**Molecular Plant**
Research article

CellPress
Partner Journal

# The origin and evolution of salicylic acid signaling and biosynthesis in plants

Xianqing Jia[1,3], Long Wang[1,3], Hongyu Zhao[1], Yibo Zhang[1], Zhixiang Chen[2], Lei Xu[1,*] and Keke Yi[1,*]

[1]Key Laboratory of Plant Nutrition and Fertilizer, Ministry of Agriculture and Rural Affairs, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[2]Department of Botany and Plant Pathology, Center for Plant Biology, Purdue University, West Lafayette, IN 47907-2054, USA

[3]These authors contributed equally to this article.

***Correspondence: Lei Xu (xulei@caas.cn), Keke Yi (yikeke@gmail.com)**

https://doi.org/10.1016/j.molp.2022.12.002

## ABSTRACT

**Salicylic acid (SA) plays a pivotal role in plant response to biotic and abiotic stress. Several core SA signaling regulators and key proteins in SA biosynthesis have been well characterized. However, much remains unknown about the origin, evolution, and early diversification of core elements in plant SA signaling and biosynthesis. In this study, we identified 10 core protein families in SA signaling and biosynthesis across green plant lineages. We found that the key SA signaling receptors, the nonexpresser of pathogenesis-related (NPR) proteins, originated in the most recent common ancestor (MRCA) of land plants and formed divergent groups in the ancestor of seed plants. However, key transcription factors for SA signaling, TGACG motif-binding proteins (TGAs), originated in the MRCA of streptophytes, arguing for the stepwise evolution of core SA signaling in plants. Different from the assembly of the core SA signaling pathway in the ancestor of seed plants, SA exists extensively in green plants, including chlorophytes and streptophyte algae. However, the full isochorismate synthase (ICS)-based SA synthesis pathway was first assembled in the MRCA of land plants. We further revealed that the ancient abnormal inflorescence meristem 1 (AIM1)-based β-oxidation pathway is crucial for the biosynthesis of SA in chlorophyte algae, and this biosynthesis pathway may have facilitated the adaptation of early-diverging green algae to the high-light-intensity environment on land. Taken together, our findings provide significant insights into the early evolution and diversification of plant SA signaling and biosynthesis pathways, highlighting a crucial role of SA in stress tolerance during plant terrestrialization.**

**Key words:** salicylic acid, SA signaling, SA biosynthesis, plant evolution, high irradiance

**Jia X., Wang L., Zhao H., Zhang Y., Chen Z., Xu L., and Yi K.** (2023). The origin and evolution of salicylic acid signaling and biosynthesis in plants. Mol. Plant. **16**, 245–259.

## INTRODUCTION

Plants have developed sophisticated defense mechanisms to overcome a wide variety of phytopathogens and abiotic stresses. One such counteractive mechanism is systemic acquired resistance mediated by salicylic acid (SA), which is linked to the expression of pathogenesis-related (PR) genes (Zhou and Zhang, 2020). During pathogen infection or abiotic stress, SA is induced and accumulates in both infected and systemic tissue, and a large number of genes, including those encoding PR proteins and enzymes involved in oxidative stress protection, are upregulated concomitantly (Kachroo et al., 2020). In addition to its well-characterized roles in immune response, SA has been reported to be involved in diverse abiotic stresses, such as salt, drought, chilling, heat, UV-B light, and high light intensity (reviewed in Khan et al., 2015). Beyond its function in defense signaling against biotic or abiotic stress, SA is also tightly associated with various aspects of plant growth and development, including photosynthesis, transpiration, ion uptake, thermogenesis, senescence, and cross talk with other hormones (Rivas-San Vicente and Plasencia, 2011).

Nonexpresser of PR (NPR) proteins have been identified as core SA receptors, and all NPRs bind TGACG motif-binding (TGA) transcription factors (Kesarwani et al., 2007). In turn, NPR–TGA protein complexes bind to the promoters of SA-responsive

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

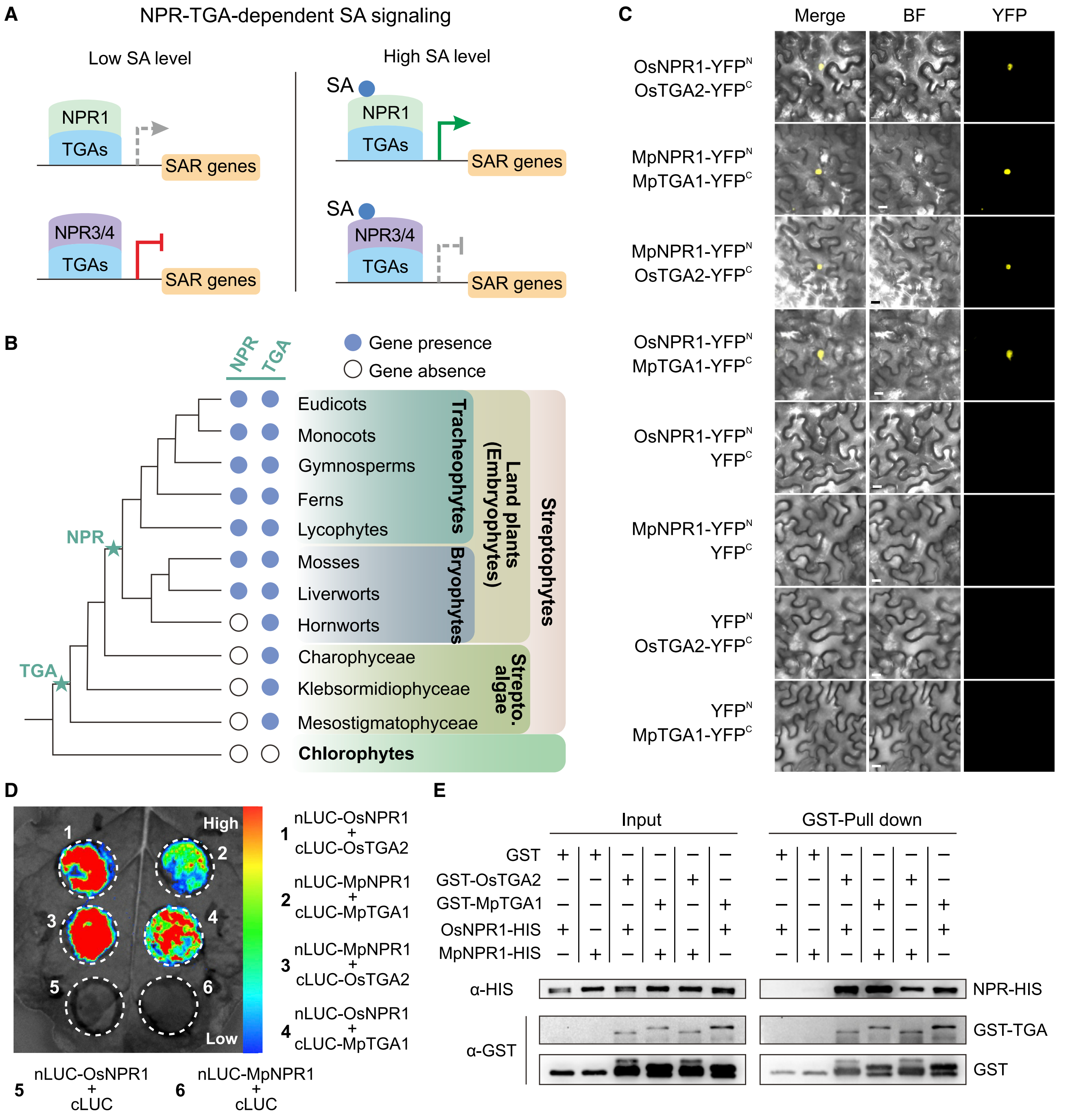

**Figure 1. Presence of core elements involved in SA signaling in green plants.**

**(A)** NPR-dependent SA signaling pathways in the model plant *Arabidopsis*, modified from Peng et al. (2021). NPRs associate with TGA transcription factors, and NPR–TGA complexes bind to the promoters of SA-responsive (SAR) genes to regulate their expression. NPR1 and NPR3/4 have opposite roles in SA signaling: NPR1 protein acts as an activator, and NPR3/4 proteins act as repressors. At a low SA level, SAR genes are poorly activated by the NPR1–TGA complex, while they are highly suppressed by the NPR3/4–TGA complex (red line); thus their expression is subdued. Under a high SA content, SAR genes are activated by the NPR1–TGA complex (green arrow), and the repression by the NPR3/4–TGA complex is relieved by SA binding, thus inducing the expression of SAR genes.

**(B)** Presence of NPR and TGA proteins in plants. Stars mark the proposed origins of the NPR and TGA families. Phylogeny was adopted from previous studies (Rensing, 2020; Su et al., 2021).

**(C)** BiFC analysis of the interaction between NPRs and TGAs of rice and *M. polymorpha*. The N-terminal fragment of YFP (termed YFP$^N$) was fused to the C terminus of OsNPR1 and MpNPR1. The C-terminal fragment of YFP (termed YFP$^C$) was fused to OsTGA2 and MpTGA1. Scale bars, 20 μm.

*(legend continued on next page)*

(SAR) genes (Wang et al., 2011) to regulate their expression. The NPR1 and NPR3/4 proteins have opposing roles in SA signaling: NPR1 proteins activate, and NPR3/4 proteins repress, the SA-activated pathway (Ding et al., 2018). When SA levels are low, SAR genes are weakly activated by the NPR1–TGA complex but strongly suppressed by the NPR3/4–TGA complex. At high SA levels, SAR genes are activated by the NPR1–TGA complex. Meanwhile, SA binding relieves NPR3/4–TGA-mediated repression, further enhancing the increased expression of SAR genes. It has been suggested that NPR-dependent SA signaling originated in the most recent common ancestor (MRCA) of land plants, and only one NPR gene was identified in *Marchantia polymorpha* (Wang et al., 2015; Baggs et al., 2020). However, less is known about how SA signaling components evolved during early land plant evolution, how NPR diverged, and whether the NPR–TGA module existed and functioned in early land plants. With recently sequenced genomes, especially for bryophytes and streptophyte algae (Nishiyama et al., 2018; Cheng et al., 2019; Li et al., 2020; Liang et al., 2020; Zhang et al., 2020), the previously lacking data can now be obtained using new genomic sources to understand the origin and evolution of the core elements of SA signaling and biosynthesis.

