



A species-specific functional module controls formation of pollen apertures

Byung Ha Lee^{1,4}, Rui Wang^{1,4}, Ingrid M. Moberg^{ID 1,2}, Sarah H. Reeder¹, Prativa Amom¹, Michelle H. Tan¹, Katelyn Amstutz^{ID 1}, Pallavi Chandna¹, Adam Helton^{ID 1}, Ekaterina P. Andrianova³, Igor B. Zhulin^{ID 3} and Anna A. Dobritsa^{ID 1}

Pollen apertures are an interesting model for the formation of specialized plasma-membrane domains. The plant-specific protein INP1 serves as a key aperture factor in such distantly related species as *Arabidopsis*, rice and maize. Although INP1 orthologues probably play similar roles throughout flowering plants, they show substantial sequence divergence and often cannot substitute for each other, suggesting that INP1 might require species-specific partners. Here, we present a new aperture factor, INP2, which satisfies the criteria for being a species-specific partner for INP1. Both INP proteins display similar structural features, including the plant-specific DOG1 domain, similar patterns of expression and mutant phenotypes, as well as signs of co-evolution. These proteins interact with each other in a species-specific manner and can restore apertures in a heterologous system when both are expressed but not when expressed individually. Our findings suggest that the INP proteins form a species-specific functional module that underlies formation of pollen apertures.

Pollen grains of flowering plants are surrounded by a robust wall, called exine. In most species, exine is deposited on the pollen surface non-uniformly, with certain regions of the surface receiving little to no exine material¹. These regions develop into pollen apertures that help pollen to hydrate, change volume and germinate^{2–6}. Across species, apertures vary greatly in their number, positions and morphology, contributing to diverse, species-specific patterns on the pollen surface^{1,2,7,8}. Recently, we and others have demonstrated that before forming apertures, developing pollen forms distinct aperture domains in their plasma membrane, which accumulate specific combinations of proteins and lipids^{6,9–11}. Apertures can thus be used to study how cells develop polarity and form membrane domains, as well as to understand how these mechanisms evolved to create the tremendous diversity of aperture patterns found in nature.

Aperture domains of the plasma membrane appear at the tetrad stage of pollen development, during which the four products of male meiosis (microspores) are transiently kept together under the common callose wall^{6,10}. The positions, number and morphology of the aperture membrane domains in microspores correspond to the aperture pattern of mature pollen. For example, in *Arabidopsis* pollen, apertures are shaped like three long and narrow meridional furrows (Fig. 1a). Accordingly, in *Arabidopsis* tetrads, each microspore develops three linear meridional domains of the plasma membrane, which attract the proteins D6 PROTEIN KINASE-LIKE3 (D6PKL3) and INAPERTURATE POLLEN1 (INP1)^{10,11}. In contrast, in rice and other grasses, pollen has only one small round aperture, positioned at the distal pole. Correspondingly, the tetrad-stage microspores in rice develop at their distal poles a single aperture domain shaped like a tiny ring, which attracts the rice orthologue of INP1 (OsINP1)⁶.

In both *Arabidopsis* and rice, as well as in maize, the INP1 protein was shown to be a major aperture factor whose loss causes a

complete loss of apertures (Fig. 1b)^{6,9,12}. INP1 is a plant-specific protein with a single recognizable domain, the DELAYED IN GERMINATION1 (DOG1) domain, whose function is unknown⁹. Although the biochemical function of INP1 remains to be identified, in both *Arabidopsis* and rice these proteins appear to play a role in keeping the aperture domains of the plasma membrane in close contact with the overlying callose wall^{6,10}, which possibly protects these domains from the deposition of exine materials.

Since the role of INP1 as an essential aperture factor is conserved in such distantly related species as *Arabidopsis*, rice and maize, it is reasonable to assume that INP1 orthologues across angiosperms are probably all involved in aperture formation. Intriguingly, though, many INP1 proteins show substantial sequence divergence and cannot substitute for the loss of *Arabidopsis* INP1 (refs. 9,12). This suggests that, despite their conserved involvement in aperture formation, INP1 proteins are probably functionally species-specific. We have previously proposed that such species specificity might be due to the presence of unknown aperture factors that have co-evolved with INP1 and help it to perform its function¹².

Here, we present an aperture factor, INP2, that fulfills the role of a species-specific partner for INP1. INP2 resembles INP1 in its protein structure, patterns of expression, trends of evolutionary divergence, mutant phenotype and genetic interactions. We provide evidence that INP2 is also functionally species-specific and that it physically interacts with INP1. Furthermore, we demonstrate that tomato orthologues of INP1 and INP2, which are unable to restore apertures in *Arabidopsis* mutants when only one of them is expressed, gain the ability to function in *Arabidopsis* when expressed together. The two INP proteins, therefore, behave as co-evolved species-specific partners that form a functional module required for the formation of pollen apertures.

¹Department of Molecular Genetics and Center for Applied Plant Sciences, Ohio State University, Columbus, OH, USA. ²Norwegian Science and Technology University, Ålesund, Norway. ³Department of Microbiology, Ohio State University, Columbus, OH, USA. ⁴These authors contributed equally: Byung Ha Lee, Rui Wang. e-mail: dobritsa.1@osu.edu

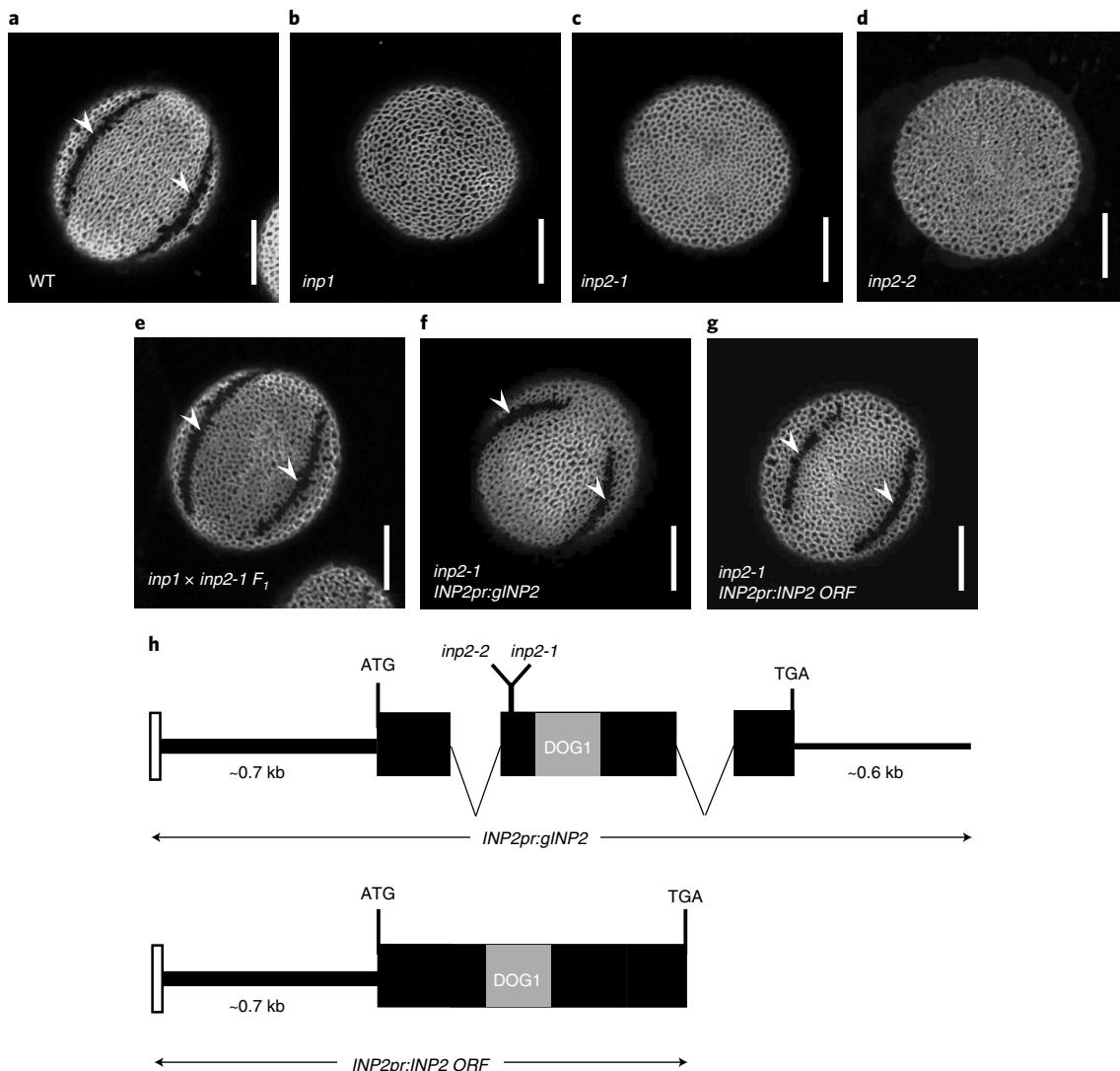


Fig. 1 | INP2 is a new factor essential for the formation of pollen apertures. **a–g,** Confocal images of pollen grains stained with auramine O. Scale bars, 10 µm. **a,** Wild-type (WT) *Arabidopsis* pollen has three equidistant furrow-like apertures (two are visible here, arrowheads). **b,** *inp1* pollen completely lacks apertures. **c,d,** Similar to *inp1*, pollen of *inp2-1* (**c**) and *inp2-2* (**d**) mutants has normal exine but completely lacks apertures (>100 pollen grains were imaged, with similar results; for *inp2-2*, two independent CRISPR plants were obtained, producing similar phenotypes). **e,** Pollen of the F₁ progeny of the cross between *inp1* and *inp2* develops normal apertures (arrowheads), indicating that mutations disrupt different genes (eight plants (≥50 pollen grains per plant) were imaged, with similar results). **f,g,** *INP2pr:gINP2* (**f**) and *INP2pr:INP2 ORF* (**g**) transgenes restore apertures (arrowheads) in *inp2* (7/8 and 15/15 independent T₁ lines, respectively; ≥50 pollen grains per line were imaged, with similar results). **h,** *INP2* gene model and structure of the *INP2pr:gINP2* and *INP2pr:INP2 ORF* complementation constructs. Black boxes indicate the protein-coding sequence of At1g15320. The region encoding the DOG1 domain is indicated by the grey box. The white box denotes a short region from the preceding gene, At1g15330, which was included in the constructs. Both the ~0.7-kb upstream region and the ~0.6-kb downstream region were included in the genomic construct. Positions of the *inp2-1* and *inp2-2* mutations are indicated on the gene model.

