymes and receptors will be activated and inactivated, which is common in many signal transduction pathways.

Protein kinases can specifically interact with corresponding residues on the substrate during certain conditions (Li *et al*., 2003; Sun *et al*., 2018). To clarify the phosphorylation process of the protein of interest, it is important to recognize its putative phosphorylation sites. However, most verification experiments are time-consuming. Additionally, high-throughput analysis is limited by the occurrences of false positive and false negative (Dou *et al*., 2014). To solve this problem, many bioinformatics tools have been developed as complementary methods, such as PhoScan (Li *et al*., 2010), NetPhos (Blom *et al*., 1999, 2004), Phospho. ELM (Diella *et al*., 2004). These tools not only recognize known phosphorylation sequences and kinases but find potential sites that have not yet been reported, which has become a hotspot in phosphorylation research.

There is a large volume of published studies describing that phosphorylation status could affect the functions of many enzymes and transcription factors in plants, including stress tolerance, transcriptional activity, and protein degradation (Ding *et al*., 2015; Wu and Li, 2017). It has also been suggested that dephosphorylation alters the subcellular localization of *Arabidopsis* HOS5 and eIF4AIII (Chen *et al*., 2013, 2015; Cui *et al*., 2016). Several AP2/ERF transcription factors, such as AtERF104 and OsEREBP1, are defined as substrates of mitogen-activated protein kinase (MAPKs) (Cheong *et al*., 2003; Gerit *et al*., 2009). BIN2, a negative regulator in the BR pathway, stabilizes AP2/ERF transcription factor TINY via phosphorylation and prevents expression of stress response genes under normal growth conditions in *Arabidopsis* (Xie *et al*., 2019). However, the phosphorylation mechanism and the phosphorylation-mediated functions for many other ethylene-responsive factors (ERFs) remain to be discovered. In this study, we predicted putative

*Address correspondence to: Lei Wang,
wleileiyu@163.com

Doi: 10.32604/biocell.2020.08242

phosphorylation sites and the serine-rich region of the ERF from *Populus simonii x P. nigra*. The protein binding capacity of dephosphorylated ERF and the ERF lacking the serine-rich region (ERFm and TrunERF, respectively) were characterized by their interactions with five proteins identified in our previous study (Wang *et al*., 2018). Then the structural differences triggered by the mutations were analyzed by homology modeling. This study attempts to elucidate the effect of dephosphorylation of PsnERF in protein binding and signal transduction networks.

Materials and Methods

Bioinformatics analysis

Multiple sequence alignments and the construction of a phylogenetic tree based on the amino acid sequences of PsnERF (Yao *et al*., 2016) and 20 ERFs from other plant species were performed using MEGA 7.0. PsnERF phosphorylation site prediction was carried out by NetPhos3.1 (http://www.cbs.dtu.dk/services/NetPhos/) (Blom *et al*., 1999, 2004). Serine-rich region prediction was carried out by Motif Scan (https://myhits.isb-sib.ch/cgi-bin/motif_scan) (Sigrist *et al*., 2010). According to the prediction results, *ERFm* was generated by replacing all the S/T by alanine (A), while the deletion of the serine-rich region was performed for *TrunERF*. The secondary structure was estimated by SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page = npsa_sopma.html) (Geourjon and Deléage, 1996). Protein structure homology modeling was obtained in three different ways: SWISS-MODEL (https://swissmodel.expasy.org/interactive) (Guex *et al*., 2009; Benkert *et al*., 2011; Bertoni *et al*., 2017; Bienert *et al*., 2017; Waterhouse *et al*., 2018), Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (Kelley *et al*., 2015), and I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy *et al*., 2010; Yang *et al*., 2015a, 2015b).

Gene cloning and vector construction

The nucleotide sequences of *ERFm* and *TrunERF*, constructed in pUC57-Kan, were synthesized by GENEWIZ (Suzhou, China). Additional restriction sites of *Eco*RI at the 5' end and *Bam*HI at the 3' end were added by PCR using the following primers: 5'-TTCATTCGAATTCCATGTGCGTATT-3'/5'-CGGGATCCCGATCGGGGAAATT-3' for *ERFm* and 5'- TTCATTCGAATTCCATGTGCGTATT-3'/5'- CGGGATCCCGTTAACCAGTGGAGGAAGGAC-3' for *TrunERF*. Following double enzyme digestion, the fragments were cloned into the pGBKT7 vector.

Transcriptional activation activity assay

All the GAL4-BD fusion constructs and empty pGBKT7 vector were separately transformed into AH109 yeast competent cells. The cells plated on synthetic dropout (SD)/-Trp (Coolaber) were then resuspended in 10 μL ddH_2O and transferred to SD/-Trp medium containing X-α-gal.

Yeast two-hybrid assay

The bait vectors were transformed into the Y2HGold yeast strain, which was followed by the transformation of pGADT7-prey plasmids. The blue colonies which grew on SD/-Trp/-Leu/X-gal/AbA (DDO/X/A) medium were resuspended in 10 μL ddH_2O and then transferred to SD/-Ade/-His/-Trp/-Leu/X-gal/AbA (QDO/X/A) medium.

Results

Analysis of PsnERF amino acid sequences

The ORF of PsnERF encodes a putative protein of 445 amino acids (aa). It contains a 60-aa AP2/ERF domain. The 14th residue is alanine and the 19th is aspartic acid, which is the typical feature of the ERF subfamily (Fig. 1(A)). Phylogenetic analysis revealed that PsnERF shares high sequence similarity with PtrERF76 and PeERF (Fig. 1(B)). PtrERF76 has been identified as the transcriptional regulator in plant stress pathways (Yao *et al*., 2016). PsnERF might play a similar role in transcriptional activation and plant stress response.

Phosphorylation sites and serine-rich region prediction results

17 S residues and one T residue were predicted as the putative phosphorylation sites of PsnERF by NetPhos3.1 (Fig. 1(C)). The detailed information for each site is listed in Tab. 1. The phosphorylation sites of some other ERFs that have been identified by experiments in a previous study can be found in Fig. 1(D), such as AtERF6 and AtERF104, which share homology N-terminal motif with plant MAPK substrates (Gerit *et al*., 2009; Meng *et al*., 2013). It has also been proved that Ser-62 is the phosphorylation site of AtERF110, and phosphorylation of AtERF110 is closely related to ethylene pathways (Zhu *et al*., 2013). Compared with these ERFs, no similar conserved phosphorylation motifs were observed, nor were any specific protein kinases indicated for PsnERF. Two serine-rich regions (aa 63-133 and aa 366-402) identified by Motif Scan is also shown in Fig. 1(C). Both regions were deleted to generate TrunERF; meanwhile, dephosphorylation was mimicked by mutating all the predicted sites (S/T to A).