Biosynthesis of SA in plants starts in the chloroplast with chorismate (CA) via two major metabolic routes, the isochorismate (IC) synthase (ICS) pathway and the phenylalanine ammonia-lyase (PAL) pathway (reviewed in Huang et al., 2020; Peng et al., 2021). IC is then transported by the multidrug and toxin extrusion transporter, enhanced disease susceptibility 5 (EDS5), to the cytoplasm (Nawrath et al., 2002). This step is followed by conjugation of IC to glutamate (Glu) by *avrPphB* susceptible 3 (PBS3) to produce IC-9-Glu, and IC-9-Glu is further broken down into SA by enhanced *Pseudomonas* susceptibility 1 (EPS1) (Torrens-Spence et al., 2019). In the PAL pathway, CA is converted into phenylalanine (Phe), and PALs then convert Phe to *trans*-CA (Ro and Douglas, 2004), which is further converted to benzoic acid (BA) by abnormal inflorescence meristem 1 (AIM1)-dependent β-oxidation (Richmond and Bleecker, 1999; Qualley et al., 2012; Xu et al., 2017). SA is subsequently produced by hydroxylation of BA via presumptive BA-2-hydroxylases (León et al., 1995). However, recent studies showed that the biosynthesis of *trans*-CA from Phe could be mediated by an alternative pathway catalyzed by aromatic amino acid aminotransferase (AAAT) and phenylpyruvic acid reductase (PPAR) (Wang et al., 2019; Zeng et al., 2020). Thus, in this study, we redefined the PAL pathway as the β-oxidation-dependent pathway. Despite knowledge about SA biosynthesis, the origin and evolution of the ICS and β-oxidation-dependent pathways in plants remain unclear. The ICS pathway is designated as the major route of SA biosynthesis for immunity in Brassicaceae (Peng et al., 2021), but it is not essential for SA biosynthesis in some monocots, such as barley (Qin et al., 2019) and rice (Xu et al., 2017, 2022). These data suggest that the contribution of the ICS pathway to SA biosynthesis during the immune response may be restricted to eudicots or even specific orders within eudicots. Therefore, studying SA biosynthesis and its evolution in plant lineages is necessary to fully understand its function. Furthermore, SA is widely present in higher land plants, but the presence or absence of SA in unicellular green algae and streptophyte algae has not been established.

Here, we try to elucidate the evolutionary origin of core proteins in SA biosynthesis and signaling by performing phylogenetic analysis across the Archaeplastida. We also investigate the functionality of the NPR–TGA module by analyzing the conservation of interactions and subcellular localization. We find that the core NPR regulators of SA signaling originated in the MRCA of land plants. The conserved SA signaling module NPR–TGA functions in the nuclei of *Marchantia* and has diverged during long-term plant evolution. Unlike the core SA signaling components, which were fully assembled in the ancestor of seed plants, SA is found extensively across green plants, including chlorophytes and streptophyte algae. The canonical ICS-based SA biosynthesis pathway was first assembled in the ancestors of land plants. We further reveal that the ancient AIM1-based β-oxidation pathway is crucial for the biosynthesis of SA in chlorophyte algae and that the extensive presence of this biosynthesis pathway in green plant lineages was crucial for plant adaptation to high-irradiance environments after plant terrestrialization. Taken together, our findings provide significant insights into the early evolution and diversification of plant SA signaling and biosynthesis pathways.

## RESULTS AND DISCUSSION

### NPR–TGA is an ancient and conserved module for SA signaling

The NPR–TGA module is widely reported as the core machine for SA signaling (Ding et al., 2018) (Figure 1A). Previous studies have shown that NPR originated in the MRCA of land plants (Wang et al., 2015; Baggs et al., 2020), but these studies are limited by the small number of plant species examined and particularly by the lack of data from hornwort and early-diverging streptophyte algae. To investigate the presence of core elements involved in SA signaling in plants, we used a similarity search combined with a phylogenetic approach to systematically identify these core elements in 47 representative plants (Archaeplastida) with well-assembled genomes, namely 4 rhodophytes, 10 chlorophytes, 3 charophytes, 5 bryophytes (i.e., 2 hornworts, 1 liverwort, and 2 mosses), 2 ferns, 1 lycophyte, 6 gymnosperms, and 16 angiosperms (Supplemental Figure 1 and Supplemental Table 1). To ensure the integrity of the protein sequences derived from the genome assemblies, proteins from each core family harboring the respective conserved domains (listed in Supplemental Table 2) were selected for further analysis. Our

**(D)** Luciferase complementation imaging (LCI) assay of the interaction between NPRs and TGAs, performed in tobacco leaves. The vectors were co-infiltrated into tobacco leaves, and the LUC activities were analyzed 48 h after infiltration. Other negative controls are shown in Supplemental Figure 2. Similar results were obtained in at least three repeats.

**(E)** Pull-down analysis of the interaction between NPRs and TGAs of rice and *M. polymorpha*. Purified GST, GST–OsTGA2, GST–MpTGA1, OsNPR1–HIS, and MpNPR1–HIS proteins were used to perform the pull-down assay. Immunoblots were developed with anti-HIS antibodies to detect OsNPR1–HIS and MpNPR1–HIS and anti-GST to detect GST, GST–OsTGA2, and GST–MpTGA1.

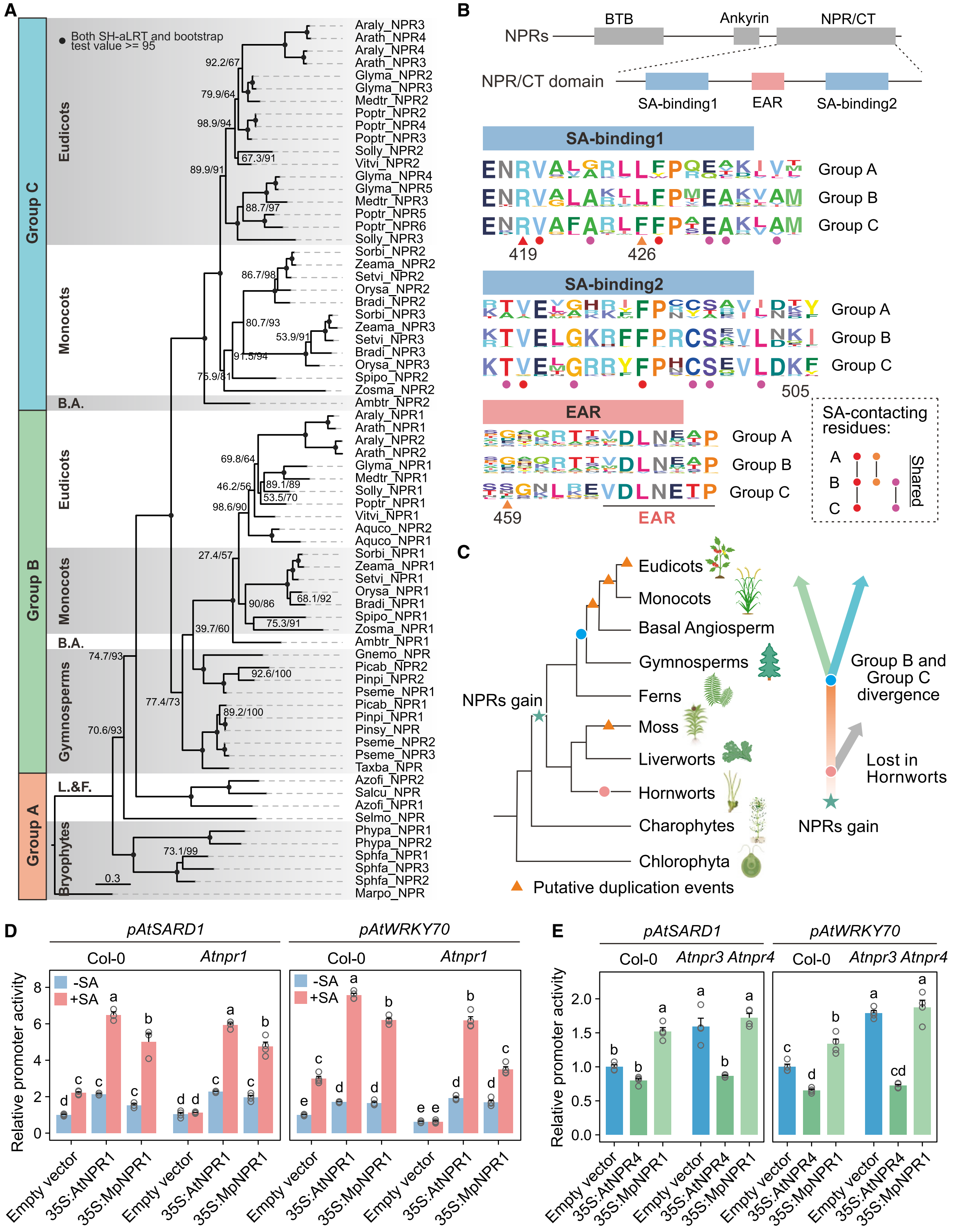

**Figure 2. Phylogeny of NPR proteins and proposed scenario of NPR evolution in land plants.**

**(A)** Phylogeny of NPR proteins was reconstructed from their full-length protein sequences. Multiple sequence alignment was performed by Clustal Omega, and the tree was built by IQ-TREE2 using the maximum-likelihood method with JTT + I + G4 as the best-fitting model and 1000 replicates. Nodes

*(legend continued on next page)*

results showed that sequences encoding NPRs and TGAs emerged in the MRCA of land plants and Streptophyta, respectively (Figure 1B). Interestingly, the copy number of NPR and TGA genes increased drastically in land plants after the split between the charophyte and land plant lineages (Supplemental Figure 1). Given the crucial roles of SA in plant defense and growth, the increased complexity of SA signaling pathways was relevant to adapting to the intricate terrestrial environments after the colonization of land.

To examine whether this crucial NPR–TGA interaction module exists in the early living relatives of land plants, we next tested for interaction between MpNPR1 and MpTGA1 (there is only one NPR and one TGA in *M. polymorpha*) using bimolecular fluorescence complementation (BiFC), luciferase (LUC) complementation imaging (LCI), and pull-down assays (Figure 1C–1E). The previously reported interaction between OsNPR1 and OsTGA2 in rice was used as a positive control (Moon et al., 2018). The co-expression of MpNPR1–YFP$^N$/MpTGA1–YFP$^C$ led to a fluorescence signal in the nucleus of tobacco leaf cells, as was observed for OsNPR1–YFP$^N$/OsTGA2–YFP$^C$. In contrast, the co-expression of OsNPR1–YFP$^N$/YFP$^C$ or MpNPR1–YFP$^N$/YFP$^C$ did not produce any detectable fluorescence signal (Figure 1C). The results from the LCI and pull-down assays verified the interactions between MpNPR1 and MpTGA1, with OsNPR1 and OsTGA2 as positive controls (Figure 1D and 1E and Supplemental Figure 2). Furthermore, crossed-interaction assays showed that MpNPR1 interacted with OsTGA2, and OsNPR1 interacted with MpTGA1 (Figure 1C–1E). Taken together, these results indicate that NPR–TGA is an ancient and conserved SA signaling module present in the ancestor of land plants.