Results

A new *Arabidopsis* mutant has the inaperturate pollen phenotype identical to the phenotype of the *inp1* mutant. To discover genes involved in the formation of pollen apertures, we performed a forward genetic screen on an M₂ population of *Arabidopsis* plants mutagenized with ethyl methanesulfonate. Since changes in pollen shape can serve as a proxy for aperture formation defects^{13,14}, we screened these plants for unusual pollen shapes under dissecting microscopes. One mutant produced pollen that looked much rounder than the wild-type pollen, strongly resembling the phenotype of the *inp1* mutants. An examination by confocal microscopy showed that, like *inp1*, pollen of this mutant completely lacks apertures (inaperturate phenotype) but had otherwise normal exine (Fig. 1c).

To test whether the mutation represented an allele of *INP1* or disrupted another gene, we crossed the new mutant with the *inp1-1* null mutant. In the F₁ progeny of this cross, all pollen had normal apertures (Fig. 1e), demonstrating that the defect affected a gene other than *INP1*. This result also showed that, similar to *inp1* and other previously discovered aperture mutants, the new mutation affected a gene with the sporophytic function. Because of the similarities with the *inp1* mutant, we named the new gene *INAPERTURATE POLLEN2* (*INP2*) and its mutant allele *inp2-1*.

The *inp2-1* mutation disrupts the At1g15320 gene. Using positional cloning, we mapped the *inp2-1* defect to a 146-kilobase (kb) region at the top of chromosome 1, containing 51 genes. To

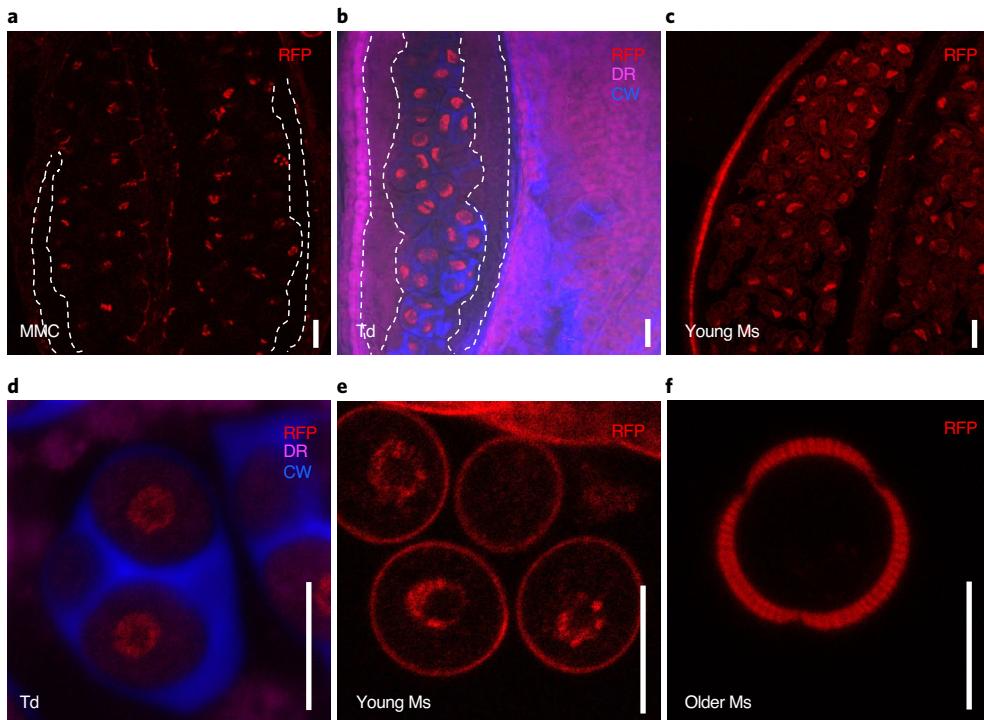


Fig. 2 | INP2 is expressed in the male reproductive lineage at the time of aperture formation. Images of anthers at different developmental stages (a–c) and magnified images of the cells from the male reproductive lineage at different developmental stages (d–f) expressing the transcriptional fusion construct *INP2pr:H2B-RFP*. Nuclear signal of H2B-RFP (red) is found in dividing microspore mother cells (MMC; a), tetrads of microspores (Td; b,d) and young free microspores (Ms; c,e). Older microspores (f) do not show nuclear H2B-RFP signal (peripheral red signal is due to the autofluorescence of the developing exine). No signal was observed in the tapetal layer of the anther (outlined by the white dashed lines in a,b). Besides RFP, the images in b,d show staining for callose wall (blue, CW, calcofluor white) and membranous structures (magenta, DR, CellMask Deep Red). Five independent T_1 lines were imaged, with similar results. Scale bars, 10 μm .

narrow down the list of gene candidates, we inspected their predicted identities as well as patterns of their messenger RNA expression reported in the TRAVA RNA-seq database¹⁵. We focused on the genes expressed in young flower buds (flowers 12–18 in the TRAVA nomenclature), as these buds include the tetrad stage of development associated with aperture formation. One gene, At1g15320, was prioritized as a particularly strong candidate as it is predicted to be expressed nearly exclusively in young buds (Extended Data Fig. 1) and encodes a protein with structural similarities to INP1 (below). Sequencing of this gene from *inp2-1* revealed a G-to-A substitution which created an early stop codon (Trp84Stop) (Fig. 1h). To independently confirm that *INP2* is At1g15320, we targeted At1g15320 in the wild-type Col-0 background with CRISPR-Cas9 and generated an allele (*inp2-2*) with a two-nucleotide deletion that caused a frame shift after the amino acid 83 (Fig. 1h). The CRISPR mutant displayed the same inaperturate pollen phenotype as the original *inp2-1* allele (Fig. 1d).

To further verify the identity of At1g15320 as *INP2* and to define its regulatory regions, we created transgenic constructs containing either the genomic region of At1g15320 (including introns and the ~0.6-kb region downstream of the stop codon) or its open reading frame (ORF) (Fig. 1h). These constructs were placed under the control of the putative native promoter (a region of ~0.7-kb between the start codon of At1g15320 and the preceding gene At1g15330) and transformed into *inp2-1*. Both constructs successfully restored apertures in transgenic plants—15/15 T_1 plants with the ORF construct and 7/8 T_1 plants with the genomic construct (Fig. 1f–g). Taken together, our results demonstrate that (1) At1g15320 encodes INP2 (a new factor essential for aperture formation) and (2) that the 0.7-kb upstream region is sufficient to drive functional expression

of *INP2*. This promoter region was then used for all subsequent *INP2* constructs transformed into *Arabidopsis*.

INP2 shares structural similarity with INP1. INP2 is a plant-specific protein of unknown biochemical function which shares certain similarities with INP1. Both proteins have similar size (273 amino acids for INP1 versus 307 amino acids for INP2), are usually encoded in angiosperm genomes by single-copy genes and contain the same domain—the plant-specific DELAYED IN GERMINATION1 (DOG1) domain (PFam14144) (Fig. 1 and Extended Data Fig. 2a). This domain, typically associated with seed dormancy proteins and TGA bZIP transcription factors¹⁶, is the only recognizable domain in both INP proteins. Interestingly, although INP1 and INP2 share only limited homology with each other (23% sequence identity; Extended Data Fig. 2a), the protein-fold recognition software Phyre2 (ref. ¹⁷) selected the same template for homology modeling of both proteins and predicted similar structures, with three alpha-helices, for their C-terminal regions (Extended Data Fig. 2b,c).

Protein alignments of INP1 and INP2 with their respective orthologues from other plants also revealed that, in eudicots, these proteins typically contain a region enriched in Asp and Glu residues. However, these acidic regions are positioned differently between INP1 and INP2. In the INP1 proteins, the acidic region follows the DOG1 domain^{9,12}, whereas in the INP2 proteins it is located ahead of the DOG1 domain (Extended Data Fig. 3).