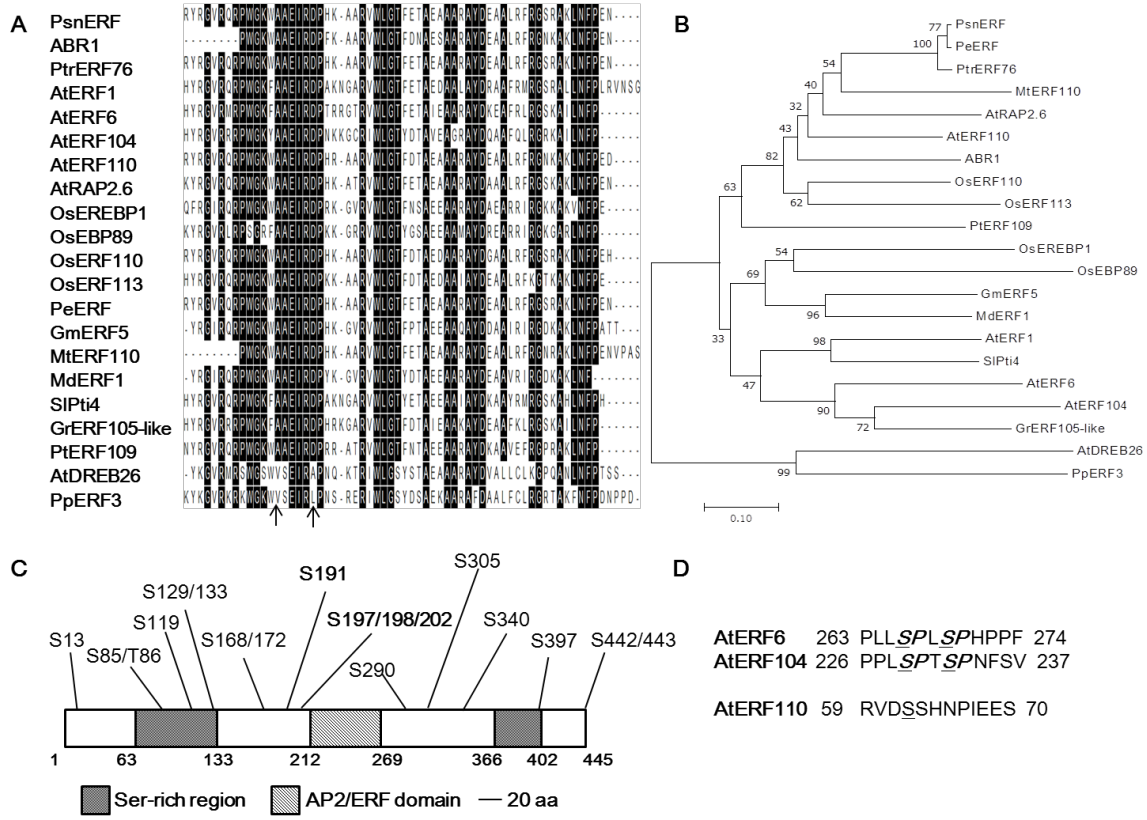

FIGURE 1. Analysis of PsnERF amino acid sequence. (A) Alignment of AP2/ERF domain of PsnERF with other ERFs, including *Populus trichocarpa* PtrERF76 (Potri.005G195000.1); *Populus euphratica* PeERF (XP_011044262.1); *Arabidopsis thaliana* ABR1 (AT5G64750.1), AtERF1 (AT4G17500.1), AtDREB26 (AT1G21910.1), AtERF6 (AT4G17490.1), AtERF104 (AT5G61600.1), AtERF110 (AT5G50080.2) and AtRAP2.6 (AT1G43160.1); *Oryza sativa L.* OsEREBP1 (Os02g54160.1), OsEBP89 (Os03g08470.1), OsERF110 (Os12g0168100) and OsERF113 (Os04g0398000); *Glycine max* GmERF5 (Glyma.09G041500.1); *Medicago truncatula* MtERF110 (MTR_1g098460); *Malus domestica* MdERF1 (MDP0000128979); *Solanum lycopersicum* SlPti4 (Solyc05g052050.1.1); *Prunus persica* PpERF3 (Prupe.7G194400); *Gossypium raimondii* GrERF105-like (Gorai.003G112200.1); *Poncirus trifoliata (L.) Raf* PtERF109 (MH779873.1). The arrow indicates the 14th and 19th amino acid, the core residues to distinguish ERF and DREB subfamily. Conserved sites at 90% level were marked with black background. (B) A phylogenetic tree based on the amino acids of PsnERF and other ERFs was constructed using neighbor-joining method. (C) Diagram displaying the putative phosphorylation sites and serine-rich region of PsnERF. The phosphorylation sites and serine-rich region were respectively predicted by NetPhos3.1 (http://www.cbs.dtu.dk/services/NetPhos/) and Motif Scan (https://myhits.isb-sib.ch/cgi-bin/motif_scan). The predicted phosphorylation sites were marked, and AP2/ERF conserved domain (aa 212-269) and two serine-rich regions (aa 63-133, 366-402) were shown as different boxes. The bar indicates the length of 20 amino acids. (D) Scheme of the phosphorylation sites of other ERFs that have been verified by experiments. The phosphorylation sites were underlined, and the conserved MAPK-phosphorylation sites were shown in bold and italics.