## Evolution and divergence of NPRs and TGAs

It has been shown that NPR1 and NPR3/NPR4 have opposite roles in the transcriptional regulation of the plant immunity (Ding et al., 2018). We, therefore, investigated when this functional divergence arose. To this end, we performed a phylogenetic analysis with the maximum-likelihood method based on all 70 identified NPR sequences across liverworts, mosses, ferns, gymnosperms, and angiosperms. The topology of the tree showed that the NPR family comprises three distinguishable groups, which we termed Group A, Group B (containing the AtNPR1 protein), and Group C (containing the AtNPR3/4 proteins) (Figure 2A). Group A includes NPR proteins from liverworts, mosses, lycophytes, and ferns. The NPR protein family expanded in both mosses and ferns, resulting in two monophyletic groups in mosses and ferns, respectively. Group C NPR proteins are present only in flowering plants, whereas members of Group B are widely found in seed plant lineages. All monocots retained one copy of a Group B gene, but two copies were found in several eudicots, indicating that a gene duplication event occurred during eudicot evolution. In Group C NPRs, two apparent monophyletic ancestors were formed in grasses, suggesting that a recent duplication event occurred before the generation of grass and that the resulting NPR genes were retained after duplication. Furthermore, in the eudicots, several monophyletic groups of Group C proteins formed, indicating that this group underwent both loss and duplication during long-term diversification. These results suggested that the divergence of NPR proteins of Group B and Group C occurred before or along with the diversification of seed plants.

NPR proteins typically contain three conserved domains, a BTB/POZ domain, ankyrin repeats, and an NPR1/NIM1-like defense protein C-terminal domain (hereafter called the NPR domain, also known as the CT domain) (Figure 2B). The NPR domain has been shown to bind SA (Cao et al., 1997; Wang et al., 2020), and it contains three main features, two SA-binding regions (designated SA-binding1 and SA-binding2 here) and a putative ethylene-responsive element-binding-factor-associated amphipathic repression (EAR) motif. After multiple sequence alignments using these three key regions of the NPR domain (Supplemental Figure 3), we found that, of the vital SA-contacting amino acid sites, about 30% (4 of 13) of residues are conserved in the three NPR groups. The other 70% (9 of 13) of residues are highly conserved in NPRs of Group B and Group C (Figure 2B and Supplemental Figure 3). Intriguingly, this divergence in the SA binding sites is mainly linked to sequences from mosses and ferns, supporting possible differences in SA binding and signaling between mosses/ferns and seed plants (Peng et al., 2017). Conserved SA binding residues of AtNPR1 in *Arabidopsis* have been identified by recent structural assessment, and the three residues AA419, AA426, and AA459 (corresponding to the positions in AtNPR1) had the most substantial effect on SA binding (Wang et al., 2020). Here, we found that AA419 is conserved in all three NPR clades, while AA426 and AA459 are conserved between Group A and Group B proteins but are distinct in Group C proteins (Figure 2B). We examined another essential regulatory element, the EAR motif, and found that it is highly conserved in Group C but varies in Group A and Group B, indicating that the functional divergence between the NPR proteins occurred after the divergence of seed plants.

with less and greater than 95% SH-aLRT and bootstrap support are shown with numbers and points, respectively. B.A., basal angiosperms; L.&F., lycophytes and ferns.

**(B)** Conserved amino acid sites of three core parts of the NPR domain required for SA binding. AA 419, 426, and 459, the residues with the most substantial effect on SA binding according to the structural analysis of AtNPR1 by Wang and colleagues (Wang et al., 2020), are marked by triangles. Residues that contact SA are marked by dots. The meaning of colors of triangles or dots are shown in the dotted box. EAR, ethylene-responsive element-binding-factor-associated amphipathic repression.

**(C)** A proposed scenario of NPR evolution. Canonical NPR proteins originated in the last common ancestor of land plants and were likely lost in hornworts, then NPRs of Group B and Group C emerged and diverged before or along with the formation of seed plants. Putative duplication events are marked by orange triangles. Phylogeny was adopted from previous studies (Rensing, 2020; Su et al., 2021).

**(D and E)** Luciferase activities in protoplasts of *Arabidopsis* Col-0 and *Atnpr1* **(D)** and *Atnpr3 Atnpr4* **(E)** co-transformed with the indicated constructs together with the *pAtSARD1:LUC* or *pAtWRKY70:LUC* reporter gene. The promoter–luciferase assay was performed as described previously (Ding et al., 2018). +SA and −SA indicate treatment with and without SA, respectively. Error bars indicate SE. Differences were tested by ANOVA using the LSD method with a Bonferroni correction at $\alpha = 0.05$. Means with the same letter in each group are not significantly different.

To verify the evolutionary and functional divergence of NPRs, we performed a promoter–LUC assay to examine the transcriptional regulation activities of representative NPRs from the three NPR groups. Previous studies showed that the AtNPR1 and AtNPR3/4 proteins'respectively activate and repress *AtSARD1* and *AtWRKY70* expression in *Arabidopsis* (Ding et al., 2018; Liu et al., 2020). We, therefore, tested the transcriptional regulation activities of a representative NPR in Group A, MpNPR1 (the only NPR protein in *M. polymorpha*). In brief, plasmids expressing NPR genes were co-transfected with constructs expressing a LUC reporter gene under the control of the promoter of *AtSARD1* or *AtWRKY70* into Col-0 protoplasts. Consistent with the findings of a previous study (Ding et al., 2018), when the *pAtSARD1:LUC* and *pAtWRKY70:LUC* reporters were co-transfected with plasmids overexpressing *AtNPR1* LUC expression was significantly increased, and when the reporters were co-transfected with plasmids overexpressing *AtNPR4, LUC expression was significantly* repressed (Figure 2D and 2E). Similar to *AtNPR1*, *MpNPR1* significantly activated, rather than repressed, the expression of *pAtSARD1:LUC* and *pAtWRKY70:LUC*. This indicates that *MpNPR1* might be a functional homolog of *AtNPR1*. To further test this, we determined the activity of MpNPR1 in the *Atnpr1* and *Atnpr3 Atnpr4* mutants. As expected, the SA application could induce the *AtSARD1* and *AtWRKY70* promoter activities in the Col-0 background but not in the *Atnpr1* mutant background (Figure 2D). Like *AtNPR1*, *MpNPR1* could restore the SA-induced *SARD1* and *WRKY70* promoter activities when expressed in the *Atnpr1* mutant background (Figure 2D). On the other hand, the *SARD1* and *WRKY70* promoter activities were enhanced in the *Atnpr3 Atnpr4* double-mutant background. However, unlike *AtNPR4*, overexpression of *MpNPR1* failed to rescue the defects of the *Atnpr3 Atnpr4* double mutant (Figure 2E). This clearly demonstrated that *MpNPR1* is functionally homologous to *AtNPR1* but not to *AtNPR4*. Likewise, the *Physcomitrella patens* NPR gene, another Group A member, partially rescued the mutant phenotypes of the *Atnpr1* mutant but did not complement the constitutive defense phenotype of the *Atnpr3 Atnpr4* double mutant (Peng et al., 2017). Taken together, these results suggested that Group B NPRs are functionally closely related to Group A NPRs.

One major difference between AtNPR1 and AtNPR3/4 is their subcellular localization. AtNPR1 mainly localizes to the cytoplasm and translocates to the nucleus after SA treatment, and AtNPR3/4 is mostly localized in the nucleus even without SA treatment (Ding et al., 2018; Liu et al., 2020). We, therefore, tested the impact of SA on the subcellular localization of two NPRs in Group A (MpNPR1 and ScNPR1 from *Salvinia cucullata*) and two rice NPR proteins (Group B OsNPR1 and Group C OsNPR3). This analysis also tested whether monocot NPRs work in a manner similar to *Arabidopsis* NPRs. As found for *Arabidopsis* NPR1 (Ding et al., 2018; Liu et al., 2020), OsNPR1 was mainly localized in the cytoplasm (77% of the total) in the absence of SA but was increasingly localized in the nucleus (65% of the total) after exogenous SA treatment. OsNPR3 was principally localized in the nucleus (90% of the total) even without SA treatment (Supplemental Figure 4A). However, MpNPR1 and ScNPR1 were always nuclear localized, with or without SA (Supplemental Figure 4B and 4C). Taken together, these data indicate that the nucleocytoplasmic distribution of NPRs of Group B and Group C diverged during seed plant evolution and that this divergence between NPRs of Group B and Group C is conserved in monocots and eudicots. These results also suggest that the evolution and maintenance of differences in SA receptors are most likely associated with the intricate control of SA responses.

TGA proteins, a class of basic-region leucine-zipper transcription factors, are required to establish the SA-dependent defense response through interactions with NPRs in plants (Kesarwani et al., 2007; Gatz, 2012). To further trace the evolutionary trajectory of TGA proteins, we reconstructed the TGA-homolog phylogeny with 259 TGA-homolog protein sequences identified in 32 plant species (Supplemental Figure 5). The phylogenic analysis showed that TGA homologs in angiosperms underwent several duplications and were divided into four main clades (termed Groups 1–4) after the divergence in land plants. In the *Arabidopsis* genome, 10 members of the TGA family have been identified, 6 of which (AtTGA1, AtTGA3, and AtTGA4 in Clade 3; AtTGA2, AtTGA5, and AtTGA6 in Clade 4) have been implicated in NPR-dependent defense signaling (reviewed in Garner et al., 2016). Expansion and functional differentiation of the TGA proteins in plants may have been necessary for adaptation to complex terrestrial environments, allowing plants to respond to diverse biotic and abiotic stresses.

Taking these findings together, we propose a possible NPR evolutionary route (Figure 2C): canonical NPR proteins originated in the MRCA of land plants (Embryophyta) but were likely lost in hornworts. The NPR proteins in Group B and Group C then emerged and diverged before or along with the formation of seed plants. The evolutionary conservation and divergence of NPR proteins indicate the vital roles of SA signaling in adapting to complicated terrestrial environments as land plants evolved. TGA, a known partner of NPR, originated in the MRCA of Streptophyta plants and subsequently diverged into four main clades in land plants. Interestingly, even though the NPR–TGA module acts as the core regulator of SA signaling, the origins of NPR and TGA proteins occurred at different stages of plant evolution. Exogenous SA significantly induced the expression of TGA genes in the Streptophyta algae (Supplemental Figure 6). Our data indicate that SA might have been present in Streptophyta algae before the origin of NPR proteins, raising the possibility that there are SA receptors other than NPRs that mediate the induction of TGA genes, which regulate SAR genes, in Streptophyta algae and perhaps in other plants as well.