For *Arabidopsis thaliana* INP2 (AtINP2), multiple algorithms also predicted the existence of a transmembrane (TM) domain at its N terminus (Extended Data Fig. 4), with most of the protein expected to be outside the cell, facing the extracellular space. Yet,

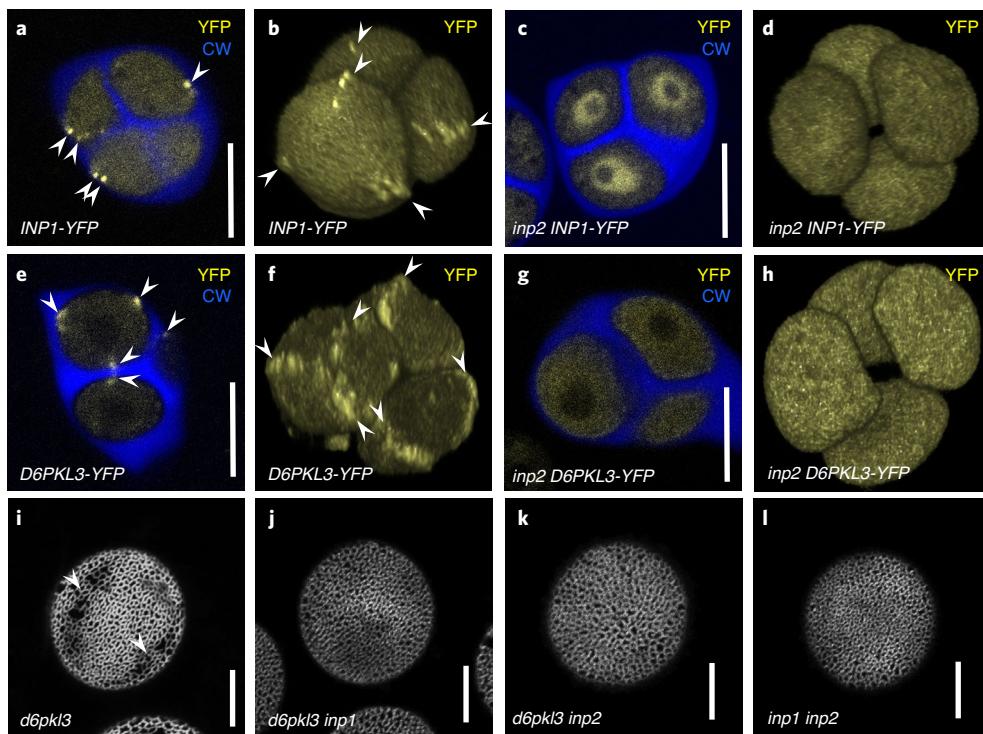


Fig. 3 | INP2 is required for INP1 and D6PKL3 accumulation at the aperture domains and both *inp1* and *inp2* are epistatic to *d6pk3*. **a–h**, INP1-YFP and D6PKL3-YFP localization in tetrads of microspores in the presence and absence of INP2. Confocal optical sections (**a,c,e** and **g**) and three-dimensional reconstructions of tetrads of microspores (**b,d,f** and **h**). YFP signal is shown in yellow and callose wall (CW, stained by calcofluor white) is shown in blue. Arrowheads point to the YFP signal at the aperture domains. INP1-YFP localizes to the aperture PM domains in the wild type (**a,b**) but loses this localization in the *inp2* mutant (**c,d**), instead becoming enriched in the nucleoplasm. Experiments in (**c,d**) were repeated three times, with similar results. Likewise, D6PKL3-YFP localizes to the aperture domains in the wild type (**e,f**) but loses this localization in the *inp2* mutant (**g,h**). Experiments in **g,h** were repeated two times, with similar results. **i–l**, *inp1* and *inp2* mutations are epistatic to *d6pk3* and do not cause additional phenotypic changes when combined. Confocal images of pollen grains stained with auramine O. The *d6pk3* mutant pollen often develops apertures partially covered with exine (arrowheads) (**i**), whereas double mutants *d6pk3 inp1* (**j**), *d6pk3 inp2* (**k**) and *inp1 inp2* (**l**) completely lack apertures. Three or more plants (≥ 50 pollen grains per plant) were imaged in **i–l**, with similar results. Scale bars, 10 μ m.

the algorithms failed to identify a TM domain in many orthologues of AtINP2, including the highly related proteins from *A. lyrata* and other members of the Brassicaceae family, suggesting that this is not a common feature of INP2 proteins. No lipid modifications are predicted for INP2.

INP2 is expressed in the developmental lineage of pollen at the time of aperture formation. Publicly available RNA-seq data show that, like INP1, INP2 is expressed nearly exclusively in young buds containing pollen at or around the tetrad stage during which apertures form (Extended Data Fig. 1). To test whether

in these buds INP2 is expressed in the male reproductive lineage, we expressed the nuclear marker histone H2B tagged with red fluorescent protein (RFP) under the control of the INP2 promoter (*INP2pr:H2B-RFP*) in the wild-type Col-0 plants. This reporter, with its concentrated localization in the nucleus, was specifically chosen to help visualize the expression from the INP2 promoter, since, like INP1, INP2 is predicted to be expressed at low levels (Extended Data Fig. 1). The nuclear RFP signal was found in the dividing microspore mother cells, tetrad-stage microspores and young free microspores (Fig. 2). The signal was absent in older microspores, the surrounding somatic tapetal cell layer and other

Fig. 4 | INP1 and INP2 physically interact. **a**, Yeast two-hybrid assay of interaction between INP1 and INP2^{ΔN} (lacking the N-terminal region). BD, DNA-binding domain; AD, activating domain; SD, synthetic defined medium. To test for the presence of both BD and AD constructs, leucine (L) and tryptophan (W) were excluded from the medium. To test for protein interaction, yeast were grown on media lacking L, W and histidine (H) and containing 20 mM 3-aminotriazole (3-AT). **b**, BiFC experiments. INP1 and INP2 proteins fused, respectively, to the N- and C-terminal parts of YFP (YN and YC) were cotransformed into tobacco leaves to test for interaction. Cotransformation of INP1-YN with only YC and cotransformation of INP2-YC with only YN were used as negative controls. Top panels show YFP signal in leaf epidermis. Bottom panels show merged YFP and bright-field images. Scale bars, 50 μ m. **c**, Co-immunoprecipitation experiments. INP1-HA₃/INP2-GFP and INP1-GFP/INP2-HA₃ pairs (or just single tagged proteins as negative controls) were co-expressed in tobacco leaves, precipitated with anti-GFP and visualized with anti-GFP or anti-HA. IP, immunoprecipitation; IB, immunoblot. ‘Mock’ indicates protein extract from leaves infiltrated only with buffer. **d**, Split-luciferase assay. Tobacco leaves were divided into sectors co-expressing indicated proteins containing the N-terminal (NLuc) and C-terminal (CLuc) parts of the firefly luciferase. Panels on the left show the bright-field images and panels on the right show the corresponding luminescence images. **e**, Y2H assay in which the DOG1 domains of INP1 (INP1^{DOG1}) and INP2 (INP2^{DOG1}) were tested for interaction with each other, self-interaction and interaction with the full-length INP1 and with INP2^{ΔN}. The description is the same as for **a**, except that 3 mM 3-AT was used here. Experiments in **a–c** and **e** were repeated three times and experiments in **d** were repeated two times (each time using multiple leaves from multiple plants), with similar results.

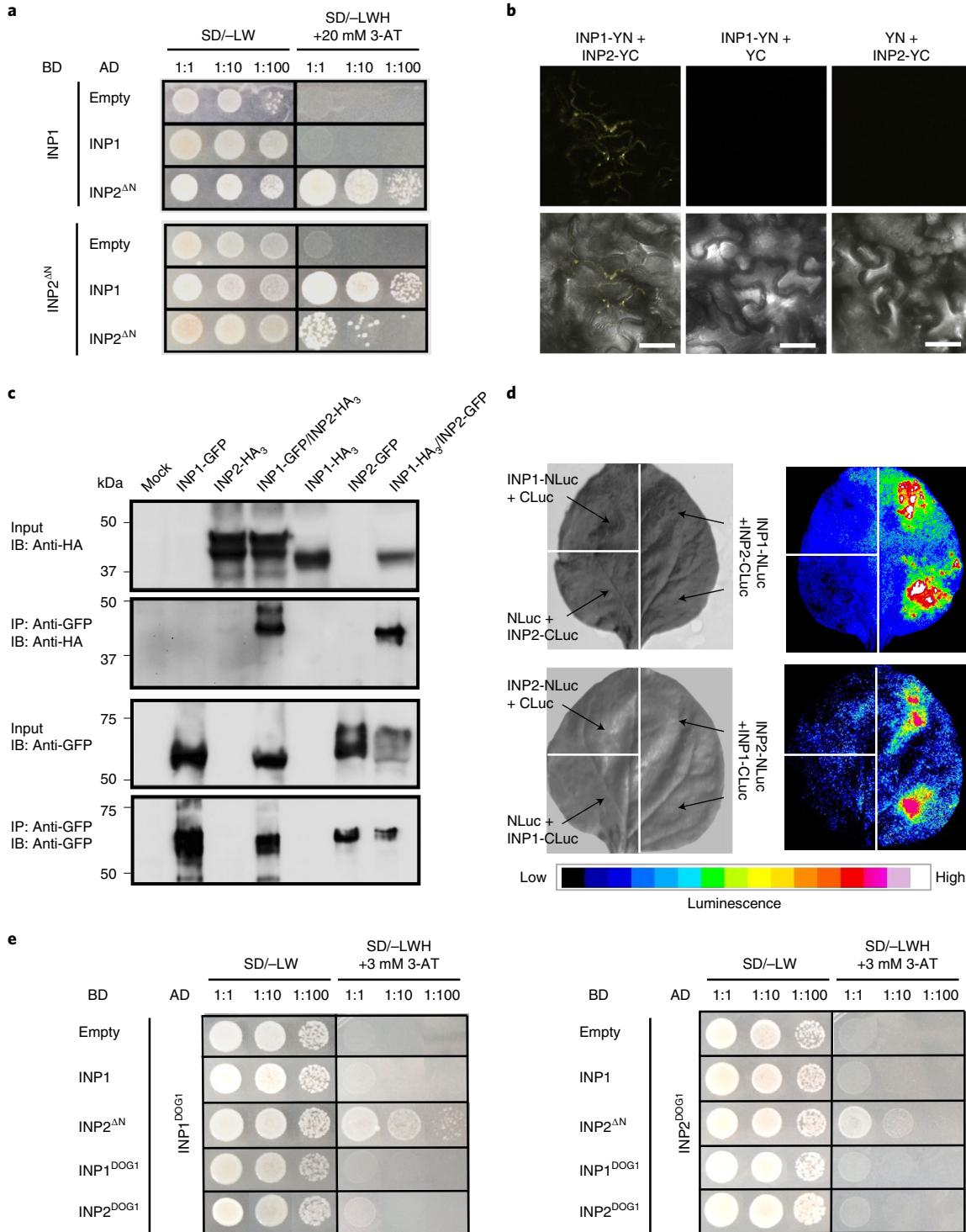
anther layers (Fig. 2). This expression pattern matches that of INP1 (refs. ^{9,10}).