TABLE 1 Prediction of phosphorylation sites of PsnERF

The prediction analyzes the phosphorylation probability of each serine, threonine and tyrosine. For each residue only the best prediction scoring >0.9 will be displayed. Details such as residue position, residue types, sequence context of 9 residues, scores and kinases for the positive predictions were listed

Position	Site	Context	Score	Kinase	Position	Site	Context	Score	Kinase
13	S	PRGN<u>S</u>GEYS	0.985	unsp	197	S	TATP<u>S</u>SETA	0.958	unsp
85	S	GHRG<u>S</u>TSDW	0.996	unsp	198	S	ATPS<u>S</u>ETAS	0.962	unsp
86	T	HRGS<u>T</u>SDWG	0.914	unsp	202	S	SETA<u>S</u>LGET	0.990	unsp
119	S	TSPA<u>S</u>PSLS	0.952	unsp	290	S	QVPV<u>S</u>RSQL	0.979	unsp
129	S	YSST<u>S</u>GSGS	0.929	unsp	305	S	QPIS<u>S</u>PRQQ	0.927	unsp
133	S	SGSG<u>S</u>WIGQ	0.993	unsp	340	S	QLLQ<u>S</u>SGDF	0.958	unsp
168	S	DFRS<u>S</u>LGDS	0.998	unsp	397	S	SSTL<u>S</u>PSAS	0.994	unsp
172	S	SLGD<u>S</u>SSSG	0.952	unsp	442	S	SRRP<u>S</u>STG--	0.996	unsp
191	S	TLVF<u>S</u>TTAT	0.924	unsp	443	S	RRPS<u>S</u>TG--	0.989	unsp

Dephosphorylation affected the transcriptional activation effect and protein binding ability of PsnERF

ERFm and *TrunERF* genes were cloned into the pGBKT7 vector. The expected bands amplified by PCR confirmed the insertion of the fragments (Fig. S1). The recombinant bait vectors were transformed into AH109 and Y2HGold competent cells, respectively. The colonies containing pGBKT7-*ERFm* or pGBKT7-*TrunERF* grew well on SD/-Trp medium, revealing that the baits were not toxic to the host. However, in AH109, BD-ERFm was unable to activate the expression of reporter *MEL1*, which was the same as the negative control (Fig. 2(A)). This result revealed that PsnERF and TrunERF have trans-activation ability, but dephosphorylation of PsnERF impedes its binding to the *cis*-acting elements.

The interactions between PsnERF and five proteins were identified in our previous work (Wang *et al*., 2018). In this study, no self-activation for the recombinant baits were observed, or were any interactions between ERFm and prey proteins detected. By comparison, the normal growth of *TrunERF* yeast cells on the selection plates, indicated that the interactions were not influenced (Figs. 2(B), S(2) and S(3)). These results imply that dephosphorylation blocks the protein interactions between PsnERF and prey proteins, while the interactions are not disturbed by the deletion of serine-rich regions.

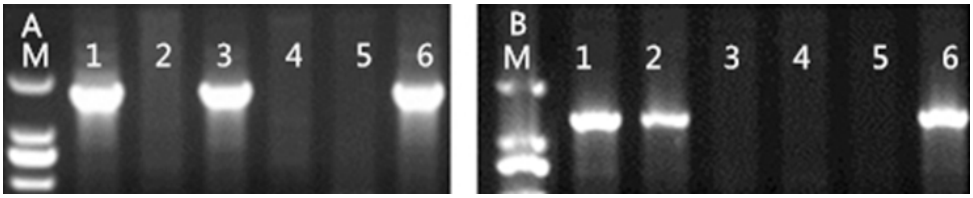

FIGURE S1. Identification of recombinant bait vector by colony PCR. M5-BD/M3-BD were used as primers A:pGBKT7-ERFm B: pGBKT7-trunERF. M: DL2000. 1-6: Detected samples.

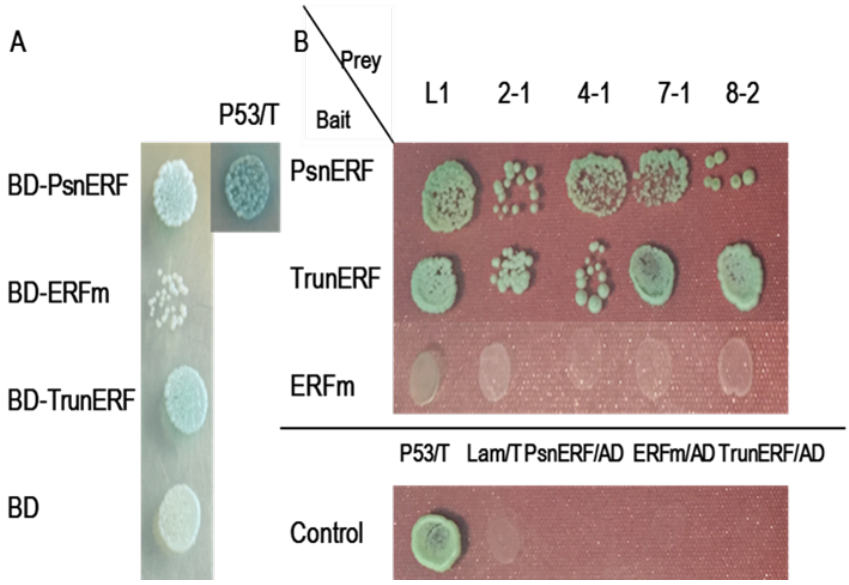

FIGURE 2. Trans-activation activity of PsnERF and protein interactions with preys. (A) Trans-activation assay on SD/-Trp/x-gal medium. BD: the plasmid containing the GAL4 DNA-binding domain (pGBKT7). P53/T: positive control. (B) Y2H assay on QDO/X/A medium. The blue colonies on DDO/X/A were resuspended in 10 μL ddH_2O and transferred to QDO/X/A plates. For ERFm/preys, few blue colonies on DDO/X/A plates were unable to grow on QDO/X/A after restreaking. P53/T: positive control. Lam/T: negative control. AD: empty prey vector containing GAL4 transcriptional activation domain (pGADT7). Prey proteins: L1: Galactinol–sucrose galactosyl transferase 2; 2-1: Senescence-associated protein; 4-1: Unknown protein; 7-1: Phosphoinositide-specific phospholipase C family protein; 8-2: MLP-like protein 328.