## SA is present in the early-diverging plant species

The above results showed that the two components of the core SA signaling machinery, NPRs and TGAs, emerged in the MRCA of land plants and Streptophyta plants, respectively, and that the expression of TGAs is responsive to exogenous SA treatment in the Streptophyta algae. Thus, SA was likely already present in the Streptophyta algae, whereas it remains unclear whether it was also present in green algae. To clarify the presence of SA across plant lineages, here, we measured SA concentrations in five Chlorophyta algae (*Chlamydomonas reinhardtii*, *Pediastrum duplex*, *Scenedesmus bijuga*, *Scenedesmus quadricauda*, and *Scenedesmus spinosus*), three Streptophyta algae (*Mesostigma viride*, *Spirogyra gracilis*, and *Klebsormidium* sp.), and a liverwort (*M. polymorpha*), which were cultured in sterile medias (Figure 3 and Supplemental Table 3). SA was detected

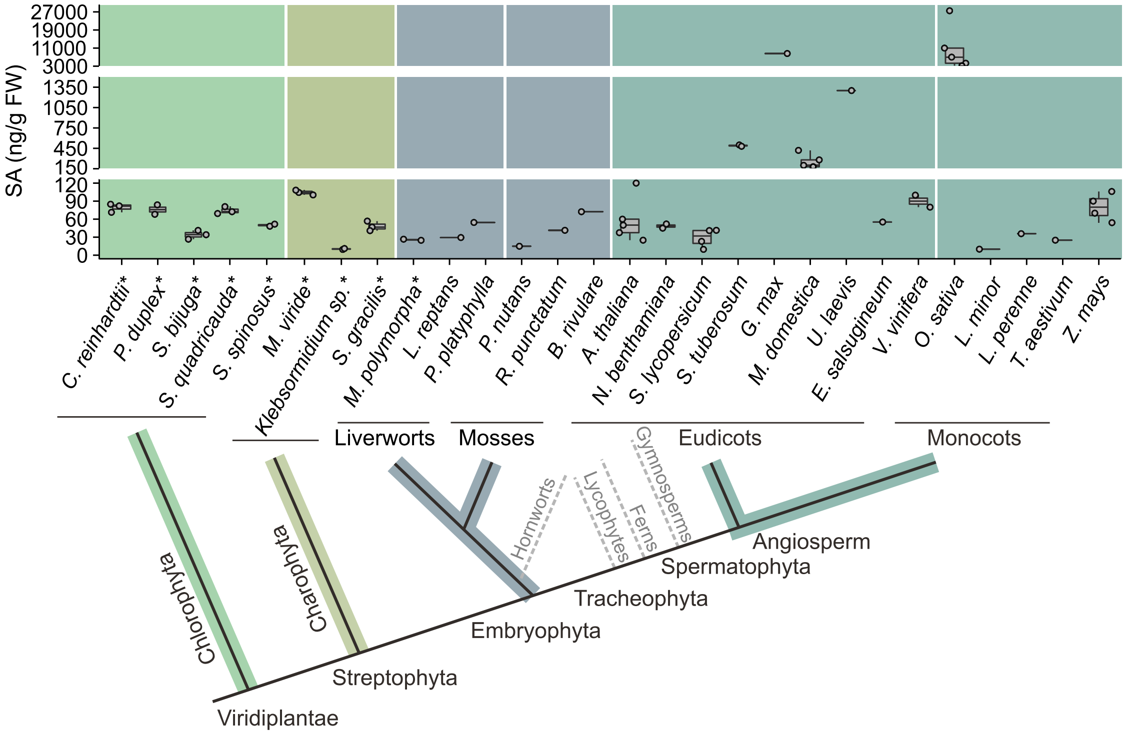

**Figure 3. SA concentrations of representative species in plant lineages.**
Data with asterisks were obtained in this study using liquid chromatography–tandem mass spectrometry. FW, fresh weight. Dashed lines mean that no data on SA concentration were collected in this lineage. The full list of collected SA concentration data is shown in Supplemental Table 3. Phylogeny was adopted from previous studies (Rensing, 2020; Su et al., 2021).

in all these clades, indicating that the SA biosynthesis pathway was already established in the early-diverging plant lineages, including green algae and bryophytes. These results combined with the SA concentrations in different species based on published data (Supplemental Table 3) indicate that SA exists extensively in green plants and that its concentrations vary greatly in different plants. For example, we found that the SA concentration in *C. reinhardtii* is comparable to that in *Arabidopsis thaliana*, around 80 ng/g (fresh weight). However, rice has hundreds of times more SA than *C. reinhardtii* (Supplemental Table 3). Taken together, these results indicate that SA is pervasive in green plants.

## ICS-based biosynthesis was present in the MRCA of land plants but not involved in SA synthesis in chlorophytes

Biosynthesis of SA in plants starts with CA in chloroplasts, with two major metabolic routes, the ICS pathway and the β-oxidation-dependent pathway (Figure 4A). For the ICS pathway, the conversion of CA to IC in plants, which occurs in the plastids, is catalyzed by ICSs with high similarity to bacterial ICS (Wildermuth et al., 2001; Garcion et al., 2008). This argues that plant ICS genes originated early in evolution and preceded the origin of plants. Based on the investigation of ICS coding sequences across Archaeplastida genomes, we found that ICS genes are present in all plant lineages (Figure 4B and Supplemental Figure 1). To further infer ICS evolutionary routes, phylogenetic analysis of a total of 42 ICS protein sequences was carried out. We found that the ICS phylogeny is clearly aligned with the species phylogeny (Supplemental Figure 7). Angiospermous ICS proteins are divided into two clades, one comprising ICS genes from monocots and the other ICS genes from eudicots, with ICS genes in *Amborella* as the base branch. Interestingly, while the ICS gene family originated very early in evolution, it did not undergo expansion via whole-genome duplication or tandem duplication. Monocots contain single copies of ICS genes, while the ICS genes in the eudicot *Arabidopsis* and Leguminosae lineages underwent a one-time duplication (Supplemental Figure 7).

To trace the evolution of other key elements in the ICS pathway, we reconstructed the phylogenetic trees of the EDS5, PBS, and EPS families based on identified genes (Supplemental Figures 8, 9, and 10). We found that EDS5 originated during the formation of streptophyte algae, PBS originated in diverging streptophyte algae, and EPS first occurred in the MRCA of land plants. Like ICS, EDS5 did not undergo multiple duplications, and in monocots only one copy was retained. In contrast, the PBS and EPS genes underwent many duplication events and are found in tandem duplication clusters (Supplemental Figure 11). It is interesting to explore why only single copies of ICS and EDS5 are retained in land plants, especially in monocots. Taken together, the canonical ICS pathway was present in the MRCA of land plants, and the ICS genes are found in all green plants.

ICS is also known to be involved in pathways other than the biosynthesis of SA precursors: IC produced by ICS is also a precursor for phylloquinone, which serves as an electron acceptor in the

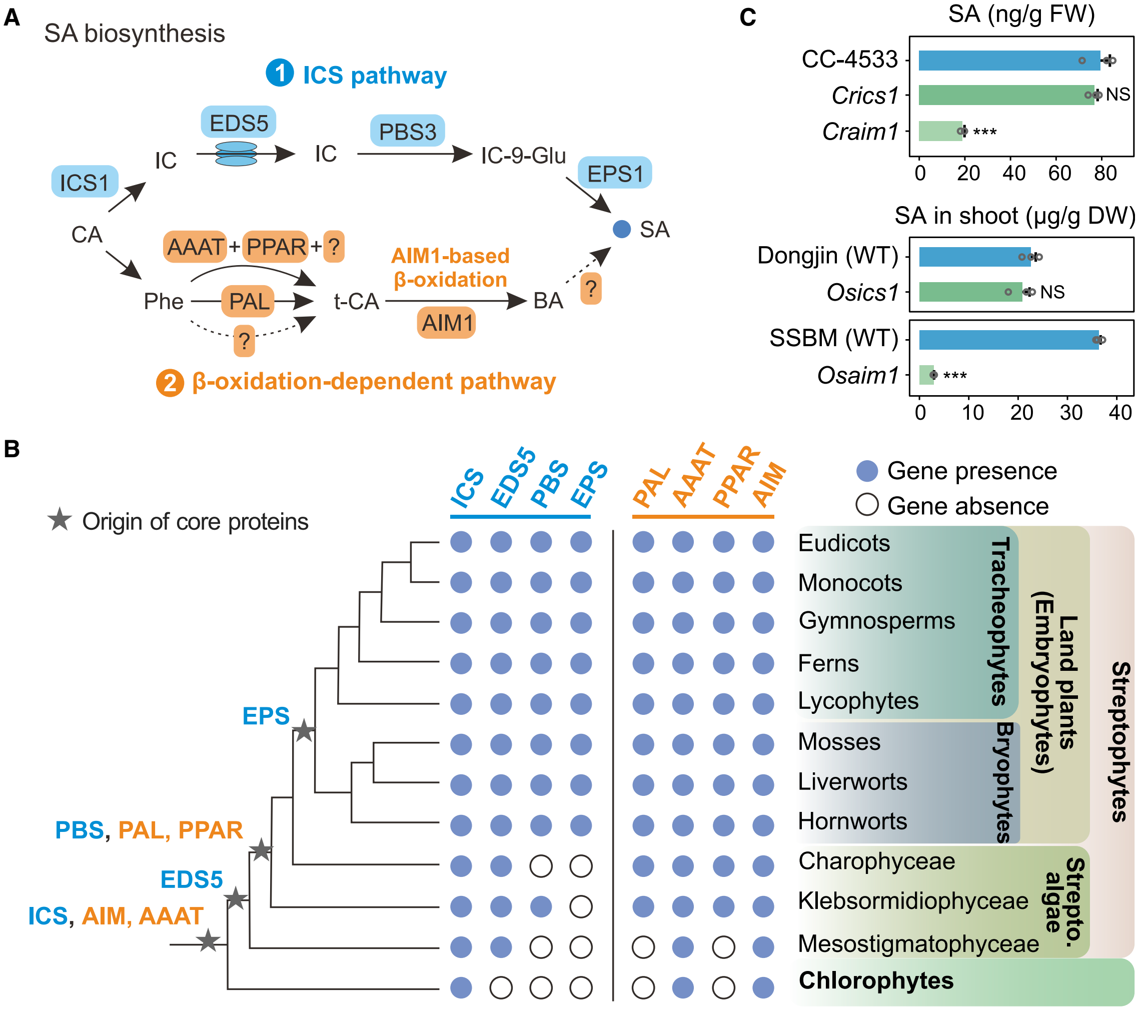

**Figure 4. Presence of core elements involved in SA biosynthesis in Plantae.**

**(A)** Two major SA biosynthetic routes in seed plants. In the ICS pathway, ICS proteins convert chorismate (CA) to isochorismate (IC) in the plastids, then IC is transported by the EDS5 transporter to the cytoplasm, and converted to isochorismate-9-glutamate (IC-9-Glu) by PBS3. Finally, IC-9-Glu spontaneously breaks down into SA or be converted to SA by EPS1. In the β-oxidation-dependent pathway, CA is converted into Phe in the chloroplasts, and Phe is converted to *trans*-CA (t-CA) by PAL or an alternative pathway based on AAAT and PPAR. t-CA is further converted to benzoic acid (BA) via AIM1-based β-oxidation. The key proteins that transform BA to SA have not been identified yet.

**(B)** Presence of core sequences involved in SA biosynthesis and signaling in plants. Phylogeny was adopted from previous studies (Rensing, 2020; Su et al., 2021).