To visualize the subcellular localization of the INP2 protein, we first created five constructs, in which INP2 was tagged with yellow fluorescent protein (YFP) at four positions: at the N terminus, at the C terminus (either directly or following an 18-amino acid linker), after the predicted TM region and internally—within the low-conservation region (below). However, none of the YFP-tagged constructs rescued the *inp2* mutant, suggesting that INP2 does not tolerate addition of sizable tags. This notion was supported by further experiments in which partial rescue of the mutant phenotype was

achieved with constructs expressing INP2 tagged at the C terminus with one or three copies of the small HA tag. Of these two types of constructs, the shorter HA₁ construct produced better rescue (Methods), yet no protein signal was detected in these lines with anti-HA in anther sections or whole-mount preparations, possibly owing to the low levels of the INP2 expression. This prevented us from determining whether INP2, like INP1, specifically localizes to the aperture domains in the plasma membrane of microspores.

Localization of INP1 and D6PKL3 to plasma-membrane aperture domains depends on the presence of INP2.

To test whether INP2



Tree scale: 1

Plant taxa

- Basal angiosperms
- Monocots—Arecales
- Monocots—Poales
- Basal eudicots
- Basal core eudicots
- Asterids—Asterales
- Rosids—Vitales
- Asterids—Gentianales
- Asterids—Ericales
- Asterids—Apiales
- Asterids—Solanales
- Asterids—Lamiaceae
- Rosids—Malvales
- Rosids—Fagales
- Rosids—Rosales
- Rosids—Malpighiales
- Rosids—Sapindales
- Rosids—Brassicaceae

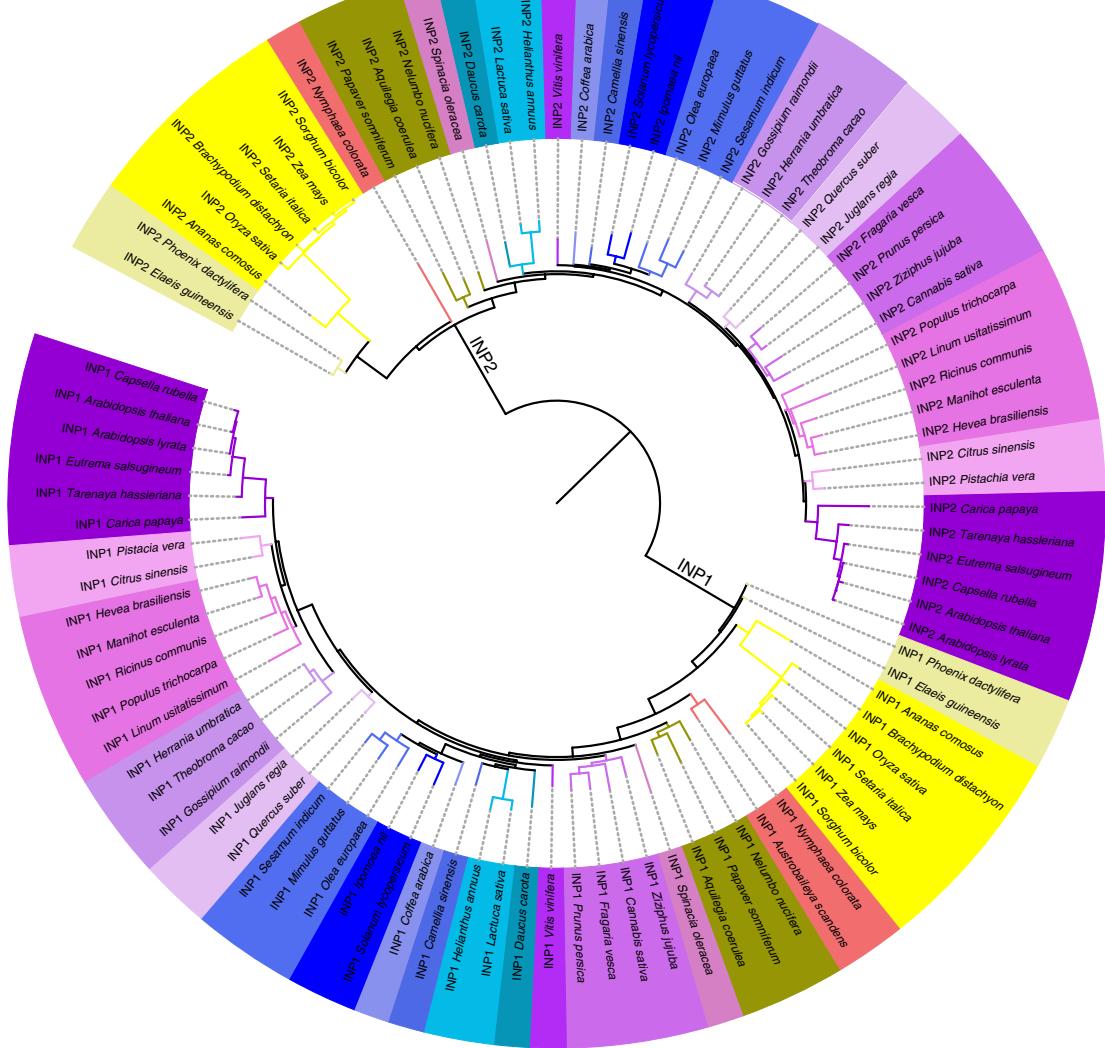


Fig. 5 | INP1 and INP2 exhibit similar trends of evolutionary sequence divergence. Maximum likelihood phylogenetic tree of INP1 and INP2 sequences from a variety of angiosperm taxa (indicated by colour coding). The INP1 and INP2 sequences cluster into two separate clades, which display similar topology.

contributes to the distinct positioning of INP1 and another recently identified aperture factor, D6PKL3, both of which accumulate at the microspore aperture domains (Fig. 3a,b,e,f), we introgressed the previously characterized reporter constructs *DMC1pr:INP1-YFP* (ref. ¹⁰) and *D6PKL3pr:D6PKL3-YFP* (ref. ¹¹) into the *inp2* mutant background. In the absence of INP2, INP1-YFP failed to localize to the aperture domains of the plasma membrane, instead showing notable enrichment in the nucleoplasm (Fig. 3c,d). This result suggests that INP2 is involved either in targeting INP1 to the aperture domains or in keeping it there. Likewise, in the absence of INP2, the membrane-associated kinase D6PKL3-YFP lost its association with the aperture domains, instead displaying diffuse cytoplasmic localization (Fig. 3g,h). As D6PKL3 reacts the same way to the absence of INP1 (ref. ¹¹), both INP1 and INP2 are thus required to keep it at the aperture domains.

To test for epistatic relationships between these aperture factors, we created double mutants of *inp1 d6pkl3* and *inp2 d6pkl3*. Single mutations in D6PKL3 do not completely abolish aperture formation, instead producing ‘shadows of apertures’ that are partially

covered with exine (Fig. 3i)¹¹. Yet both double mutants produced completely inaperturate pollen (Fig. 3j,k), indicating that *inp1* and *inp2* are both epistatic to *d6pkl3*. To investigate the possibility of synergistic interactions between INP1 and INP2, we also created the *inp1 inp2* double mutant. Its phenotype, however, was identical to those of single mutants (Fig. 3l), showing that the simultaneous loss of INP1 and INP2 does not cause any additional observable effects (for example, in the exine deposition) and suggesting that these proteins behave as bona fide aperture factors. Taken together, the results presented so far are consistent with the notion that INP1 and INP2 occupy very similar positions in the aperture formation pathway and might coordinate their activities.

INP1 and INP2 are interacting proteins. Since INP1 and INP2 exhibit similarities in their protein structures, patterns of expression, mutant phenotypes and genetic interactions, we suggested that they might physically interact. To explore this possibility, we used several approaches. An initial yeast two-hybrid (Y2H) assay with the full-length INP1 and INP2 did not result in yeast growth