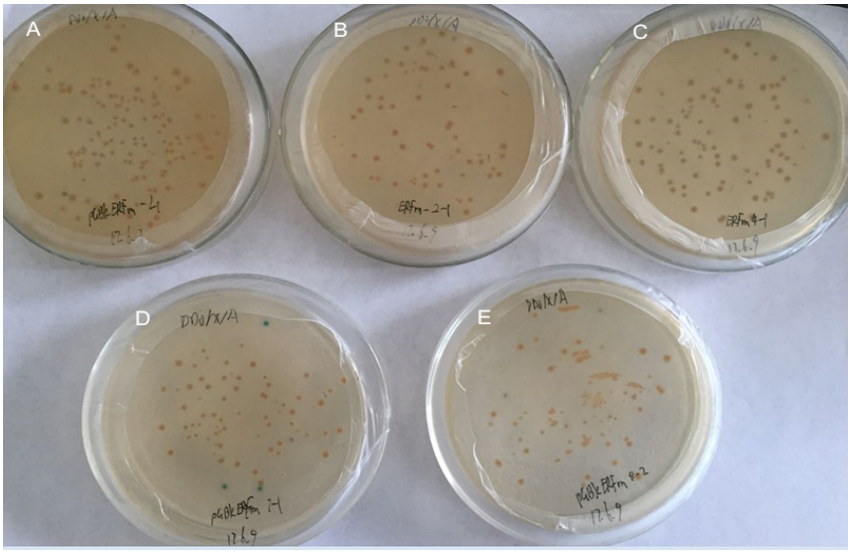

FIGURE S2. Y2H assay for ERFm-BD/Prey-AD on DDO/X/A plates. Prey proteins: A: Galactinol-sucrose galactosyl transferase 2. B: Senescence-associated protein. C: Unknown protein. D: Phosphoinositide-specific phospholipase C family protein. E: MLP-like protein 328. Few blue colonies were unable to grow on QDO/X/A after restreaking.

Secondary structure prediction

The dynamics of protein dephosphorylation would lead to conformational changes. Hence, we suggested that the defect in the protein-protein interactions might result from an inability to recognize the protein-binding domain. To clarify the conformational differences, we estimated the secondary structure of ERFm, PsnERF, and TrunERF by SOPMA. In parallel alignment with PsnERF, a remarkable structural change for TrunERF was observed, which was predictable due to the loss of over 100 amino acids. Unexpectedly, most structures of ERFm were not severely affected by dephosphorylation. It was also noted that the secondary structures of the AP2/ERF conserved domain in three sequences were roughly the same (aa 212-269 in PsnERF and ERFm and aa 141-198 in TrunERF).

Disregarding the deleted region, some unique structures in ERFm, were supposed to be the key factors affecting the protein binding ability, such as the alpha helix region (marked by the red arrow) located near aa 200 (Fig. 3). Four phosphorylation sites located around this α-helix region can be divided into two subgroups: S191 and S197/S198/S202 (shown in bold in Fig. 1(C)), which are likely to be the substrates of kinases. To further analyze the putative sites, we estimated the secondary structure of *psnerf*^{191a} and *psnerf*$^{197/198/202a}$. As shown in Fig. 3, *psnerf*^{191a} showed higher similarity with PsnERF rather than ERFm, while a similar secondary structure was observed between ERFm and *psnerf*$^{197/198/202a}$, especially the α-helix region at aa 200. These results suggested that S197/S198/S202 residues of PsnERF might be the major phosphorylation sites, and dephosphorylated mutations may attenuate the protein binding ability.

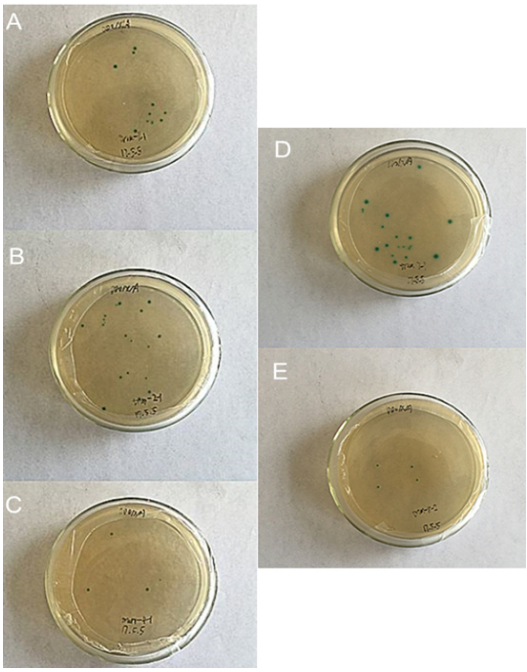

FIGURE S3. Y2H assay for TrunERF-BD/Prey-AD on DDO/X/A plates. Prey proteins: A: Galactinol-sucrose galactosyl transferase 2. B: Senescence-associated protein. C: Unknown protein. D: Phosphoinositide-specific phospholipase C family protein. E: MLP-like protein 328.

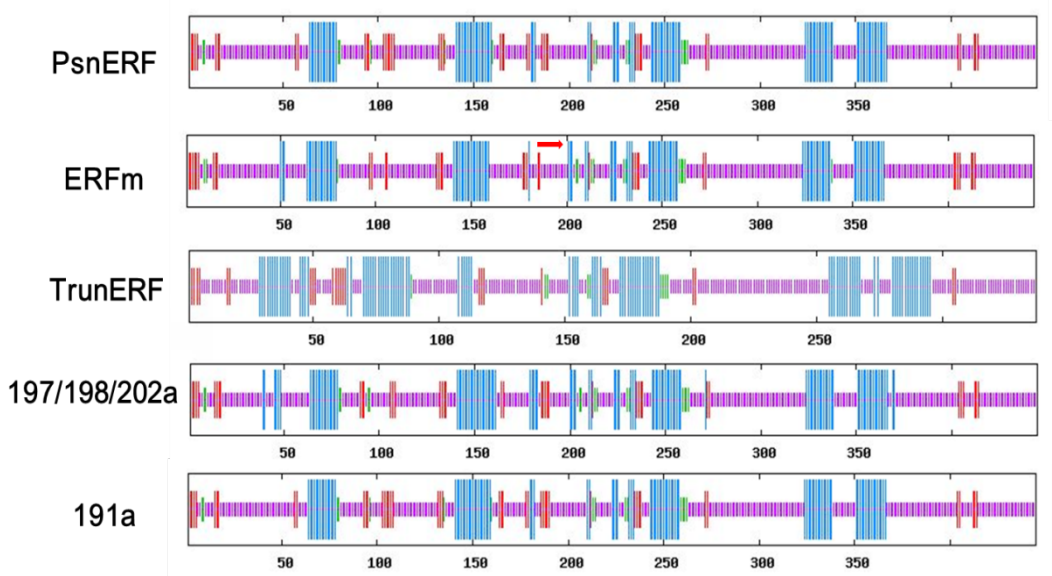

FIGURE 3. The secondary structure of PsnERF, ERFm, TrunERF, *psnerf*$^{197/198/202a}$ *and psnerf*^{191a} estimated by SOPMA. The red arrow indicates the missing alpha helix in PsnERF (aa 200) and TrunERF (aa 130). Blue: alpha helix. Red: extended strand. Green: beta turn. Pink: random coil.

