Convergent gene loss in aquatic plants predicts new components of plant immunity and drought response

Baggs EL [1,2], Thanki AS[1], O'Grady R[3], Schudoma C[1], Haerty W[1], Krasileva KV [1,2,3*]

1 Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, United Kingdom

2 University of California Berkeley, Berkeley, California, 94720, United States of America

3 The Sainsbury Laboratory, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

Corresponding author: kseniak@berkeley.edu

Abstract

The transition of plants from sea to land sparked an arms race with pathogens. The increased susceptibility of land plants is largely thought to be due to their dependence on micro-organisms for nutrients; the ensuing co-evolution has shaped the plant immune system. By profiling the immune receptors across flowering plants, we identified species with low numbers of NLR immune receptors. Interestingly, four of these species represent distinct lineages of monocots and dicots that returned to the aquatic lifestyle. Both aquatic monocot and dicot species lost the same well-known downstream immune signalling complex (EDS1-PAD4). This observation inspired us to look for other genes with a similar loss pattern and allowed us to predict putative new components of plant immunity. Gene expression analyses confirmed that a group of these genes was differentially expressed under pathogen infection. Excitingly, another subset of these genes was differentially expressed upon drought. Collectively, our study reveals the minimal plant immune system required for life under water, and highlights additional components required for the life of land plants.

Author summary

Plant resistance to pathogens is commonly mediated by a complex gene family, known as NLRs. Upon pathogen infection, changes in the cellular environment trigger NLR activation and subsequent defence responses. Despite the dependence of agricultural practices on NLR genes to control pathogen load, relatively little is known about this gene family outside of model crop species. In this study, we identified a convergent reduction in the NLR gene family among two lineages of aquatic plants. Furthermore, we established that NLR reduction occurred in conjunction with the loss of a common immune signalling pathway. Subsequently, we identified other genes convergently lost in aquatic species and propose these as candidate components of the plant immune signalling pathway. In addition, we revealed components of the agronomically important drought response to be lost in aquatic plants. This study adds to our understanding of the complex interactions between environment and response to biotic stress, widely known as the disease triangle. The pathways identified in this study shed further light on the link between responses to drought and disease.

Introduction

Plants evolved from a common ancestor with charophyte green algae upon a major change in lifestyle -- the transition from water to land -- over 450 million years ago (MYA) (1,2). Extant plant lineages, such as bryophytes (mosses, liverworts and hornworts), terrestrial non-vascular plants (ferns), gymnosperms (pine, conifers) and angiosperms (monocots, dicots) diverged from the ancestor of terrestrial plants over 300 MYA (3). Modern agriculture depends heavily on two prominent branches of angiosperms, monocots and dicots. Monocots range from cereal crops to tropical palms, whilst the most widely consumed dicots include soybean, potato and tomato (4). Crops are continuously exposed to both biotic and abiotic stresses, which can result in major yield loss. Such losses are expected to increase in frequency with climate change (5). Adaptation of crops to new and evolving pathogen threats often utilises introgressing *Nucleotide Binding Leucine Rich Repeat* (*NLR*) genes, which are involved in pathogen recognition.

The first evidence of pathogenic infection of land plants comes from fossils dating back 400 million years (6), prior to the divergence of angiosperms from mosses and gymnosperms (7). Similarly, major components of the immune system, such as *NLR* genes, were prevalent in plants early upon transition to land (8). Plant pathogen interactions therefore predate the divergence of angiosperms and for 200 million years angiosperms relied on common pathways to defend against pathogens. Since the divergence of monocots and dicots 120-180 MYA lineage specific response to stress including pathogen infection have emerged (9).