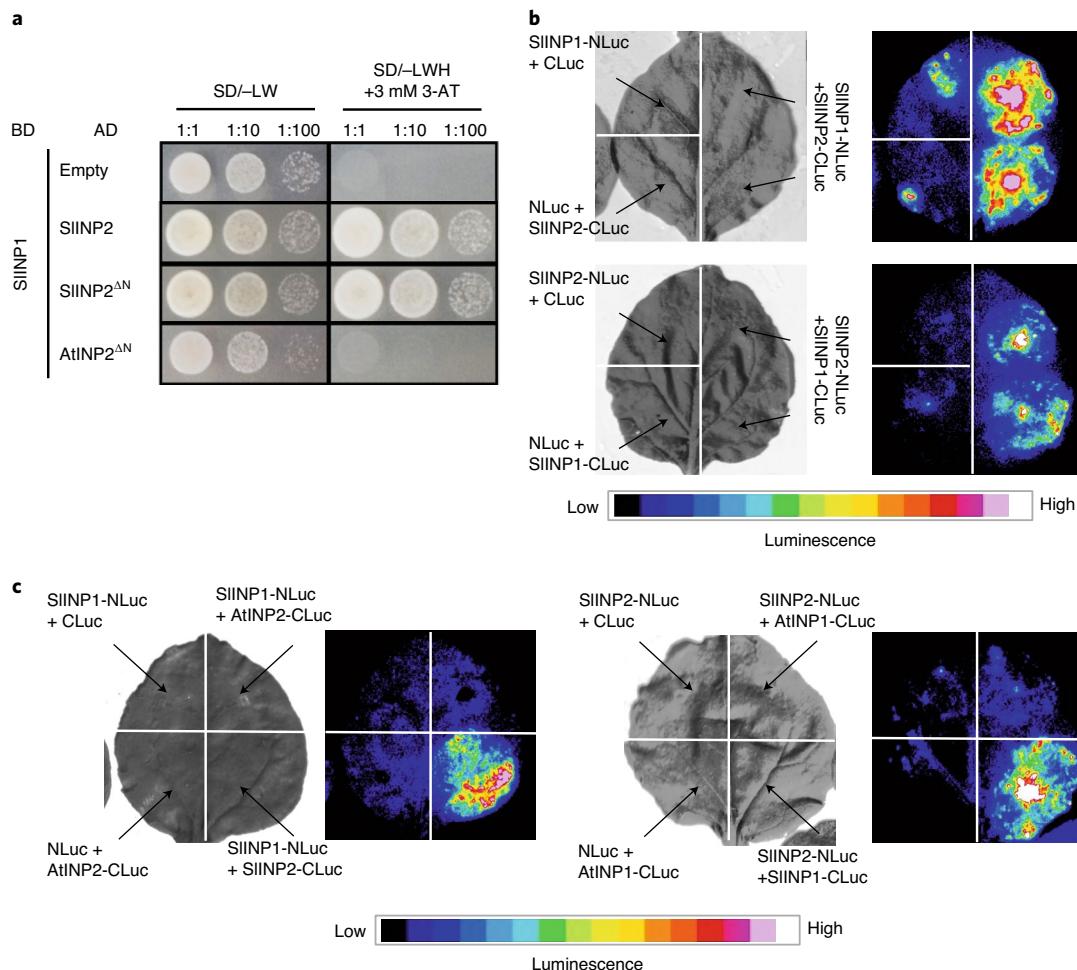


Fig. 6 | INP1 and INP2 interact in a species-specific manner. **a**, Y2H assay testing SIINP1 interactions with SIINP2 (or SIINP2^{ΔN} lacking the N-terminal region) and AtINP2^{ΔN}. To test for the presence of both BD and AD constructs, leucine (L) and tryptophan (W) were excluded from the medium. To test for protein interaction, yeast were grown on media lacking L, W and histidine (H) and containing 3 mM 3-aminotriazole (3-AT). **b**, Split-luciferase assay testing the ability of SIINP1 and SIINP2 to interact. Tobacco leaves were divided into sectors co-expressing indicated proteins containing the N-terminal (NLuc) and C-terminal (CLuc) parts of the firefly luciferase. Panels on the left show the bright-field images and panels on the right show the corresponding luminescence images. **c**, Split-luciferase assay testing the ability of INP1 and INP2 from *Arabidopsis* and tomato to interact with a protein from another species. Only the same-species interactions were observed. The description is the same as for **b**. All experiments were repeated at least twice, with similar results.

indicative of protein interaction. We reasoned, however, that lack of yeast growth would be expected if INP2 indeed had a TM domain at its N terminus and most of the protein was extracellular.

We, therefore, expressed INP2 in yeast without its first 24 amino acids, which contained the predicted TM domain. This truncated INP2 (INP2^{ΔN}) showed strong interaction with INP1 in the Y2H system (Fig. 4a). In addition, this assay revealed that INP2 may be able to self-interact (Fig. 4a). We further verified the ability of INP1 and INP2 to interact in planta by expressing them in tobacco leaf cells and performing co-immunoprecipitation, bifluorescent molecular complementation (BiFC) and a split-luciferase assay (Fig. 4b–d).

The DOG1 domain is the only recognizable protein domain present in these proteins. Although its function is unknown, it has been proposed that this domain might participate in protein–protein interactions¹⁸. We therefore used the Y2H assay to test the ability of the DOG1 domains from INP1 and INP2 to interact with each other and with the full-length (or nearly full-length in the case of INP2^{ΔN}) proteins (Fig. 4e). INP1^{DOG1} was able to interact with INP2^{ΔN}. In contrast, INP2^{DOG1} failed to interact with INP1 but showed some ability to interact with INP2^{ΔN}, consistent with the finding that INP2

may self-interact. However, no interactions occurred when only the DOG1 domains were present (Fig. 4e), suggesting that these regions probably interact with other portions of INP2.

INP1 and INP2 exhibit similar trends of evolutionary sequence divergence. We previously reported that INP1 greatly diversified in angiosperm lineages^{9,12}. Still, in several species these divergent orthologues were found to be involved in the formation of pollen apertures and able to localize to specific plasma-membrane aperture domains^{9,12}, suggesting that, despite the substantial difference in primary sequences, all INP1 proteins in angiosperms probably function as aperture factors. However, INP1 proteins appear to exhibit a notable degree of functional species specificity, since the divergent INP1 orthologues were not able to complement the aperture defects of the *Arabidopsis* *inp1* mutant¹². A possible interpretation of this result is that INP1 proteins might require the presence of co-evolved partners to perform their function.

To see whether INP2 shows signs of co-evolution with INP1, we performed BLAST searches for INP2 homologues followed by phylogenetic analysis, revealing notable parallels between INP1 and INP2. Although proteins with the DOG1 domain appeared as

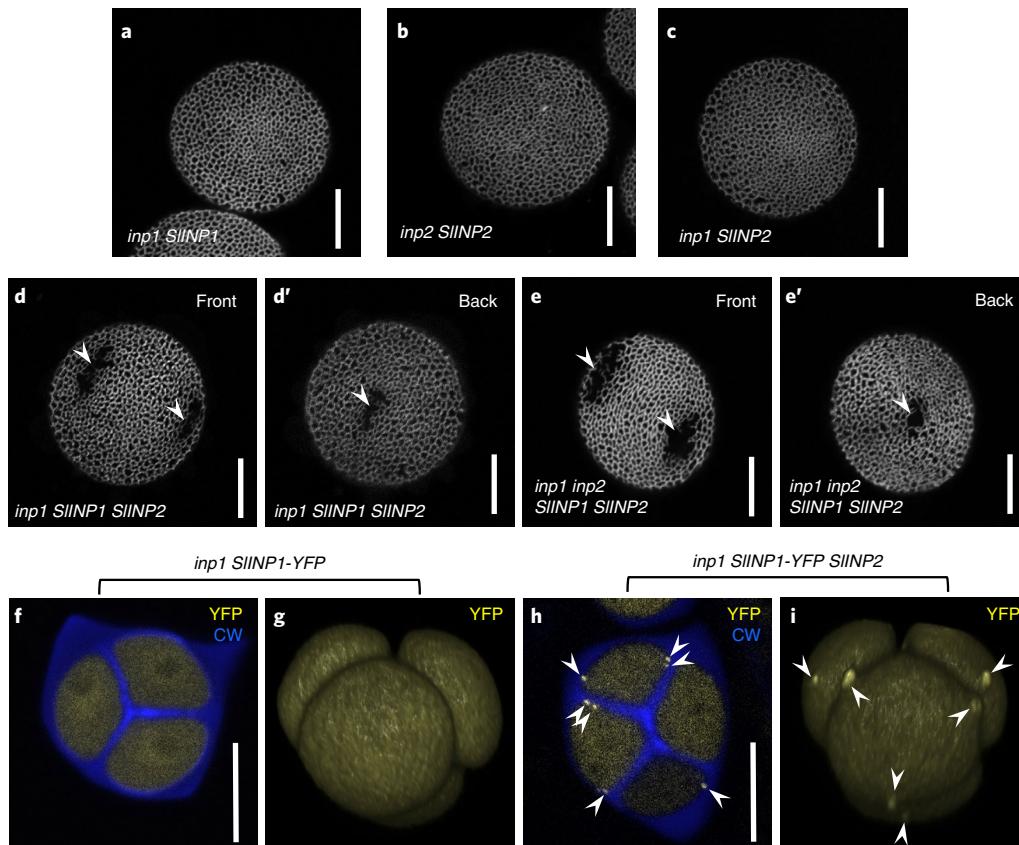


Fig. 7 | Tomato orthologues of INP1 and INP2 fail to function in *Arabidopsis* when expressed individually but gain this ability when co-expressed.

a–c, Neither SIINP1 (**a**) nor SIINP2 (**b,c**) are able to restore apertures in *Arabidopsis* pollen when expressed on their own. More than ten T₁ plants (≥ 50 pollen grains per plant) were analysed, with similar results. **d–e'**, When both SIINP1 and SIINP2 are expressed in *Arabidopsis*, they restore short to medium apertures (arrowheads) in the *inp1* (**d,d'**) and *inp1 inp2* (**e,e'**) *Arabidopsis* mutants. Confocal images of pollen grains stained with auramine O. Both front and back views are shown for the same pollen grains in (**d,d'**) and (**e,e'**) to demonstrate positions of apertures. Experiments were repeated twice, with similar results (~90% of plants had short- to medium-size apertures and the rest had no apertures). **f–i**, SIINP1-YFP localizes to the aperture domains in the presence of SIINP2 (**h,i**) but not when expressed on its own (**f,g**). Confocal optical sections (**f,h**) and three-dimensional reconstructions of tetrads of microspores (**g,i**). YFP signal is shown in yellow and callose wall (CW, stained by calcofluor white) is shown in blue. Arrowheads point to the YFP signal at the aperture domains. Experiments were repeated twice, with similar results. Scale bars, 10 μ m.

early as green algae, we found distinct, well-supported INP1 and INP2 protein lineages only in gymnosperms and angiosperms. In angiosperms, they have greatly diversified and display similar trends of evolutionary divergence, resulting in phylogenetic trees of similar topology (Fig. 5). Orthologues of both INP1 and INP2 exist in various families of rosids, asterids, basal eudicots, monocots and magnoliids. The INP1 sequence is also present among the transcripts from two ANA-grade basal angiosperms, *Austrobaileya* and *Nymphaea*. Failure to find an INP2 homologue in *Austrobaileya*, despite finding one in *Nymphaea*, could be due to the incompleteness of the database. Interestingly, both INP1 and INP2 are absent from the genome of *Amborella*, another basal angiosperm whose genome was published several years ago¹⁹.