Homology modeling

To further investigate the effect of S197/S198/S202 mutations, protein structures were analyzed by homology modeling, i.e., SWISS-MODEL, Phyre2, and I-TASSER. These tools can directly detect homologous sequences of query peptides and compare them with the reference model for protein structure analysis. The secondary structures of PsnERF, ERFm, TrunERF, and *psnerf*$^{197/198/202a}$ in I-TASSER are displayed in Fig. 4. Among all the models, there were some subtle differences around the ERF conserved domain region, such as the helix region at aa 150 (aa 80 in TrunERF, Fig. 4(C)). Additionally, in ERFm (Fig. 4(B)) and *psnerf*$^{197/198/202a}$ (Fig. 4(D)), the helix region around aa 290 was more pronounced. The 3D structure of PsnERF, ERFm, and *psnerf*$^{197/198/202a}$ were quite similar to each other, except the aa 255-361 region, where a pair of "tweezers" was formed. Interestingly, the "tweezers" of PsnERF and *psnerf*$^{197/198/202a}$ were open compared to ERFm (Fig. 5(A)). Despite the structural similarity between PsnERF and *psnerf*$^{197/198/202a}$, the determined calcium-ion ligand binding sites were different (Tab. 2), indicating a remarkable correlation between ligand binding affinity and mutations at S197/S198/S202.

For Phyre2, erf096 (PDB: 5wx9) was identified as the template for all input sequences, with model confidence of 99%. All the predictions output a similar 3D model, except *psnerf*$^{197/198/202a}$, where two helix regions in the C-terminus (orange and red) were missing (Fig. 5(B)). Clearly, both predictions mentioned above emphasize the importance of S197/S198/S202. However, a different result was displayed in SWISS-MODEL. In this method, different templates were applied for structure evaluation (i.e., PDB id:3gcc for PsnERF, TrunERF, and *psnerf*$^{197/198/202a}$, and PDB id:1gcc for ERFm). Accordingly, the predicted C-terminal structure of ERFm (ssccccscc) was different from the other three peptides (csscccscc), as shown in Tab. 3. This result does not explain the effect of dephosphorylation of S197/S198/S202 on the structure of the aligned domain; nevertheless, it is possible that the disordered region not displayed in SWISS-MODEL may be affected by mutations. Overall, these results indicate a link between protein structure and dephosphorylation for PsnERF. Some residues, especially S197/S198/S202, may be key sites that can affect protein binding activity.

TABLE 2

Calcium binding sites of PsnERF, ERFm, and *psnerf*$^{197/198/202a}$

	PDB Hit	Ligand Binding Site Residues
PsnERF	3iprD	227,228
ERFm	4jo2H	223,224
psnerf$^{197/198/202a}$	3sqgA	249,250,252

TABLE 3

C-terminal structure of PsnERF, ERFm, TrunERF and *psnerf*$^{197/198/202a}$ predicted by SWISS-MODEL. c = loop or irregular, s = bend

	PDB id	C-terminal structure
PsnERF	1gcc	csscccscc
ERFm	3gcc	ssccccscc
TrunERF	1gcc	csscccscc
psnerf$^{197/198/202a}$	3gcc	csscccscc

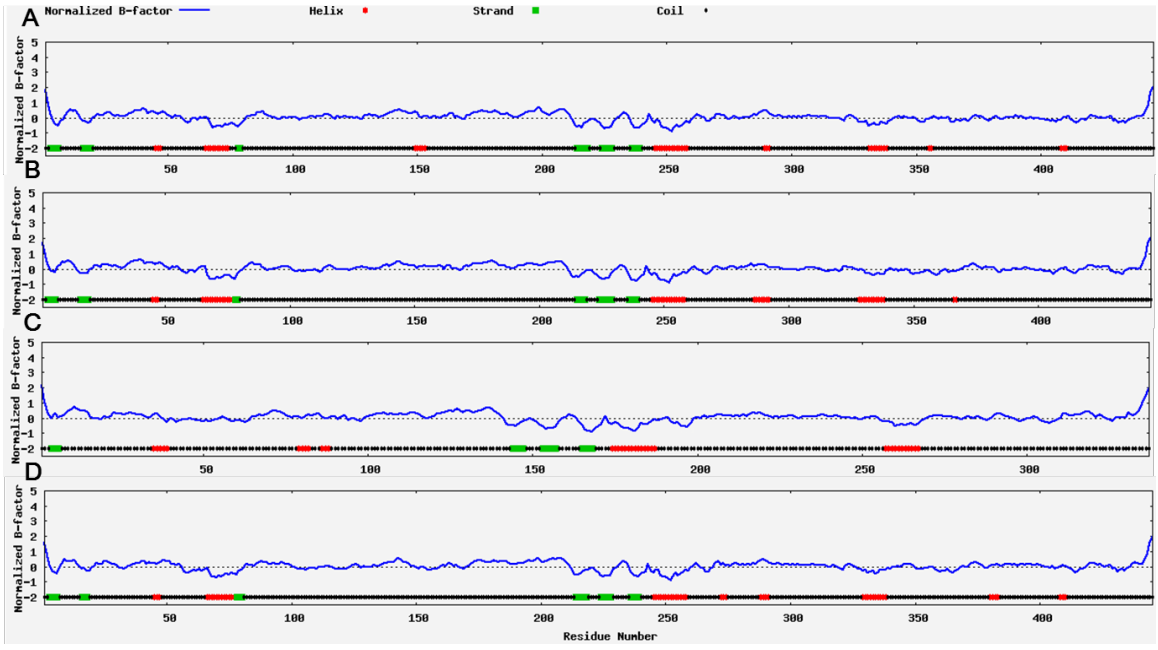

FIGURE 4. The secondary structure of PsnERF (A), ERFm (B), TrunERF (C) and *psnerf*$^{197/198/202a}$ (D) estimated by I-TASSER. X-axis indicates the residue number, and Y-axis is the normalized B-factor. The bars on the X-axis represent different structures displayed at the top of the illustration. The reported B-factor in this figure corresponds to the normalized B-factor of the target protein, defined by B = (B'-u) / s, where B' is the raw B-factor value, u and s are respectively the mean and standard deviation of the raw B-factors along the sequence. (A) PsnERF. (B) ERFm. (C) TrunERF. (D) *psnerf*$^{197/198/202a}$.