**(C)** The SA concentration was significantly reduced in the *Craim1* mutant but not in the *Crics1* mutant compared with wild-type (WT) *C. reinhardtii*, and similar results were obtained in rice. Rice SA concentration data were collected from Xu et al. (2022). *Osaim1* and *Osics1* mutants were generated in the SSBM and Dongjin genetic backgrounds, respectively. Error bars indicate SE. Differences between each mutant and its WT were tested by *t*-test. ***$P <$ 0.001; NS, not significant.

chloroplast electron transport chain (Gross et al., 2006). Notably, the ICS domain in the earliest diverging plants (including chlorophytes, streptophyte algae, bryophytes, and ferns) is fused with MenC, MenD, and α/β-hydrolase domains (or some of them) (Supplemental Figure 7), which are involved in phylloquinone biosynthesis (Gross et al., 2006), indicating that in early-diverging plant lineages, such as in unicellular green algae, ICS is responsible for important metabolic pathways other than SA synthesis. To test this, we identified a knockdown strain for only one ICS (*CrICS1*, Cre16.g659050) in the model unicellular green algae *C. reinhardtii* (Supplemental Figure 12). SA content determination revealed no significant difference in the *CrICS1* knockdown strain (~76.8 ng/g fresh weight [FW]) compared with the wild type (CC-4533, ~79.5 ng/g FW) (Figure 4C). Similar results were obtained for the rice *Osics1* mutant (Xu et al., 2022). Thus, ICS may not be involved in SA biosynthesis in chlorophytes, and ICS-dependent SA biosynthesis was possibly established along with the stepwise gains of downstream biosynthetic genes during plant evolution, specifically in some eudicots.

## The AIM1-based β-oxidation-dependent pathway is an ancient SA biosynthesis pathway

Given that the canonical elements of the ICS-dependent SA biosynthesis pathway were initially assembled in the last ancestor of land plants, and ICS is not involved in SA biosynthesis in chlorophytes, we inferred that other SA biosynthesis pathways are responsible for SA synthesis in early-diverging green algae. We, therefore, focused on the β-oxidation-dependent pathway, which is also widely found in land plants (Barros and Dixon, 2020). AIM1 was demonstrated to be the key enzyme in β-oxidation; it was shown to convert *trans*-CA to BA (the precursor of SA) (Xu et al., 2017, 2022) (Figure 4A), and the loss of function of AIM1 in rice dramatically reduced SA levels (Xu et al., 2017). Here, we first traced the evolutionary trajectory of AIM1 and the divergence of AIM1 homologs by reconstructing the AIM1-homolog phylogeny with 99 AIM1 homologs identified from 44 plant species (Supplemental Figure 13). The results of phylogenetic analysis showed that AIM1 homologs emerged in the MRCA of green plants and that this family subsequently expanded and diverged in the land plants. To clarify the roles of AIM1 homologs in SA biosynthesis in chlorophytes, we identified a *C. reinhardtii* strain containing a mutation in AIM1 (Cre16.g695050) (Supplemental Figure 12). Similar to the rice *Osaim1* mutant (Xu et al., 2022), the *Craim1* mutant had SA levels (∼19 ng/g FW) that were greatly lower than those in the wild type (∼79.5 ng/g FW) (*t*-test, $P$ = 0.001472). Taken together, the β-oxidation-dependent SA biosynthetic pathway probably originated in the MRCA of green plants, and AIM1-based β-oxidation is required for SA biosynthesis in chlorophytes.

There are two reported pathways for the biosynthesis of *trans*-CA, the substrate of β-oxidation: the PAL-based pathway and the AAAT/PPAR-based pathway (Figure 4A). Our investigation showed that PAL is present in the late-diverging streptophyte algae but not in the rhodophytes, chlorophytes, or early-diverging streptophyte algae (e.g., *M. viride*) (Figure 4B and Supplemental Figure 1). PAL copy numbers drastically increased in liverworts, mosses, and seed plants. To further understand the duplication of PAL genes, we performed a phylogenetic analysis of 226 PAL protein sequences with the maximum-likelihood method (Supplemental Figure S14). Angiosperm PAL genes are divided into two main clades, one from monocots and the other from eudicots. PAL gene duplication events occurred several times after the divergence of monocots and eudicots. Interestingly, consistent with a previous study (Bagal et al., 2012), gymnosperms have retained a diverse set of PAL genes distributed into four major clades that arose from gene duplications predating the divergence of the two seed plant lineages. Genes encoding AAAT and PPAR, members of the other known pathway for biosynthesis of *trans*-CA, originated in the MRCA of green plants and late-diverging streptophyte algae, respectively (Figure 4B and Supplemental Figures 15 and 16).

Taken together, these results indicate that the β-oxidation-dependent pathway is an ancient SA biosynthesis route that originated in the MRCA of green plants and is responsible for the biosynthesis of SA in green algae. Further analysis will be necessary to identify additional components involved in SA biosynthesis and the biological roles of the moderate levels of SA in chlorophytes.

## SA is crucial for tolerance to increased light intensity after terrestrialization

In land plants, both the SA signaling and synthesis pathways have evolved into very complex networks, likely because of their vital roles in plant defense and stress responses. It is unclear, however, whether the ubiquitous presence of SA in green plants is related to terrestrial adaptation. Notably, light intensity dramatically increased during the terrestrialization of the ancestors of green plants, as their habitat changed from the deep sea to shallow waters and from the aquatic environment to land, which is directly exposed to sunlight (Maberly, 2014; Han et al., 2019; Zhang et al., 2022). To prevent photo-oxidative damage and to acclimate to excess light, land plants have evolved direct and indirect mechanisms for sensing and responding to excess light (Li et al., 2009; Schwenkert et al., 2022). A recent study showed that an expanded repertoire of genes for cell-wall biosynthesis in *Penium margaritaceum*, a member of the sister lineage to land plants, help limit water loss and protect against high radiation, contributing to the process of terrestrialization (Jiao et al., 2020). Previous studies also showed that endogenous SA protects rice plants from oxidative damage caused by high light stress (Yang et al., 2004) and that SA could protect the photosystem by alleviating photoinhibition in *A. thaliana* under high light (Chen et al., 2020).

To test whether the ubiquitous presence of SA throughout green plants is functionally associated with the adaptation to increased light after terrestrialization, we analyzed the role of SA in protecting photosystem II (PSII) against high irradiance in four representative species of the green plant linage, *C. reinhardtii*, *M. polymorpha*, *A. thaliana*, and *Oryza sativa*. First, we found that after exposure to high light, SA levels in all four plant species increased significantly (Figure 5A). No such increase was detected in the two SA-biosynthesis-defective mutants *Chlamydomonas Craim1* and rice *Osaim1* (Figure 5A). Second, the maximum quantum efficiency of PSII photochemistry (ratio of variable fluorescence/maximal fluorescence [Fv/Fm]) decreased in all four of these plant species after high light treatment, and this decrease could be partially rescued by SA treatment (Figure 5B–5F). More importantly, the *Craim1* and *Osaim1* mutants, but not their complementation lines, displayed a significantly lower Fv/Fm ratio than wild-type plants after high light treatment. Taken together, these results indicate that SA plays a vital role in the tolerance to high light intensity and that the AIM1-dependent β-oxidation pathway is crucial for tolerance to high irradiance in both chlorophytes and land plants. In addition, the presence of the AIM1-based β-oxidation pathway may have facilitated the adaptation of early-diverging green algae to the high-light-intensity environment on land.

In this study, we examined the evolutionary origin of core proteins involved in SA signaling and SA biosynthesis using phylogenetic analysis across green plants. We found that the core receptors of SA signaling, NPR proteins, originated in the MRCA of land plants, while their partners, TGAs, originated in the MRCA of streptophytes. This NPR–TGA core SA signaling module originated later than SA itself, and its two components originated at different evolutionary stages. The NPR–TGA interaction occurred when this module was first assembled in the land plants. The differentiation of NPR functions occurred during the divergence of spermatophytes, which helped build complex stress-response networks. However, SA exists extensively in green plants, including chlorophytes and streptophyte algae, although the ICS-based SA biosynthesis pathway was first present in the MRCA of land plants. We also found that the core process of the β-oxidation-dependent SA biosynthesis pathway, AIM1-based β-oxidation, originated in the

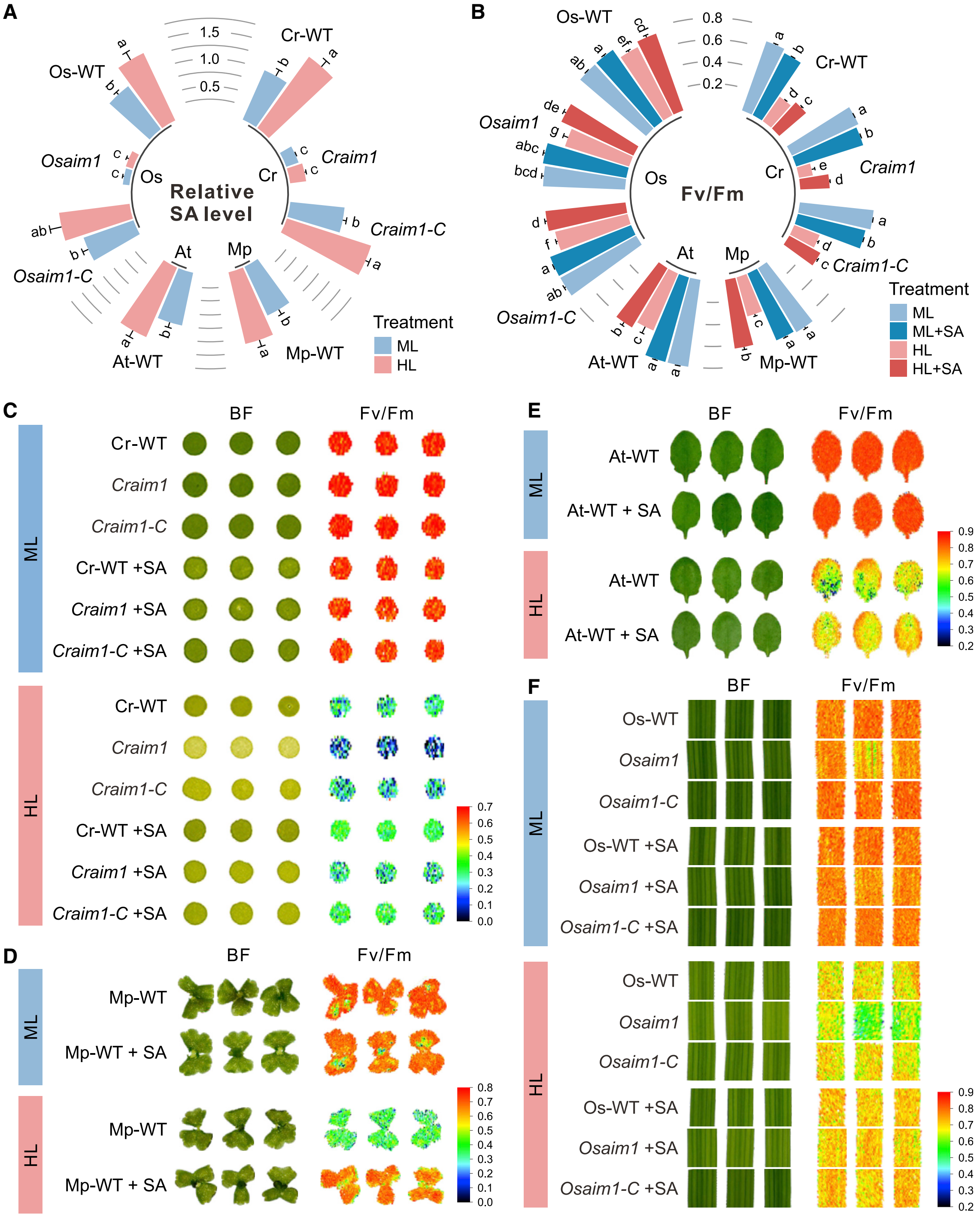

**Figure 5. SA can alleviate high light stress.**

**(A)** Relative SA levels measured for each treatment. Error bars indicate SD. Differences were tested by ANOVA using the LSD method with a Bonferroni correction at $\alpha = 0.05$. Means with the same letter in each group are not significantly different.