Degrees of sequence divergence within the INP1 and INP2 angiosperm lineages are generally consistent with the phylogenetic relationships between species (Supplementary Table 1 and Fig. 5). Both *Arabidopsis* INP1 and INP2 (AtINP1 and AtINP2) share between ~95 and ~70% protein sequence identity with their respective orthologues from closely related species in the Brassicaceae and Cleomaceae families. Sequence identity with orthologues from more distantly related eudicots drops to ~40–50%. In monocots, the similarities to AtINP1 and AtINP2 are further reduced: proteins from Arecales and Bromeliaceae families (for example, palms

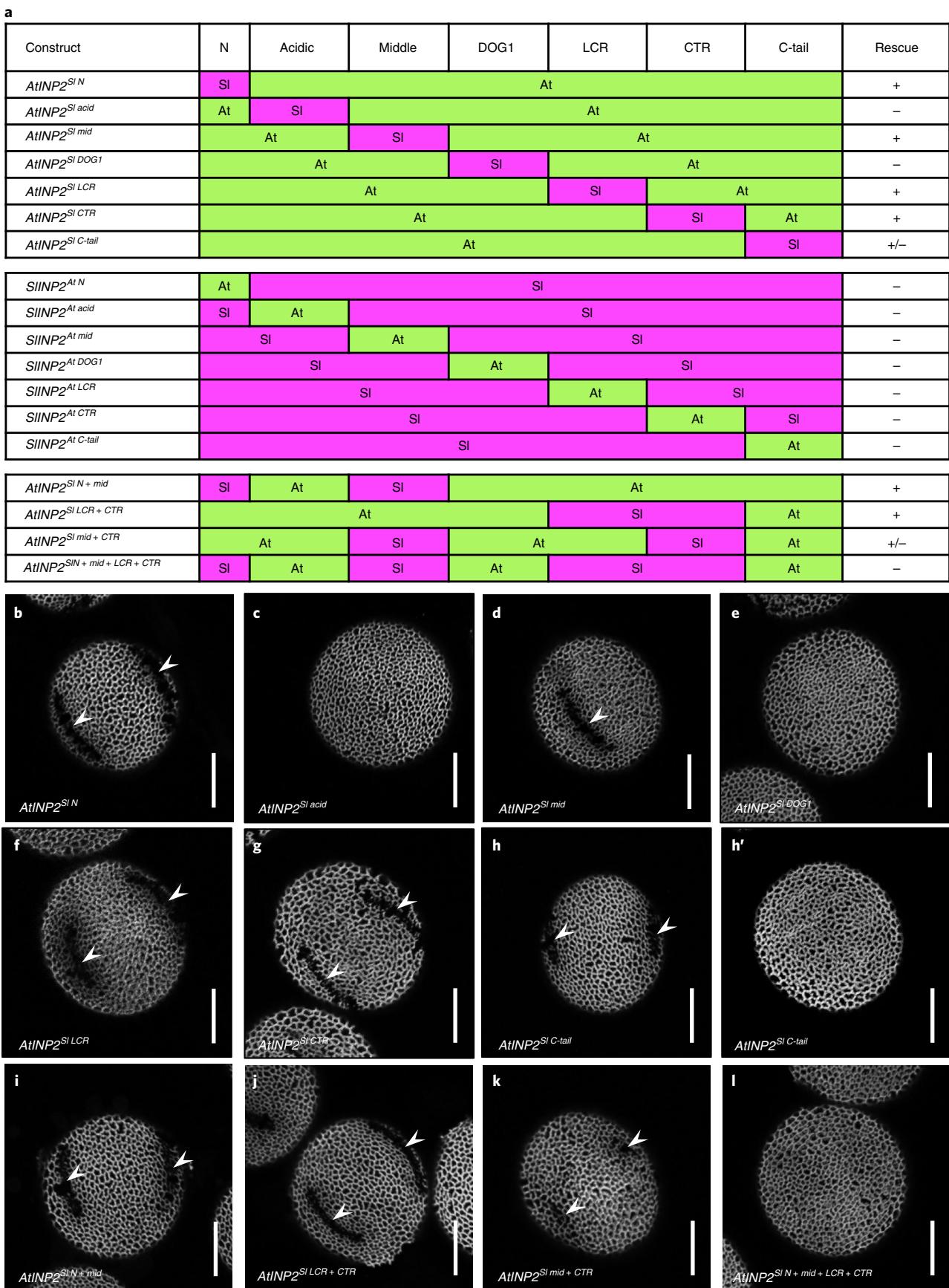
and pineapple) exhibit ~45 to ~30% sequence identity with AtINP1 and AtINP2. In both INP1 and INP2 lineages, particularly distinct clades are formed by proteins from grasses (Poaceae) (Fig. 5 and Supplementary Table 1): within each INP group, these proteins diverged greatly from the rest of their lineages (showing ~35 and ~20–25% identity, respectively, to AtINP1 and AtINP2) but retained >80% identity to their orthologues from other species of Poaceae despite the long evolutionary history of this monocot family²⁰.

INP1 and INP2 are functionally species-specific. The similar evolutionary trends displayed by INP1 and INP2, as well as the ability of these proteins to interact, led us to suggest that INP2 might serve as a species-specific partner for INP1. We tested this idea using the orthologues of INP1 and INP2 from tomato *Solanum lycopersicum* (SIINP1 and SIINP2) which both share ~45% amino acid identity with their *Arabidopsis* counterparts. Using the Y2H and split-luciferase assays, we confirmed the ability of SIINP1 and SIINP2 to interact (Fig. 6a,b). Furthermore, in both assays, the tomato INP proteins specifically interacted with each other and not with the *Arabidopsis* proteins (Fig. 6a,c).

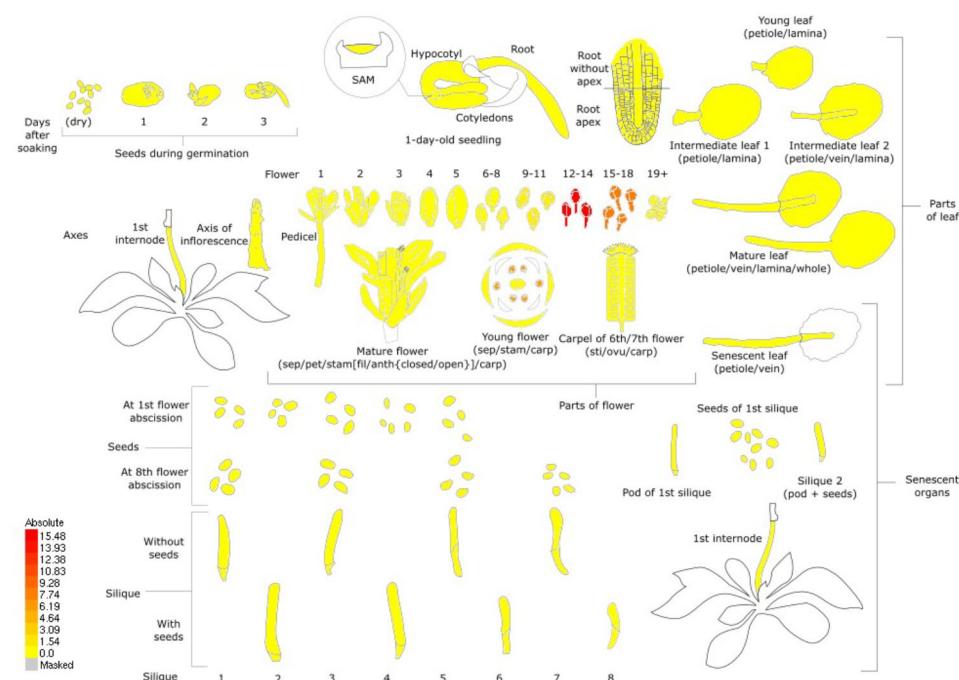
We demonstrated previously that SIINP1 was unable to localize to aperture domains and restore apertures when expressed in the *Arabidopsis* *inp1* mutant¹² (Fig. 7a). Here, we placed SIINP2

they are co-expressed but not when expressed individually (Fig. 7). We have also demonstrated that SIIINP1, which, on its own, does not assemble at the aperture domains in *Arabidopsis* tetrads¹², gains

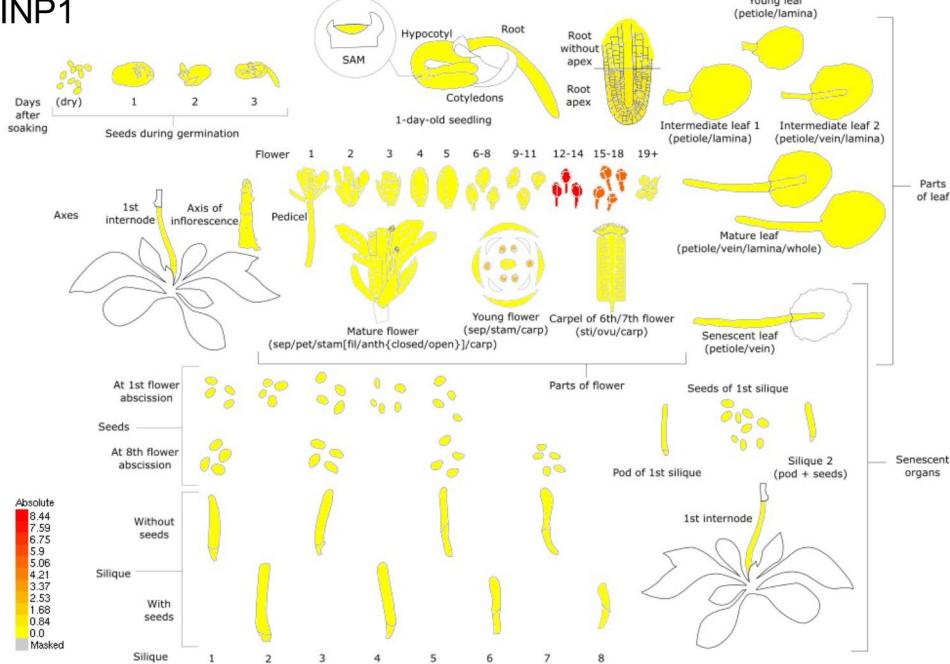
the ability to do this in the presence of SIIINP2 (Fig. 7h,i). Our data show that several regions of INP2, including the DOG1 domain, contribute to its species specificity (Fig. 8).