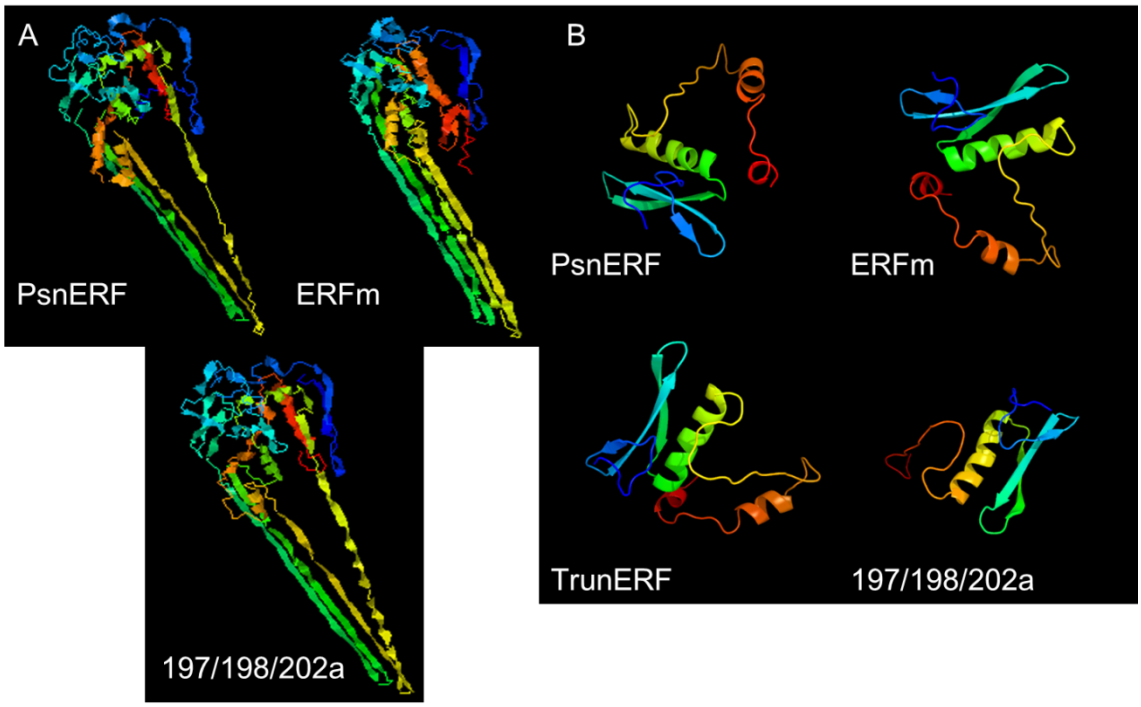

FIGURE 5. 3D structure predicted by I-TASSER (A) and Phyre2 (B). (A) The illustrations are displayed in the rainbow style, and the "tweezers" in each peptide are located from yellow to orange. (B) Only aligned sequences were displayed by Phyre2, and the rest were reported as disordered, which cannot be meaningfully predicted.

Discussion

Phosphorylation status influences protein biological functions by switching the relay signals on and off in plants. In poplar, several phosphorylation substrates and phosphorylation-mediated regulations have been identified. MAPK, which preferentially phosphorylates MBP, is suggested to play a central role in the crosstalks between many biotic and abiotic stresses in *Populus trichocarpa x P. deltoides* (Hamel *et al.*, 2005). Histidine-aspartate kinase (HK1), the binding partner of histidine phosphotransfer proteins, can act as an osmosensor and transfer the phosphosignal in response to drought stress (Héricourt *et al.*, 2013, 2016). Moreover, monolignol biosynthesis of *P. trichocarpa* can be regulated by the reversible phosphorylation of PtrAldOMT2 (Wang *et al.*, 2015). It has also been reported that several ERFs' functions are related to phosphorylation. For example, tomato ERF Pti4, which can be induced by ethylene and pathogen infection, can be phosphorylated by the Pto kinase. This reaction will enhance the GCC-box binding ability of Pti4, regulating the expression of many other defense-related genes (Gu *et al.*, 2000, 2002). Similarly, phosphorylation of *Arabidopsis* ERF6 by MPK3/MPK6 induces resistance to pathogen attack (Meng *et al.*, 2013). Also, TaERF1 is identified as the substrate of TaMAPK1, implying its role in TaMAPK1 cascades in multiple stress responses (Xu *et al.*, 2007). Besides, the phosphorylation status of Ser-62 in ERF110 is related to the bolting time of *Arabidopsis* (Zhu *et al.*, 2013).

In this study, the protein sequence alignment and construction of the phylogenetic tree were carried out between ERFs in *P. simonii x P. nigra* and other plant species. The results suggested that PsnERF is a member of the ERF TF family. It contains an AP2/ERF conserved domain, which shares a high similarity with PtrERF76 and PeERF (Figs. 1(A)-1(B)). These ERFs provide the reference for PsnERF function prediction. It has been reported that PtrERF76 is an ethylene signaling-dependent TF that could regulate the expression of genes from diverse subsets (Caroline *et al.*, 2018). The connection between PtrERF76 and salt stress response was also indicated in a previous study (Yao *et al.*, 2016). We also predicted the putative phosphorylation sites and serine-rich region of PsnERF via the bioinformatics method (Tab. 1, Fig. 1(C)). No similar conserved phosphorylation motifs that have been identified by experiments were observed for PsnERF (Fig. 1(D)). Dephosphorylated PsnERF was mimicked by S/T to A mutations, which introduce a nonphosphorylated amino acid dynamically similar to the polar uncharged S/T (Barnes and Gray, 2003). Meanwhile, TrunERF was generated by the deletion of serine-rich regions. The transcriptional activation activities of the mutated PsnERF were assayed in yeast strain AH109. The yeast colonies containing ERFm did not turn blue when the medium was added with X-α-gal (Fig. 2(A)), implying that PsnERF failed to bind to the *cis*-acting element after dephosphorylation, while the trans-activation activity was not diminished by the deletion of serine-rich regions. Similarly, interactions between ERFm, TrunERF, and five proteins previously screened from *P. simonii x P. nigra* cDNA library were studied by Y2H assay (Fig. 2(B)). Although a similar secondary structure pattern of ERF conserved domain (aa 212-269 in PsnERF and ERFm, and aa 141-198 in TrunERF) was displayed in these three peptides (Fig. 3), no interactions were detected between ERFm and preys, implicating that the dynamics of PsnERF dephosphorylation will block the interactions. The results demonstrate a similar structural model between *psnerf*$^{197/198/202a}$ and ERFm, indicating that S197/S198/S202 might be key residues for

PsnERF phosphorylation.