**(B)** Fv/Fm ratios measured for each treatment. Error bars indicate SD. Differences were tested by ANOVA using the LSD method with a Bonferroni correction at $\alpha = 0.05$. Means with the same letter in each group are not significantly different.

**(C–F)** Images of chlorophyll fluorescence (presented as Fv/Fm ratio) for each treatment. Cr-WT, Mp-WT, At-WT, and Os-WT indicate the wild-type plants of *C. reinhardtii* **(C)**, *M. polymorpha* **(D)**, *A. thaliana* **(E)**, and *O. sativa* **(F)**, respectively. *Craim1-C* and *Osaim1-C* are complementation lines of the *Craim1* and *Osaim1* mutants, respectively. +SA, treated with SA; BF, bright field; ML, moderate light; HL, high light. Experiments were repeated three times with similar results.

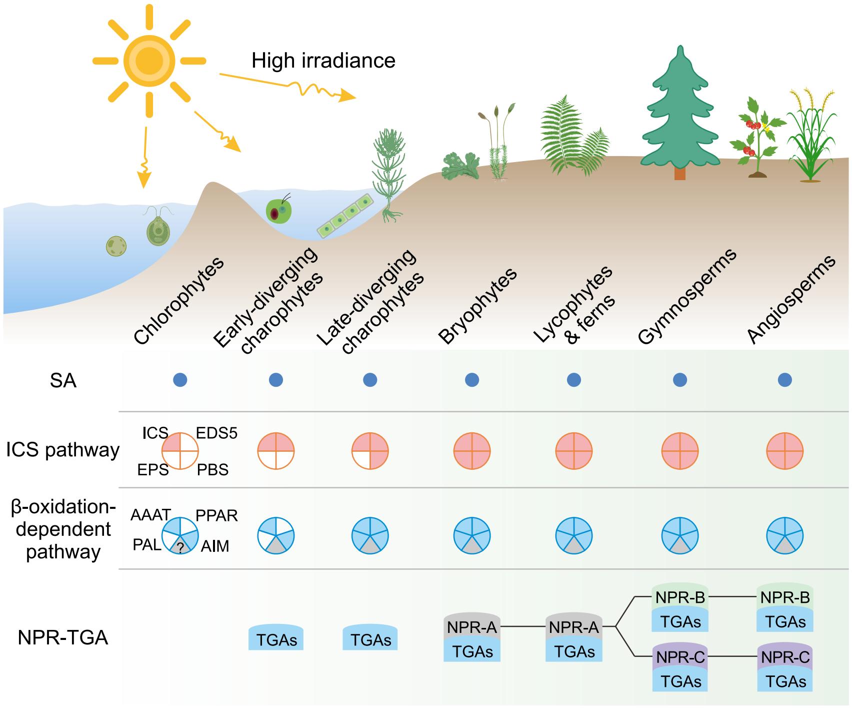

**Figure 6. Proposed trajectories of the origin and evolution of SA signaling and biosynthesis.**

The core receptors of SA signaling, NPR proteins, originated in the most recent common ancestor (MRCA) of land plants, while their partners, TGAs, originated in the MRCA of streptophytes. The NPR–TGA interaction occurred when this module was first assembled in the land plants. The differentiation of NPR functions occurred during the divergence of spermatophytes, which helped to build intricate stress-response networks. However, SA exists extensively in green plants, including chlorophytes and streptophyte algae, although the ICS-based SA biosynthesis pathway was first present in the MRCA of land plants. The core process of the β-oxidation-dependent SA biosynthesis pathway, AIM1-based β-oxidation, originated in the MRCA of green plants. The extensive presence of SA in green lineages might be important for their adaptation to high-irradiance environments after plant terrestrialization. Other possible unidentified elements are marked by question marks with a gray background, and the absence of elements is indicated by a blank background.

MRCA of green plants. Mutant strain analysis showed that AIM1-based β-oxidation is the key process contributing to SA biosynthesis in unicellular green algae. The known proteins performing synthesis of SA intermediates were not found to work together in chlorophytes, indicating that there are other unknown elements involved in the biosynthesis of SA and its intermediates in early-diverging green algae. Furthermore, our results suggest that extensive presence of SA in green lineages is crucial for their adaptation to high-irradiance environments after plant terrestrialization (Figure 6). Taken together, our findings provide significant insights into the early evolution and diversification of plant SA signaling and biosynthesis pathways.

## METHODS

### Materials and growth conditions

*C. reinhardtii* wild-type strain CC-4533 and the *Crics1* (LMJ.RY0402.243919) and *Craim1* (LMJ.RY0402.141054) mutants were obtained from the *Chlamydomonas* Resource Center (Li et al., 2019). Cells were cultured in standard Tris-acetate-phosphate (TAP) media (Harris, 1989) at pH 7.0 under continuous illumination on a rotating platform (150 rpm) at 24°C. Strains of *P. duplex* (FACHB-2909), *S. bijuga* (FACHB-76), *S. quadricauda* (FACHB-508), *S. spinosus* (FACHB-1268), *S. gracilis* (FACHB-354), and *Klebsormidium* sp. (FACHB-1489) were obtained from the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB, Chinese Academy of Sciences [CAS]) and cultured in the standard BG11 medium (Castenholz, 1988). *M. polymorpha* accession Takaragaike-1 (Tak-1, male) was provided by Dr. Lin Xu (Center for Excellence in Molecular Plant Sciences, CAS) and was grown primarily on half-strength Gamborg's B5 medium containing 1% agar. *A. thaliana* Col-0 plants were grown under long-day conditions (22°C, 16 h light/8 h dark) and approximately 80% humidity. Rice (*O. sativa*) materials (wild type and the *Osaim1*, *Osaim1-C*, and *Osics1* lines) were described in Xu et al. (2022). All rice plants were grown in a greenhouse with a 12-h day (30°C)/12-h night (22°C) photoperiod and approximately 60% humidity. Detailed values of the illumination intensity of all species are listed in Supplemental Table 4.

### Complementation and transformation in *Chlamydomonas*

The *CrAIM1* cDNA was cloned and ligated downstream of the Hsp70A–Rbcs2 promoter in the pHsp70a-Rbcs2-Ble vector. Nuclear transformations were performed by introducing 2 μg linearized plasmid into the *Craim1* mutant by electroporation (capacitance of 25 μF, voltage of 800 V, using a standard 0.4-cm cuvette; Bio-Rad). Transformants were selected on a solid TAP medium containing 20 μg ml$^{-1}$ paromomycin and 6 μg ml$^{-1}$ bleomycin (Shimogawara et al., 1998). Positive transformants were further validated by assessing the relative expression level of *CrAIM1* using quantitative real-time PCR.

### Identification of core elements of SA signaling and biosynthesis

Genome and gene annotation data of all 47 selected plant species were collected from the Phytozome database (Goodstein et al., 2012), NCBI Assembly, PLAZA (Van Bel et al., 2018), FernBase (Li et al., 2018), and published data (Li et al., 2020) (Supplemental Table 1). Identification of sequences encoding core SA-related elements was performed using pipelines described in our previous study (Wang et al., 2021). In brief, these core proteins were identified by InterProScan 5 (Jones et al., 2014) and validated by BLASTP with an e-value cutoff at 1e−10. The protein sequences withconserved domains (Supplemental Table 2) were selected and extracted using homemade scripts. Logos of conserved amino acids were created using kpLogo (Wu and Bartel, 2017).

### Construction of phylogenetic trees

To reconstruct the phylogenetic tree of all NPR proteins, full-length protein sequences were aligned using Clustal Omega (Sievers et al., 2011), and the maximum-likelihood gene phylogeny was reconstructed by IQ-TREE2 (Minh et al., 2020) with JTT + I + G4 as the best-fitting model. Both the SH-aLRT test and ultrafast bootstrap with 1000 replicates were conducted in IQ-TREE2 to obtain the supporting values for each internal node of the tree. Phylogenetic reconstruction for other core proteins was performed in a similar manner, and the phylogeny of plant species was adopted from previous studies (Rensing, 2020; Su et al., 2021). All sequence alignments and original phylogenetic trees were uploaded to https://doi.org/10.6084/m9.figshare.14140691.

### Genomic synteny analysis and visualization

To analyze the genomic synteny relationships among plants, protein sequences for each of the plant genomes were compared with one another using the Diamond program with an e-value cutoff at 1e−5. Only the top five non-self BLASTP hits were retained as input for MCScanX (Wang et al., 2012) analyses. The genomic synteny plots for the associated chromosomal regions were generated using the Python JCVI utilities (Tang et al., 2015) and homemade R scripts available by request.

### Determination of SA levels

Fresh plant samples were harvested, immediately frozen in liquid nitrogen, ground into powder (30 Hz, 1 min), and stored at −80°C until needed. Fifty milligrams of plant sample was weighed into a 2-ml plastic microtube and frozen in liquid nitrogen, dissolved in 1 ml methanol/water/formic acid (15:4:1, v/v/v). Ten microliters of internal standard mixed solution (100 ng/ml) was added into the extract as an internal standard for quantification. The mixture was vortexed for 10 min and then centrifuged for 5 min (12 000 rpm and 4°C). The supernatant was transferred to clean plastic microtubes, followed by evaporation to dryness, dissolved in 100 $\mu$l 80% methanol (v/v), and filtered through a 0.22-$\mu$m membrane filter. Total SA contents were measured by liquid chromatography–tandem mass spectrometry as previously reported (Floková et al., 2014). SA contents were calculated by dividing the whole amount by the FW (ng/g FW).