PAMP-triggered immunity (PTI) is the initial plant disease resistance response to detection of a pathogen through monitoring the extracellular environment, however pathogen effectors can suppress PTI to facilitate virulence. Hence, a second intracellular monitoring system of effector

triggered immunity (ETI) is essential for resistance to many pathogens. Plant NLR immune receptors convey ETI, upon detection of intracellular pathogen molecules. The NLR proteins are typically composed of three or more domains. A central Nucleotide Binding (NB-ARC) domain is a component of all NLRs due to it being essential for receptor activation, similar to the NAIP, CIITA, HET-E, and TP1 (NACHT) domains found in animal NLR immune receptors (10). The NB-ARC domain is commonly followed by a series of Leucine Rich Repeats (LRRs), previously shown, in some but not all instances of direct recognition, to mediate pathogen-derived effector binding (11-13). Toll-like, Interleukin-1 (TIR-1) or coiled-coil (CC) domain, typically found at the N-terminus, function in the initiation of the signalling cascade (14). NLRs containing TIR-1 are referred to as TIR-1 NLRs, while CNL refers to NLRs with a CC domain. Within the CNL class, there is a sub-clade characterized by CC_{RPW8} , which have been shown genetically to be required for signalling by other NLRs (15-17).

A subset of NLRs have undergone functional specialisation to serve downstream of other NLRs in the signalling cascade (18-21). This mechanism is referred to as an NLR sensor-helper pair, where genomically the two NLR genes are often adjacent in a head to head orientation with a shared promoter (22,23) . The sensor NLR is required to recognise changes upon infection, whilst the helper NLR is activated upon a change in its paired NLR rather than via pathogen induced change (18). The helper NLR is then able to activate the signalling cascade for cell death. Additionally, sensor and helper interaction can be in the form of a network whereby multiple sensors interact with the same helper (24). Examples of network helpers include: *NLR REQUIRED FOR CELL DEATH (NRC)* clade (NRC2/3/4) (25) and CC_{RPW8} containing; *N-REQUIRED GENE 1 (NRG1) (16,17)*, *ACTIVATED DISEASE RESISTANCE (ADR1) (26)* of *NLR*s. The *NLR* gene family is complex, with copy numbers variable by 10 fold between species of the same family (27). NLR copy numbers in crops range from as low as 30 in cucumbers to

as high as nearly 3,000 in hexaploid wheat (27,28). The variation in copy number of *NLRs* is becoming more apparent with the increasing number of available genomes, however due to the historical bias in genome sequencing favouring economically important or model species most of our information comes from the *Brassicaceae*, *Solanaceae* and *Poaceae* families (29).

While there is abundant literature on pathogen sensing and signalling by NLRs (30), the next stages of the immune signalling cascade are still unclear. Among dicots, typically over half the NLRs contain a TIR-1 domain (29,31). In contrast, no TIR-1 domain containing NLR receptors have yet been identified in monocots. The two key signalling components that are downstream of TIR-1 NLRs, ENHANCED DISEASE SUSCEPTIBILITY 1 (EDS1) and PHYTOALEXIN DEFICIENT 4 (PAD4), are present in almost all angiosperms despite the absence of TIR-1 NLRs in monocots (32). Wheat EDS1 has retained functional importance in immunity despite the absence of TIR-1 NLRs, with EDS1 overexpression contributing to increased powdery mildew resistance (33). Furthermore, EDS1 in dicots forms a complex with PAD4 and also SENESCENCE ASSOCIATED GENE 101 (SAG101) (32). *In planta* work in *Arabidopsis* showed EDS1 binding directly to PAD4 and SAG101 to form mutually exclusive heterodimeric complexes (32), with subcellular localization of EDS1 dependent on the interacting partner (34). PAD4 and SAG101 also interact with MYC2/ JASMONATE INSENSITIVE-1, a transcriptional regulator of JA signalling (35). Interestingly, while *EDS1* and *PAD4* are conserved, the *SAG101* gene is absent from available genomes of grasses (32). Regardless of over 120-180 Mya of divergent evolution, some NLRs such as barley *MILDEW LOCUS A (MLA)* are functional when transformed into *Arabidopsis* (36). This indicates overall conservation of downstream immune pathways across monocots and dicots.