Following secondary structure prediction, homology modeling was applied to explore the effect of mutations at these sites, which can align the query sequences to the template and resemble the structural model of the query target. Unlike other methods that can only identify structure based on homologous sequences, I-TASSER can remodel the full-length input sequence, including the aligned and unaligned region. It was substantiated by I-TASSER that the secondary and 3D structure of PsnERF could be affected by dephosphorylation (Figs. 4 and 5). Calcium ions were among the top five ligand results of all the target sequences, whereas the calcium ligand binding sites of PsnERF, ERFm, and *psnerf*$^{197/198/202a}$ were different (Tab. 2). It can thus be suggested that the protein structure and binding affinity of the query sequences are affected by dephosphorylation, especially S197/S198/S202. These results corroborate the finding from Phyre2, which demonstrate that mutations of S197/S198/S202 result in the absence of two helix regions (Fig. 5(B)).

However, the results of SWISS-MODEL differ from the methods mentioned above. That is, the aligned domain structure of *psnerf*$^{197/198/202a}$ was the same as PsnERF and TrunERF, and only ERFm exhibited a different structural feature in the C-terminal Tab. (3). This inconsistency may be due to the fact that only aligned sequence models are obtained, so it is unknown whether the disordered region has changed. The different structures listed in previous methods are all located at the "disordered" region, such as the helix region around the AP2/ERF conserved domain, the "tweezers" in I-TASSER, and the missing helix in Phyre2, Even so, this result still suggests a link between protein structure and dephosphorylation in PsnERF.

It should be noted that for the ligand-binding site prediction in I-TASSER, the peptide-ligand binding item was only reported in *psnerf*$^{197/198/202a}$. Therefore, the protein binding sites were not deduced in the prediction. It is also noted that the protein interactions were not affected in TrunERF (Figs. 2(B) and S(3)), despite that the deletion of the serine-rich region had remarkable effects on the secondary structure (Fig. 3). Initially, we considered that the protein binding domain of PsnERF was not located in the deletion region. However, it is possible that the serine-rich regions may act as PTM regulators and the second candidate protein-binding domain. This region may recognize and modify the target substrates following the interactions between the primary binding domain and its partners. Hence, this assay cannot deny the protein binding ability of the serine-rich regions of PsnERF.

In sum, we predicted 18 putative phosphorylation sites and serine-rich regions of PsnERF and generated the yeast constructs of dephosphorylated ERF and ERF lacking serine-rich regions. The transcriptional activation activity assay indicated that dephosphorylation may impede the PsnERF's binding to the *cis*-acting element. The Y2H assay revealed that the ERFm cannot interact with the preys, while serine-rich region deletion does not affect the protein interactions. Homology modeling results suggested that S197/S198/S202 may be the key phosphorylation sites that affect its protein binding ability. Areas of future study include the phosphorylation status, protein binding region, and corresponding kinases involved with PsnERF.

Acknowledgement

This work was supported by the Natural Science Foundation of Heilongjiang Province of China (LH2019C059).

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

Barnes MR, Gray IC (2003). Bioinformatics for geneticists. Wiley, London.

Benkert P, Biasini M, Schwede T (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**: 343-350.

Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* **7**: 10480.

Bethke G, Unthan T, Uhrig JF, Pöschl Y, Gust AA, Scheel D, Lee J (2009). Flg22 regulates the release of an ethylene response factor substrate from MAP kinase 6 in *Arabidopsis thaliana* via ethylene signaling. *Proceedings of the National Academy of Sciences* **106**: 8067-9072.

Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T (2006). The SWISS-MODEL Repository-new features and functionalities. *Nucleic Acids Research* **45**: 313-319.

Binz SK, Sheehan AM, Wold MS (2004). Replication protein A phosphorylation and the cellular response to DNA damage. *DNA Repair* **3**: 1015-1024.

Blom N, Gammeltoft S, Brunak S (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology* **294**: 1351-1362.

Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S (2010). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**: 1633-1649.

Carolin S, Bernard W, Soile JL, Björn S, Nicolas D, Judith F, Hannele T (2018). Ethylene-related gene expression networks in wood formation. *Frontiers in Plant Science* **9**: 272.

Chen T, Cui P, Chen H, Ali S, Zhang S, Xiong L (2013). A KH-domain RNA-binding protein interacts with FIERY2/CTD phosphatase-like 1 and splicing factors and is important for pre-mRNA splicing in *Arabidopsis*. *PLoS Genetics* **9**: e1003875.

Chen T, Cui P, Xiong L (2015). The RNA-binding protein HOS5 and serine/arginine-rich proteins RS40 and RS41 participate in miRNA biogenesis in *Arabidopsis*. *Nucleic Acids Research* **43**: 8283-8298.

Cheong YH, Moon BC, Kim JK, Kim CY, Kim MC, Kim IH, Park CY, Kim JC, Park BO, Koo SC, Yoon HW, Chung WS, Lim CO, Lee SY, Cho MJ (2003). BWMK1, a rice mitogen-activated protein kinase, locates in the nucleus and mediates pathogenesis-related gene expression by activation of a transcription factor. *Plant Physiology* **132**: 1961-1972.

Clifford DM, Marinco SM, Brush GS (2004). The meiosis-specific protein kinase Ime2 directs phosphorylation of replication protein A. *Journal of Biological Chemistry* **27**: 6163-6170.

Cui P, Chen T, Qin T, Ding F, Wang Z, Chen H, Xiong L (2016). The RNA polymerase II C-terminal domain phosphatase-like protein FIERY2/CPL1 interacts with eIF4AIII and is essential for nonsense-mediated mRNA decay in *Arabidopsis*. *Plant Cell* **28**: 770-785.

Diella F, Cameron S, Gemünd C, Linding R, Via A, Kuster B, Sicheritz-Pontén T, Blom N, Gibson TJ (2004). Phospho. ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* **5**: 79.

Ding S, Zhang B, Qin F (2015). *Arabidopsis* RZFP34/CHYR1, a ubiquitin E3 ligase, regulates stomatal movement and drought tolerance via SnRK2.6-mediated phosphorylation. *Plant Cell* **27**:3228-3244.