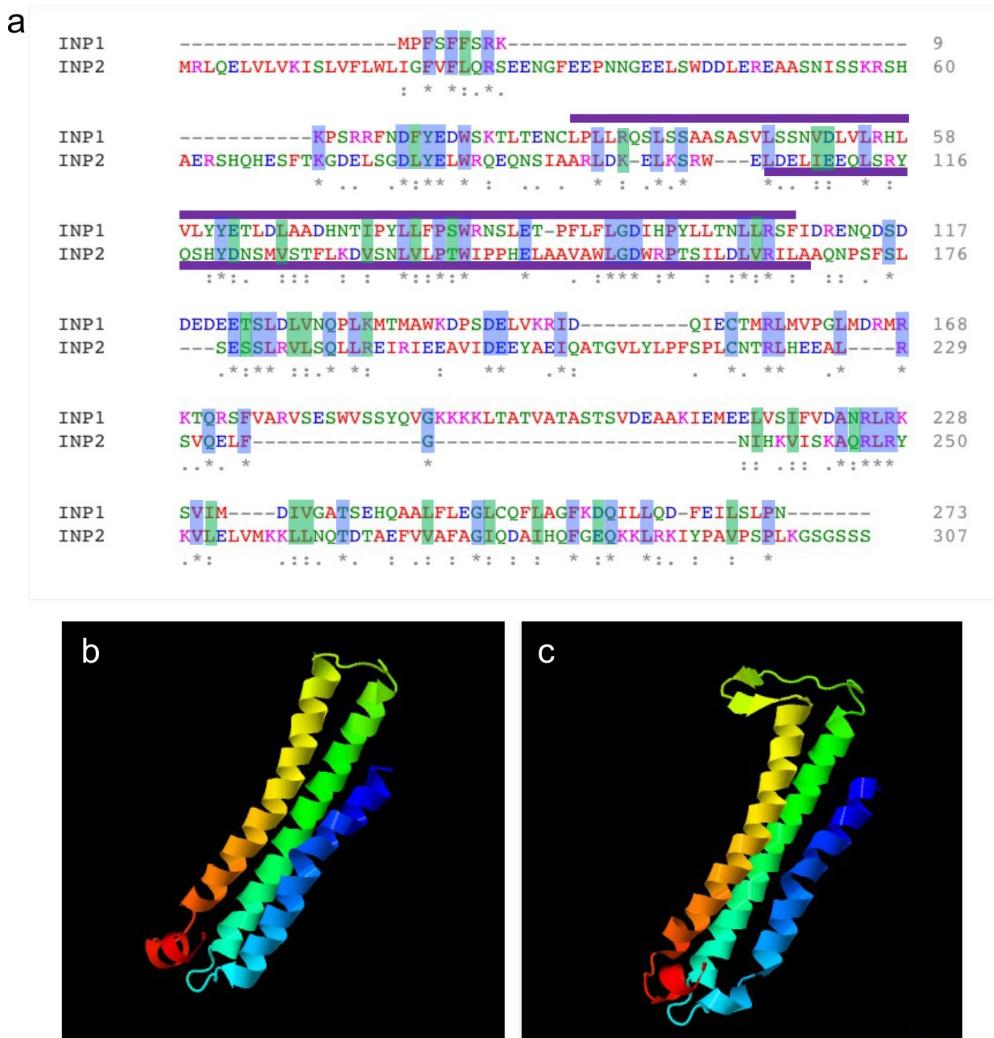
a
INP2



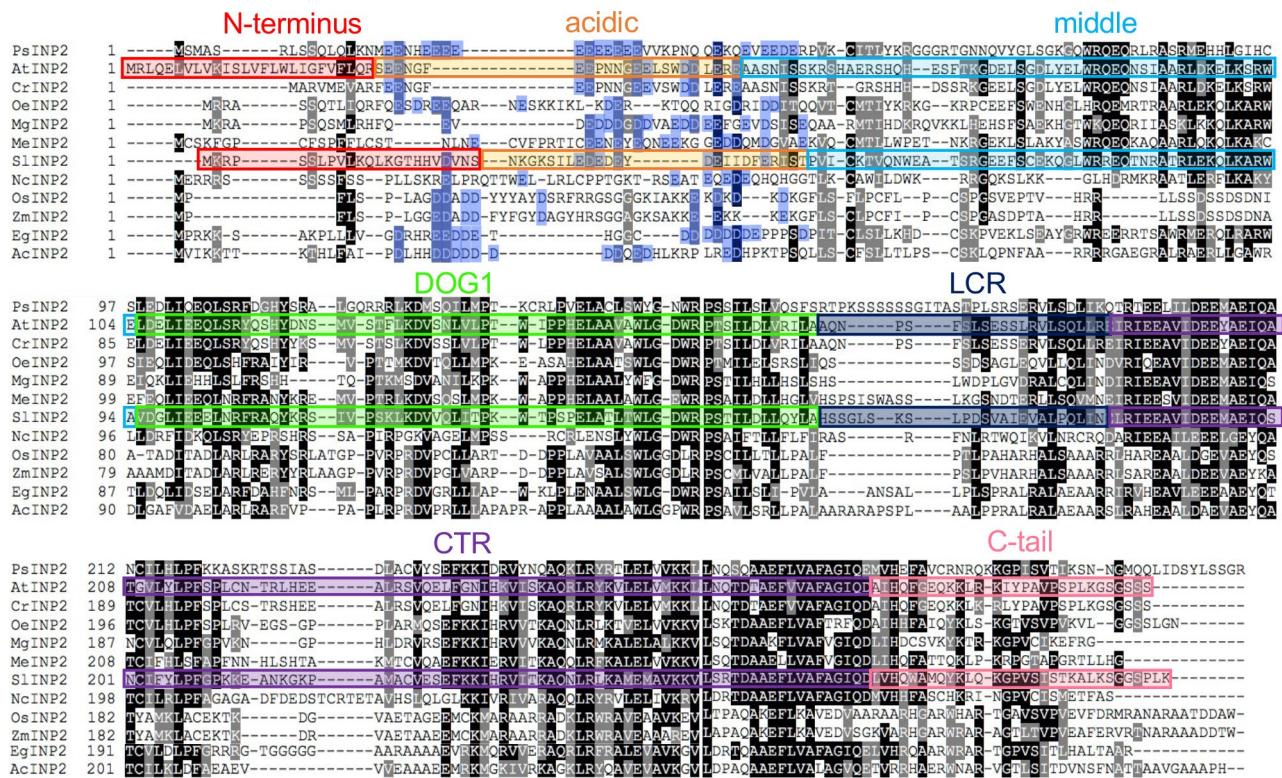
b
INP1



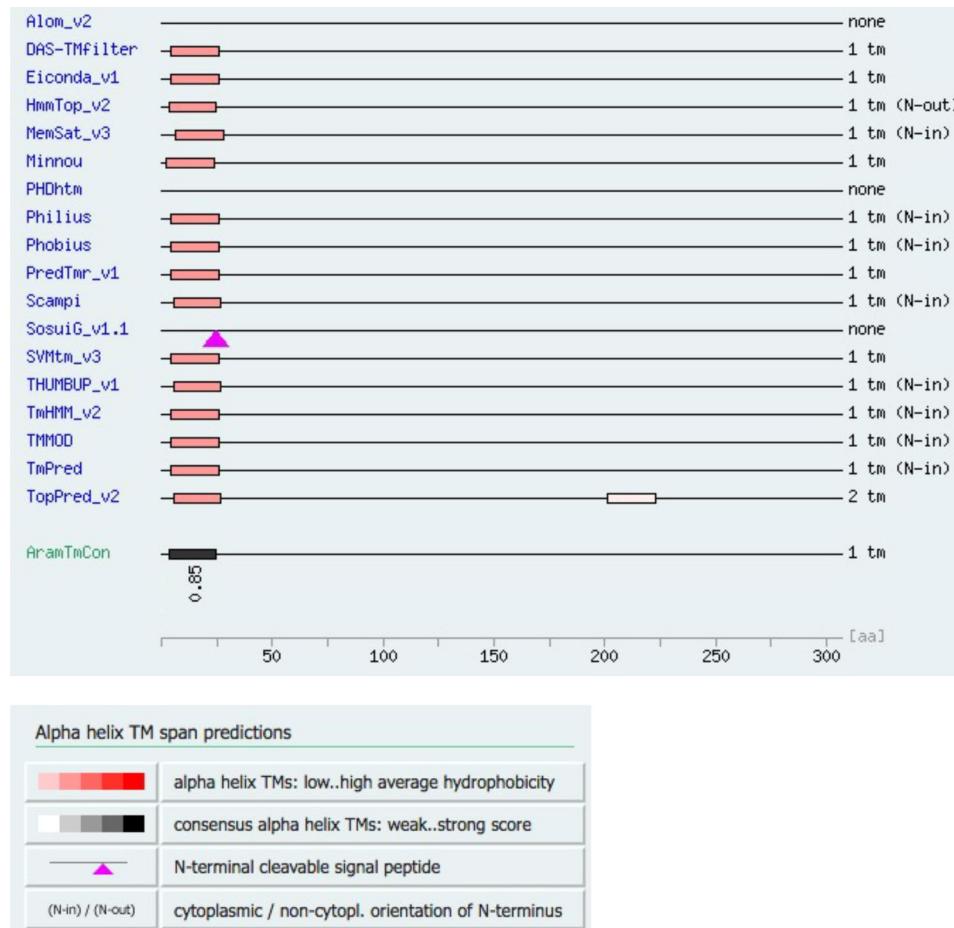
Extended Data Fig. 1 | *INP2* and *INP1* display similar expression patterns, with both genes showing highest expression in young developing buds. The RNA-seq data for *INP2* (a) and *INP1* (b) are from the dataset of Klepikova et al.¹⁵ and visualized with the BAR eFP Browser.