### High light treatment and determination of chlorophyll fluorescence

For high light stress treatment of *C. reinhardtii*, after 1 day of growth in TAP medium with or without 100 mg/l SA, cells were treated with high or normal light intensity for 1 day. For *M. polymorpha*, fresh thalli were sprayed with 500 $\mu$M SA in 0.01% Tween-20 or only 0.01% Tween-20 (mock). For *A. thaliana* and *O. sativa*, 4-week-old seedlings were sprayed with 500 $\mu$M SA in 0.01% Tween-20 or only 0.01% Tween-20 (mock). After 4 h of SA treatment, all *M. polymorpha*, *A. thaliana*, and *O. sativa* plants were treated with high or normal light intensity for 4 h. All high or normal light intensity conditions were provided by LED lights in a temperature- and humidity-controlled artificial climate chamber. The detailed photosynthetically active photon flux density values of high light treatment for each species are listed in Supplemental Table 4. The maximum operating efficiency of PSII photochemistry (Fv/Fm ratio) of all samples was measured in the dark by chlorophyll fluorometry using a FluorCam 800MF (https://fluorcams.psi.cz/) after 10 min of dark acclimation.

### Subcellular analysis of NPR–GFP fusions

The GFP fusion plasmids were transformed into protoplasts by polyethylene glycol-mediated transformation. Protoplast preparation and transfection followed previously described procedures (Xu et al., 2019). Nuclei were stained by DAPI. Confocal microscopy images were taken using a laser scanning confocal microscope (Leica TCS SP5). All primers used for the vector construction are shown in Supplemental Table 5.

### BiFC assay

The CDS sequences of OsNPR1, MpNPR1, OsTGA2, and MpTGA1 were cloned in frame with either the C-terminal or the N-terminal fragment of YFP. The specific primers used are listed in Supplemental Table 5. The constructs were transiently expressed in *Nicotiana benthamiana* leaves by *Agrobacterium*-mediated infiltration (stain EHA105), and YFP fluorescence was imaged as described in our previous study (Ruan et al., 2019). The experiment was repeated with at least three independent biological replicates.

### LUC complementation imaging assay

An LCI assay with the pCAMBIA-nLUC and pCAMBIA-cLUC vectors was used to detect protein–protein interactions in *N. benthamiana* leaves. The firefly LUC enzyme was divided into N-terminal (nLUC) and C-terminal (cLUC) parts. OsNPR1 and MpNPR1 were fused with nLUC in the pCAMBIA-nLUC vector. OsTGA2 and MpTGA1 were fused with cLUC in the pCAMBIA-cLUC vector. The vectors were co-infiltrated into *N. benthamiana* leaves, and LUC activity was analyzed 48 h after infiltration using NightSHADE LB 985 (Berthold). Before fluorescence observation, the leaves were kept in the dark for 15 min. At least three biological replicates were performed with similar results.

### Pull-down assay

OsNPR1–HIS, MpNPR1–HIS, GST–OsTGA2, GST–MpTGA1, and GST were expressed and purified from *E. coli* BL21 cell cultures using standard protocols. The pull-down assay was performed as described previously (Guo et al., 2022).

### Promoter–luciferase assay

Promoter activity assays were performed in *Arabidopsis* protoplasts by transforming the reporter constructs together with the different effector constructs, as described previously (Yoo et al., 2007; Ding et al., 2018). Protoplasts were prepared as previously described (Wu et al., 2009). In brief, a 35S-driven *Renilla* LUC reporter was used as the internal transfection control. After 16 h of incubation, protoplasts were collected, and the dual-LUC assay system (TransGen Biotech, TransDetect Double-Luciferase Reporter Assay Kit) was used to measure the activity of firefly LUC and *Renilla* LUC. The ratio of firefly LUC/*Renilla* LUC was used to calculate the relative LUC activity.

### Quantitative real-time PCR analysis

After 6 h of treatment with 10 mg/l SA, the total RNA of *Klebsormidium* sp. cells was extracted from frozen cell pellets using the Trizol reagent (Life Technologies) and further purified by phenol:chloroform:isoamyl alcohol (25:24:1) extraction and precipitation with ice-cold isopropanol. First-strand cDNA synthesis and gDNA removal were done using the TransScript cDNA Synthesis Kit (Transgen). Quantitative real-time PCR was performed using SYBR Premix kits (Roche) on a QuantStudio 6 Flex machine (Life Technologies). *KnACT1* (accession no. GAQ91093.1) and *KnPR1* (accession no. GAQ89259.1) were used as the internal control and positive control for SA treatment, respectively. The primer pairs used for quantitative real-time PCR are listed in Supplemental Table 5.

## DATA AVAILABILITY

All sequences, phylogenetic trees, and detailed information of proteins identified in this study have been submitted to the Figshare database (https://doi.org/10.6084/m9.figshare.14140691).

## SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

## FUNDING

This work was supported by the National Key Research and Development Program of China (2021YFF1000404), the National Natural Science Foundation of China (32130096), and the Central Public-interest Scientific Institution Basal Research Fund (Y2022QC14). K.Y. was supported by the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences.

## AUTHOR CONTRIBUTIONS

K.Y. conceived and supervised the project. X.J. performed the evolutionary analysis. X.J., L.W., H.Z., Y.Z., and L.X. performed the experiments. X.J., Z.C., L.X., and K.Y. wrote and revised the manuscript. All authors reviewed the manuscript.

## ACKNOWLEDGMENTS

We thank Dr. Lin Xu (CAS Center for Excellence in Molecular Plant Sciences) for providing *Marchantia polymorpha* materials. We thank Dr. Wenbin Zhou (Institute of Crop Science, Chinese Academy of Agricultural Sciences) and Ms. Huijuan Wang (Beijing EcoTech Science and Technology Ltd.) for their help with the determination of chlorophyll fluorescence. We thank the public laboratory of the Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, for SA measurement and for providing technical assistance. No conflict of interest is declared.

Received: July 7, 2022
Revised: November 10, 2022
Accepted: December 5, 2022
Published: December 6, 2022

## REFERENCES

**Bagal, U.R., Leebens-Mack, J.H., Lorenz, W.W., and Dean, J.F.** (2012). The phenylalanine ammonia lyase (PAL) gene family shows a gymnosperm-specific lineage. BMC Genom. **13**:1–9.

**Baggs, E.L., Monroe, J.G., Thanki, A.S., O'Grady, R., Schudoma, C., Haerty, W., and Krasileva, K.V.** (2020). Convergent loss of an EDS1/PAD4 signaling pathway in several plant lineages reveals coevolved components of plant immunity and drought response. Plant Cell **32**:2158–2177.

**Barros, J., and Dixon, R.A.** (2020). Plant phenylalanine/tyrosine ammonia-lyases. Trends Plant Sci. **25**:66–79.

**Cao, H., Glazebrook, J., Clarke, J.D., Volko, S., and Dong, X.** (1997). The Arabidopsis NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. Cell **88**:57–63.

**Castenholz, R.W.** (1988). Culturing methods for cyanobacteria. In Methods in Enzymology, L. Packer and A.N. Glazer, eds. (San Diego: Academic Press), pp. 68–93.

**Chen, Y.-E., Mao, H.-T., Wu, N., Mohi Ud Din, A., Khan, A., Zhang, H.-Y., and Yuan, S.** (2020). Salicylic acid protects photosystem II by alleviating photoinhibition in Arabidopsis thaliana under high light. Int. J. Mol. Sci. **21**:1229.

**Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T., Sun, W., Li, X., Xu, Y., Zhang, Y., et al.** (2019). Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. Cell **179**:1057–1067.e14.

**Ding, Y., Sun, T., Ao, K., Peng, Y., Zhang, Y., Li, X., and Zhang, Y.** (2018). Opposite roles of salicylic acid receptors NPR1 and NPR3/NPR4 in transcriptional regulation of plant immunity. Cell **173**:1454–1467.e15.

**Floková, K., Tarkowská, D., Miersch, O., Strnad, M., Wasternack, C., and Novák, O.** (2014). UHPLC–MS/MS based target profiling of stress-induced phytohormones. Phytochemistry **105**:147–157.

**Garcion, C., Lohmann, A., Lamodière, E., Catinot, J., Buchala, A., Doermann, P., and Métraux, J.P.** (2008). Characterization and biological function of the ISOCHORISMATE SYNTHASE2 gene of Arabidopsis. Plant Physiol. **147**:1279–1287.

**Garner, C.M., Kim, S.H., Spears, B.J., and Gassmann, W.** (2016). Express yourself: transcriptional regulation of plant innate immunity. Semin. Cell Dev. Biol. **56**:150–162.

**Gatz, C.** (2012). From pioneers to team players: TGA transcription factors provide a molecular link between different stress pathways. Mol. Plant Microbe Interact. **26**:151–159.

**Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S.** (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. **40**:D1178–D1186.

**Gross, J., Cho, W.K., Lezhneva, L., Falk, J., Krupinska, K., Shinozaki, K., Seki, M., Herrmann, R.G., and Meurer, J.** (2006). A plant locus essential for phylloquinone (vitamin K $_1$) biosynthesis originated from a fusion of four eubacterial genes. J. Biol. Chem. **281**:17189–17196.

**Guo, M., Ruan, W., Zhang, Y., Zhang, Y., Wang, X., Guo, Z., Wang, L., Zhou, T., Paz-Ares, J., and Yi, K.** (2022). A reciprocal inhibitory module for Pi and iron signaling. Mol. Plant **15**:138–150.

**Han, X., Chang, X., Zhang, Z., Chen, H., He, H., Zhong, B., and Deng, X.W.** (2019). Origin and evolution of core components responsible for monitoring light environment changes during plant terrestrialization. Mol. Plant **12**:847–862.

**Harris, E.H.** (1989). The Chlamydomonas Sourcebook (San Diego: Academic Press).

**Huang, W., Wang, Y., Li, X., and Zhang, Y.** (2020). Biosynthesis and regulation of salicylic acid and N-hydroxypipecolic acid in plant immunity. Mol. Plant **13**:31–41.

**Jiao, C., Sørensen, I., Sun, X., Sun, H., Behar, H., Alseekh, S., Philippe, G., Palacio Lopez, K., Sun, L., Reed, R., et al.** (2020). The Penium margaritaceum genome: hallmarks of the origins of land plants. Cell **181**:1097–1111.e12.

**Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al.** (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics **30**:1236–1240.

**Kachroo, P., Liu, H., and Kachroo, A.** (2020). Salicylic acid: transport and long-distance immune signaling. Curr. Opin. Virol. **42**:53–57.

**Kesarwani, M., Yoo, J., and Dong, X.** (2007). Genetic interactions of TGA transcription factors in the regulation of pathogenesis-related genes and disease resistance in Arabidopsis. Plant Physiol. **144**:336–346.

**Khan, M.I.R., Fatma, M., Per, T.S., Anjum, N.A., and Khan, N.A.** (2015). Salicylic acid-induced abiotic stress tolerance and underlying mechanisms in plants. Front. Plant Sci. **6**:462.