Dou Y, Yao B, Zhang C (2014). PhosphoSVM: prediction of phosphorylation sites by integrating various protein sequence attributes with a support vector machine. *Amino Acids (Vienna)* **46**: 1459-1469.

Duan JJ, Lozada AF, Gou CY, Xu J, Chen Y, Berg DK (2015). Nicotine recruits glutamate receptors to postsynaptic sites. *Molecular and Cellular Neuroscience* **68**: 340-349.

Fischer EH, Krebs EG (1955). Conversion of phosphorylase b to phosphorylase a in muscle extracts. *Journal of Biological Chemistry* **216**: 121-132.

Geourjon C, Deléage G (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer Applications in the Biosciences: CABIOS* **11**: 681-684.

Guex N, Peitsch MC, Schwede T (2010). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30**: 162-173.

Gu YQ, Yang C, Thara VK, Zhou J, Martin GB (2000). Pti4 is induced by ethylene and salicylic acid, and its product is phosphorylated by the Pto kinase. *Plant Cell* **12**: 771-785.

Gu YQ, Wildermuth MC, Chakravarthy S, Loh YT, Yang C, He X, Han Y, Martin GB (2002). Tomato transcription factors Pti4, Pti5, and Pti6 activate defense responses when expressed in *Arabidopsis*. *Plant Cell* **14**: 817-831.

Hamel LP, Miles GP, Samuel MA, Ellis BE, Séguin A, Beaudoin N (2005). Activation of stress-responsive mitogen-activated protein kinase pathways in hybrid poplar (*Populus trichocarpa x Populus deltoides*). *Tree Physiology* **25**: 277-288.

Héricourt F, Chefdor F, Bertheau L, Tanigawa M, Maeda T, Guirimand G, Courdavault V, Larcher M, Depierreux C, Bénédetti H, Morabito D, Brignolas F, Carpin S (2014). Characterization of histidine-aspartate kinase HK1 and identification of histidine phosphotransfer proteins as potential partners in a *Populus* multistep phosphorelay. *Plant Physiology* **149**: 188-199.

Héricourt F, Chefdor F, Djeghdir I, Larcher M, Lafontaine F, Courdavault V, Auguin D, Coste F, Depierreux C, Tanigawa M, Maeda T, Glévarec G, Carpin S (2016). Functional divergence of poplar histidine-aspartate kinase HK1 paralogs in response to osmotic stress. *International Journal of Molecular Sciences* **17**: 2061.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**: 845-858.

Khoury GA, Baliban RC, Floudas CA (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific Reports* **1**: 90.

Krupa A, Preethi G, Srinivasan N (2004). Structural modes of stabilization of permissive phosphorylation sites in protein kinases: distinct strategies in Ser/Thr and Tyr kinases. *Journal of Molecular Biology* **339**: 1025-1039.

Linke D, Koudelka T, Becker A, Tholey A (2015). Identification and relative quantification of phosphopeptides by a combination of multi-protease digestion and isobaric labeling: Isobaric labeling of multi-protease digested phosphoproteins. *Rapid Communications in Mass Spectrometry* **29**: 919-926.

Li L, Shakhnovich EI, Mirny LA (2003). Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 4463-4468.

Li T, Li F, Zhang X (2010). Prediction of kinase-specific phosphorylation sites with sequence features by a log-odds ratio approach. *Proteins: Structure, Function, and Bioinformatics* **70**: 404-414.

Meng X, Xu J, He Y, Yang KY, Mordorski B, Liu Y, Zhang S (2013). Phosphorylation of an ERF transcription factor by *Arabidopsis* MPK3/MPK6 regulates plant defense gene induction and fungal resistance. *Plant Cell* **25**: 1126-1142.

Roy A, Kucukural A, Zhang Y (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **5**: 725-738.

Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N (2010). PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Research* **38**: D161-166.

Sun Y, Huang GQ, Li Y, Xue JY, Wu Q, Wang L (2018). Comparison of protein phosphorylation site prediction tools. *Proceedings of International Conference on Biological Sciences & Technology (BST 2017)*: 148-154

Thingholm TE, Jensen ON, Larsen MR (2009) Analytical strategies for phosphoproteomics. *Proteomics* **9**: 1451-1468.

Wang JP, Chuang L, Loziuk PL, Chen H, Lin YC, Shi R, Qu GZ, Muddiman DC, Sederoff RR, Chiang VL (2015). Phosphorylation is an on/off switch for 5-hydroxyconiferaldehyde O-methyltransferase activity in poplar monolignol biosynthesis. *Proceedings of the National Academy of Sciences* **112**: 8481-8486.

Wang L, Sun Y, Xia XL, Jiang TB (2018). Screening of proteins interacting with ERF transcriptional factor from *Populus simonii x P.nigra* by yeast two-hybrid method. *Biotechnology & Biotechnological Equipment* **32**: 543-549.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**: 296-303.

Wu XY, Li T (2017). A casein kinase II phosphorylation site in AtYY1 affects its activity, stability, and function in the ABA response. *Frontiers in Plant Science* **8**: 323.

Xie ZL, Nolan T, Jiang H, Tang BY, Zhang MC, Li ZH, Yin YH (2019). The AP2/ERF transcription factor TINY modulates brassinosteroid-regulated plant growth and drought

responses in *Arabidopsis*. *The Plant Cell* **31**: 1788-1806.

Xu ZS, Xia LQ, Chen M, Cheng XG, Zhang RY, Li LC, Zhao YX, Lu Y, Ni ZY, Liu L, Qiu ZG, Ma YZ (2007). Isolation and molecular characterization of the *Triticum aestivum L.* ethylene-responsive factor 1 (TaERF1) that increases multiple stress tolerance. *Plant Molecular Biology* **65**: 719-732.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015a). The I-TASSER Suite: protein structure and function prediction. *Nature Methods* **12**: 7-8.

Yang J, Zhang Y (2015b). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research* **43**: 174-181.

Yao W, Wang S, Zhou B, Jiang T (2016). Transgenic poplar overexpressing the endogenous transcription factor ERF76 gene improves salinity tolerance. *Tree Physiology* **36**: 896-908.

Zhu L, Liu D, Li Y, Li N (2013). Functional phosphoproteomic analysis reveals that a serine-62-phosphorylated isoform of ethylene response factor110 is involved in *Arabidopsis* bolting. *Plant Physiology* **1161**: 904-917.