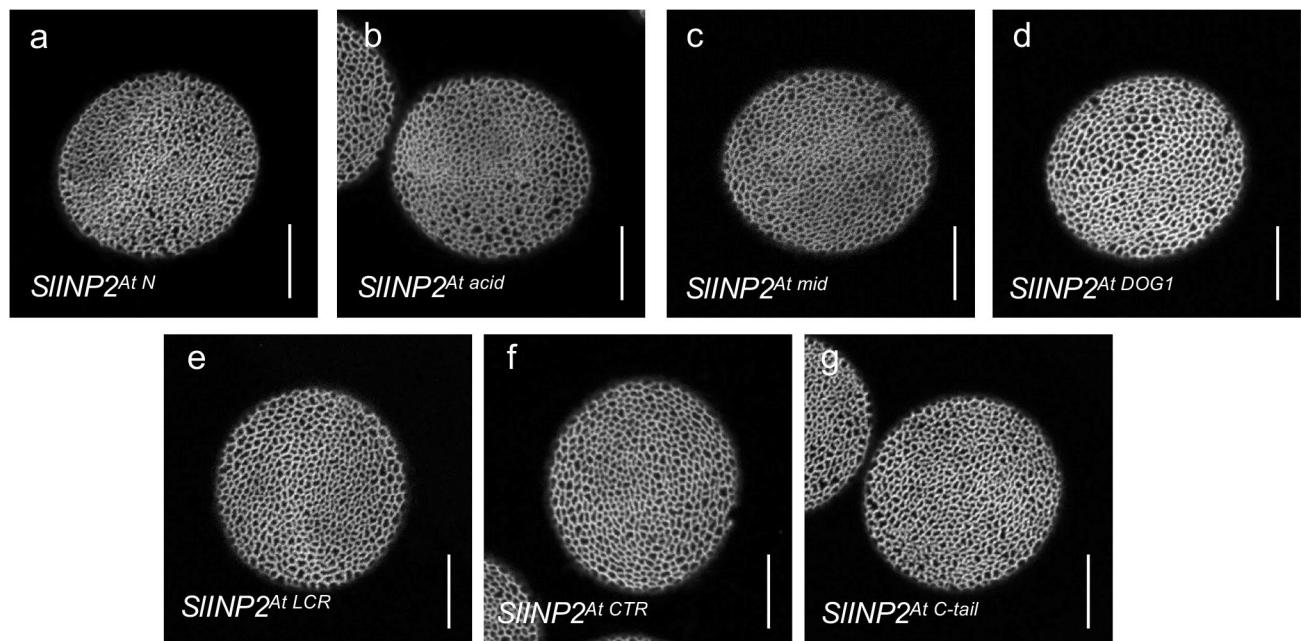
Extended Data Fig. 2 | INP1 and INP2 proteins both contain the DOG1 domain and have similar structural organization predicted for their C-terminal parts. **a**, Protein alignment between INP1 and INP2 proteins. Identical and similar (V/I/L, D/E, K/R, N/Q and S/T) residues are shaded, respectively, in blue and green. The positions of the DOG1 domains predicted by Pfam are indicated by purple lines. **b-c**, Protein structures predicted by Phyre2 for C-terminal parts of INP1 (**b**) and INP2 (**c**) (confidence: >97% for both proteins). In both cases, the modelled regions cover 114 amino acids, which constitute, respectively, 42% of INP1 and 37% of INP2. The same template (c4clvB, nickel-cobalt-cadmium resistance protein NccX from *Cupriavidus metallidurans* 31a) was selected by the program in both cases.



Extended Data Fig. 3 | Alignment of INP2 proteins from representatives of different angiosperm taxa. The following species were used (from top to bottom): *Papaver somniferum* (basal eudicots, Papaveraceae), *Arabidopsis thaliana* (rosids, Brassicaceae), *Capsella rubella* (rosids, Brassicaceae), *Olea europaea* (asterids, Oleaceae), *Mimulus guttatus* (asterids, Phrymaceae), *Manihot esculenta* (rosids, Euphorbiaceae), *Solanum lycopersicum* (asterids, Solanaceae), *Nymphaea colorata* (basal angiosperms, ANA, Nymphaeaceae), *Oryza sativa* (monocots, Poaceae), *Zea mays* (monocots, Poaceae), *Elaeis guineensis* (monocots, Arecaceae), *Ananas comosus* (monocots, Bromeliaceae). The seven regions selected for creating AtINP2/SIINP2 chimeras are indicated by differently coloured rectangles. Aspartate (D) and glutamate (E) residues in the acidic region are shaded in blue. Black shading indicates identical amino acids and grey shading indicates similar amino acids present at the same position in at least half of the aligned proteins.



Extended Data Fig. 4 | *Arabidopsis* INP2 likely contains a transmembrane domain at its N terminus. Multiple TM discovery algorithms predict existence of the transmembrane domain at the N terminus of INP2 from *Arabidopsis thaliana* (AtINP2), with the consensus score of 0.85 generated by the plant membrane protein database Aramemnon (AramTMCon).

**Extended Data Fig. 5 | None of the seven AtINP2 regions is sufficient on its own to convert SII NP2 into a protein able to function in Arabidopsis.**

Confocal images of pollen grains produced by the transgenic *inp2* plants expressing seven versions of chimeric *SII NP2* constructs in which one region at a time was replaced with the corresponding regions from *AtINP2*. At least 10 independent T_1 lines were tested for each construct (≥ 50 pollen grains per line), with similar results. Scale bars = 10 μm .

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- | | |
|-----------------|---|
| Data collection | For bioinformatics analysis, BLAST search was conducted using NCBI website (http://www.ncbi.nlm.nih.gov/), Phytozome v. 12.1 (https://phytozome.jgi.doe.gov/pz/portal.html), the 1,000 Plants project (OneKP-China National Gene Bank (https://db.cngb.org/onekp/)24 and PLAZA (https://bioinformatics.psb.ugent.be/plaza/). For confocal imaging, NIS Elements v.4.20 software (Nikon) was used. |
| Data analysis | 3D reconstruction of tetrads was done using NIS Elements v.4.20 software (Nikon). For multiple sequences alignment, MAFFT v7.017 (L-INS-i algorithm) was used (https://www.ebi.ac.uk/Tools/msa/mafft/). The IQ-TREE program v. 1.6.12, (http://www.iqtrees.org/) was used to construct phylogenetic trees. The phylogenetic trees were visualized in iTOL v. 5 (https://itol.embl.de/). Split-luciferase assay analysis was done with ImageJ v. 1.53a. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Public RNA-seq data for gene expression analysis ,were obtained from the TRAVA database (travadb.org).

Protein sequences were obtained from TAIR (<https://www.arabidopsis.org/>), NCBI (<http://www.ncbi.nlm.nih.gov/>), Phytozome v. 12.1 (<https://phytozome.jgi.doe.gov/pz/portal.html>), the 1,000 Plants project (OneKP-China National Gene Bank (<https://db.cngb.org/onekp/>)24 and PLAZA (<https://bioinformatics.psb.ugent.be/plaza/>))

bioinformatics.psb.ugent.be/plaza/.

Markers for mapping were generated using the 1,001 Genomes Project database (<http://signal.salk.edu/atg1001/index.php>)²³ and the Arabidopsis Mapping Platform (<http://amp.genomics.org.cn/>). Fig. 4C contains associated Source data.

All data supporting the findings of this study are available within the article, Supplementary Information files or from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were consistent with those in our and others' previously published similar studies (e.g. Lee et al. (2018) Plant Cell 30:2038-2056; Zhang et al. (2020) Nat. Plants 6, 394–403). All experiments were performed using at least three independent biological replicates. For transgenic lines, a minimum of eight independent lines were created and analyzed. Sample sizes and the number of times an experiment was performed are described in figure legends and main text.
Data exclusions	No data were excluded.
Replication	At least three independent biological replicates were used for each experiment, with similar results. For transgenic lines, a minimum of eight independent lines were created and analyzed, with similar results.
Randomization	Plants of different genotypes were used as study groups. When more plants were available than the required sample size, plants were chosen randomly for analysis.
Blinding	Experiments were not blinded. Data were collected according to the genotypes of samples. Since comparisons were generally qualitative, blinding was not relevant.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|--|
| n/a | Involved in the study |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Antibodies

Antibodies used

For immunoprecipitation: anti-GFP antibody (monoclonal (clone 3E6, lot 1711553), mouse, Molecular Probes by ThermoFisher Scientific, #A-11120). To detect proteins on Westerns (all antibodies were used at 1:2000 dilution): anti-GFP antibody (polyclonal, rabbit, Abcam; ab6556) anti-HA antibody (monoclonal (clone 3F10), rat, Sigma, 11867423001) anti-rabbit IgG peroxidase-conjugated antibodies (SeraCare/KPL; 5220-0283/04-15-06) anti-rat IgG peroxidase-conjugated antibodies (SeraCare/KPL; 5220-0364/04-16-06)

Validation

Anti-GFP monoclonal: <https://www.thermofisher.com/antibody/product/GFP-Antibody-clone-3E6-Monoclonal/A-11120> and publications listed thereof.
 Anti-GFP polyclonal: <https://www.abcam.com/gfp-antibody-ab6556.html> and publications listed thereof.
 Anti-HA: <https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Roche/Bulletin/1/roahahabul.pdf>

Anti-rabbit IgG peroxidase-conjugated: <https://www.seracare.com/AntiRabbit-IgG-HL-Antibody-PeroxidaseLabeled-5220-0283/>
Anti-rat IgG peroxidase-conjugated: <https://www.seracare.com/search/?q=5220-0364&searchType=product>