**León, J., Shulaev, V., Yalpani, N., Lawton, M.A., and Raskin, I.** (1995). Benzoic acid 2-hydroxylase, a soluble oxygenase from tobacco, catalyzes salicylic acid biosynthesis. Proc. Natl. Acad. Sci. USA. **92**:10413–10417.

**Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M., Eily, A., Koppers, N., Kuo, L.-Y., Li, Z., et al.** (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. Nat. Plants **4**:460–472.

**Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z., Fernandez-Pozo, N., Barker, M.S., Bennett, T., Blázquez, M.A., et al.** (2020). Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. Nat. Plants **6**:259–272.

**Li, X., Patena, W., Fauser, F., Jinkerson, R.E., Saroussi, S., Meyer, M.T., Ivanova, N., Robertson, J.M., Yue, R., Zhang, R., et al.** (2019). A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. Nat. Genet. **51**:627–635.

**Li, Z., Wakao, S., Fischer, B.B., and Niyogi, K.K.** (2009). Sensing and responding to excess light. Annu. Rev. Plant Biol. **60**:239–260.

**Liang, Z., Geng, Y., Ji, C., Du, H., Wong, C.E., Zhang, Q., Zhang, Y., Zhang, P., Riaz, A., Chachar, S., et al.** (2020). Mesostigma viride genome and transcriptome provide insights into the origin and evolution of Streptophyta. Adv. Sci. **7**:1901850.

**Liu, Y., Sun, T., Sun, Y., Zhang, Y., Radojičić, A., Ding, Y., Tian, H., Huang, X., Lan, J., Chen, S., et al.** (2020). Diverse roles of the salicylic acid receptors NPR1 and NPR3/NPR4 in plant immunity. Plant Cell **32**:4002–4016.

**Maberly, S.C.** (2014). The fitness of the environments of air and water for photosynthesis, growth, reproduction and dispersal of photoautotrophs: an evolutionary and biogeochemical perspective. Aquat. Bot. **118**:4–13.

**Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R.** (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. **37**:1530–1534.

**Moon, S.-J., Park, H.J., Kim, T.-H., Kang, J.-W., Lee, J.-Y., Cho, J.-H., Lee, J.-H., Park, D.-S., Byun, M.-O., Kim, B.-G., and Shin, D.** (2018). OsTGA2 confers disease resistance to rice against leaf blight by regulating expression levels of disease related genes via interaction with NH1. PLoS One **13**:e0206910.

**Nawrath, C., Heck, S., Parinthawong, N., and Métraux, J.P.** (2002). EDS5, an essential component of salicylic acid–dependent signaling for disease resistance in Arabidopsis, is a member of the MATE transporter family. Plant Cell **14**:275–286.

**Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K., Haas, F.B., Vanderstraeten, L., Becker, D., Lang, D., et al.** (2018). The Chara genome: secondary complexity and implications for plant terrestrialization. Cell **174**:448–464.e24.

**Peng, Y., Sun, T., and Zhang, Y.** (2017). Perception of salicylic acid in Physcomitrella patens. Front. Plant Sci. **8**:2145.

**Peng, Y., Yang, J., Li, X., and Zhang, Y.** (2021). Salicylic acid: biosynthesis and signaling. Annu. Rev. Plant Biol. **72**:761–791.

**Qin, Y., Torp, A.M., Glauser, G., Pedersen, C., Rasmussen, S.K., and Thordal-Christensen, H.** (2019). Barley isochorismate synthase mutant is phylloquinone-deficient, but has normal basal salicylic acid level. Plant Signal. Behav. **14**:1671122.

**Qualley, A.V., Widhalm, J.R., Adebesin, F., Kish, C.M., and Dudareva, N.** (2012). Completion of the core -oxidative pathway of benzoic acid biosynthesis in plants. Proc. Natl. Acad. Sci. USA. **109**:16383–16388.

**Rensing, S.A.** (2020). How plants conquered land. Cell **181**:964–966.

**Richmond, T.A., and Bleecker, A.B.** (1999). A defect in β-oxidation causes abnormal inflorescence development in Arabidopsis. Plant Cell **11**:1911–1924.

**Rivas-San Vicente, M., and Plasencia, J.** (2011). Salicylic acid beyond defence: its role in plant growth and development. J. Exp. Bot. **62**:3321–3338.

**Ro, D.-K., and Douglas, C.J.** (2004). Reconstitution of the entry point of plant phenylpropanoid metabolism in yeast (*Saccharomyces cerevisiae*): implications for control of metabolic flux into the phenylpropanoid pathway. J. Biol. Chem. **279**:2600–2607.

**Ruan, W., Guo, M., Wang, X., Guo, Z., Xu, Z., Xu, L., Zhao, H., Sun, H., Yan, C., and Yi, K.** (2019). Two RING-finger ubiquitin E3 ligases regulate the degradation of SPX4, an internal phosphate sensor, for phosphate homeostasis and signaling in rice. Mol. Plant **12**:1060–1074.

**Schwenkert, S., Fernie, A.R., Geigenberger, P., Leister, D., Möhlmann, T., Naranjo, B., and Neuhaus, H.E.** (2022). Chloroplasts are key players to cope with light and temperature stress. Trends Plant Sci. **27**:577–587.

**Shimogawara, K., Fujiwara, S., Grossman, A., and Usuda, H.** (1998). High-efficiency transformation of *Chlamydomonas reinhardtii* by electroporation. Genetics **148**:1821–1828.

**Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al.** (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. **7**:539.

**Su, D., Yang, L., Shi, X., Ma, X., Zhou, X., Hedges, S.B., and Zhong, B.** (2021). Large-scale phylogenomic analyses reveal the monophyly of bryophytes and Neoproterozoic origin of land plants. Mol. Biol. Evol. **38**:3332–3344. https://doi.org/10.1093/molbev/msab106.

**Tang, H., Krishnakumar, V., and Li, J.** (2015). jcvi: JCVI utility libraries. Zenodo **30**:2015, Advance Access published September. https://doi.org/10.5281/zenodo.31631.

**Torrens-Spence, M.P., Bobokalonova, A., Carballo, V., Glinkerman, C.M., Pluskal, T., Shen, A., and Weng, J.-K.** (2019). PBS3 and EPS1 complete salicylic acid biosynthesis from isochorismate in Arabidopsis. Mol. Plant **12**:1577–1586.

**Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., and Vandepoele, K.** (2018). Plaza 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Res. **46**:D1190–D1196.

**Wang, C., Liu, Y., Li, S.-S., and Han, G.-Z.** (2015). Insights into the origin and evolution of the plant hormone signaling machinery. Plant Physiol. **167**:872–886.

**Wang, L., Jia, X., Zhang, Y., Xu, L., Menand, B., Zhao, H., Zeng, H., Dolan, L., Zhu, Y., and Yi, K.** (2021). Loss of two families of SPX domain-containing proteins required for vacuolar polyphosphate accumulation coincides with the transition to phosphate storage in green plants. Mol. Plant **14**:838–846.

**Wang, L., Tsuda, K., Truman, W., Sato, M., Nguyen, L.V., Katagiri, F., and Glazebrook, J.** (2011). CBP60g and SARD1 play partially redundant critical roles in salicylic acid signaling. Plant J. **67**:1029–1041.

**Wang, W., Withers, J., Li, H., Zwack, P.J., Rusnac, D.-V., Shi, H., Liu, L., Yan, S., Hinds, T.R., Guttman, M., et al.** (2020). Structural basis of salicylic acid perception by Arabidopsis NPR proteins. Nature **586**:311–316.

**Wang, X., Zeng, L., Liao, Y., Zhou, Y., Xu, X., Dong, F., and Yang, Z.** (2019). An alternative pathway for the formation of aromatic aroma compounds derived from l-phenylalanine via phenylpyruvic acid in tea (Camellia sinensis (L.) O. Kuntze) leaves. Food Chem. X. **270**:17–24.

**Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.h., Jin, H., Marler, B., Guo, H., et al.** (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. **40**:e49.

**Wildermuth, M.C., Dewdney, J., Wu, G., and Ausubel, F.M.** (2001). Isochorismate synthase is required to synthesize salicylic acid for plant defence. Nature **414**:562–565.

**Wu, F.-H., Shen, S.-C., Lee, L.-Y., Lee, S.-H., Chan, M.-T., and Lin, C.-S.** (2009). Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. Plant Methods **5**:16.

**Wu, X., and Bartel, D.P.** (2017). kpLogo: positional k-mer analysis reveals hidden specificity in biological sequences. Nucleic Acids Res. **45**:W534–W538.

**Xu, L., Zhao, H., Ruan, W., Deng, M., Wang, F., Peng, J., Luo, J., Chen, Z., and Yi, K.** (2017). ABNORMAL INFLORESCENCE MERISTEM1 functions in salicylic acid biosynthesis to maintain proper reactive oxygen species levels for root meristem activity in rice. Plant Cell **29**:560–574.

**Xu, L., Zhao, H., Wan, R., Liu, Y., Xu, Z., Tian, W., Ruan, W., Wang, F., Deng, M., Wang, J., et al.** (2019). Identification of vacuolar phosphate efflux transporters in land plants. Nat. Plants **5**:84–94.

**Xu, L., Zhao, H., Wang, J., Wang, X., Jia, X., Xu, Z., Li, R., Jiang, K., Chen, Z., Luo, J., et al.** (2022). AIM1-dependent high basal SA accumulation modulates stomatal aperture in rice. Preprint at bioRxiv **25**:2022, Advance Access published June. https://doi.org/10.1101/2022.06.23.497111.

**Yang, Y., Qi, M., and Mei, C.** (2004). Endogenous salicylic acid protects rice plants from oxidative damage caused by aging as well as biotic and abiotic stress: anti-oxidative role of salicylic acid in rice. Plant J. **40**:909–919.

**Yoo, S.-D., Cho, Y.-H., and Sheen, J.** (2007). Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. **2**:1565–1572.

**Zeng, L., Wang, X., Tan, H., Liao, Y., Xu, P., Kang, M., Dong, F., and Yang, Z.** (2020). Alternative pathway to the formation of trans-cinnamic acid derived from l-phenylalanine in tea (Camellia sinensis) plants and other plants. J. Agric. Food Chem. **68**:3415–3424.

**Zhang, J., Fu, X.-X., Li, R.-Q., Zhao, X., Liu, Y., Li, M.-H., Zwaenepoel, A., Ma, H., Goffinet, B., Guan, Y.-L., et al.** (2020). The hornwort genome and early land plant evolution. Nat. Plants **6**:107–118.

**Zhang, Z., Ma, X., Liu, Y., Yang, L., Shi, X., Wang, H., Diao, R., and Zhong, B.** (2022). Origin and evolution of green plants in the light of key evolutionary events. J. Integr. Plant Biol. **64**:516–535.

**Zhou, J.-M., and Zhang, Y.** (2020). Plant immunity: danger perception and signaling. Cell **181**:978–989.