Within flowering plants, there is a huge amount of diversity that can be mined to understand plant molecular pathways. Specifically, looking at convergent evolution we can predict gains and

losses of genetic pathways associated with particular lifestyles. The switch between terrestrial and aquatic environments has occurred numerous times in both monocot and dicot lineages, although the exact points of transition remain unknown due to the lack of fossil evidence and uncertainty in phylogenetic assignment (37). Aquatic plant species genome assemblies are available for the monocot *Araceae* (*Spirodela polyrhiza, Lemna minor*) and *Zosteraceae* (*Zostera marina*) families as well as the dicot families *Lamiales* (*Utricularia gibba, Genlisea aurea*) and *Nelumbonaceae* (*Nelumbo nucifera*). Interestingly, aquatic species from both monocot and dicot lineages have reduced genome size compared to terrestrial sister species (38-41). The availability of such genomic resources allows elucidation of pathways that have independently evolved in aquatic species to facilitate acclimation to their new environment.

In this study, we used comparative genomics to identify gene families, which were convergently lost in the transition of monocot and dicot species to an aquatic environment (Fig. 1). We identified independent contractions in NLR numbers in aquatic monocots and dicots. We further analysed the known components of disease resistance pathways to identify connections between the loss of NLR genes and the loss of additional signalling pathways. Finally, we used gene family clustering methods to identify other pathways convergently lost in monocot and dicot aquatic species as well as to predict novel candidates involved in plant innate immunity and drought response.

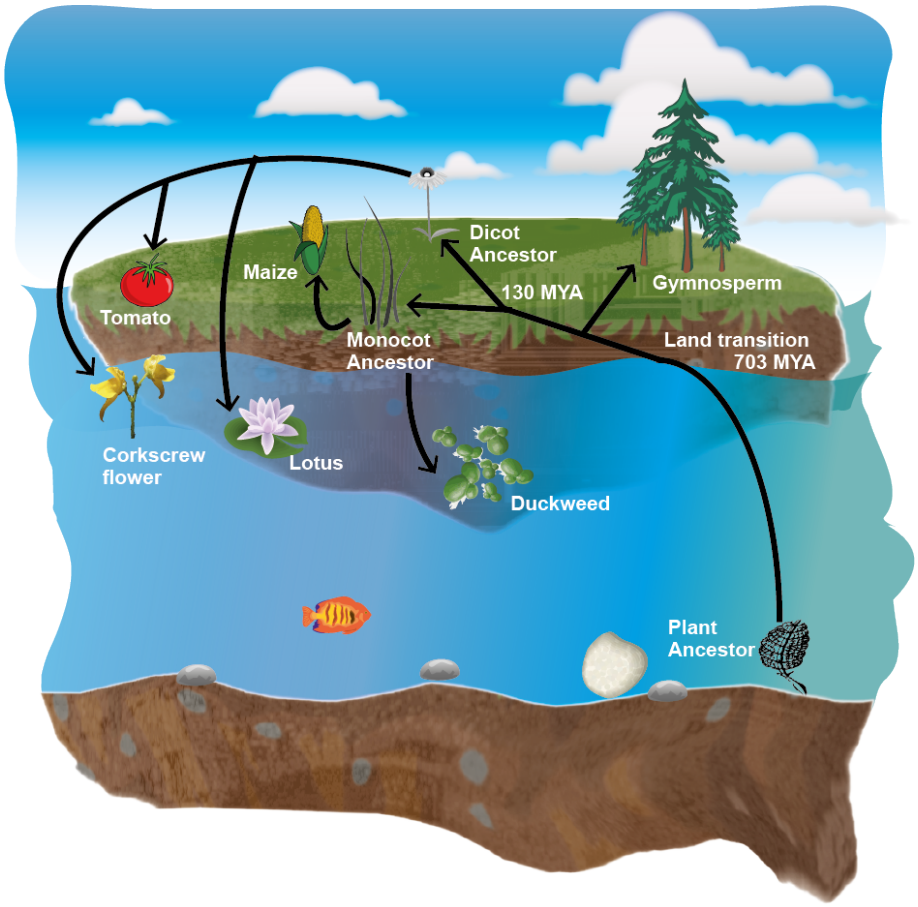


Figure 1 - **Stylized depiction of plants moving on land and back to the aquatic environment.** From right bottom: charophyte algae has the most recent common ancestor with plants before plants transition to land, the sequential divergence of gymnosperms, monocots and dicots. After divergence of monocots and dicots there are independent transitions of monocots and dicots back to the aquatic environment.

Results

Aquatic plants have lost most of the NLR plant immune receptors

We annotated and characterized NLRs across 18 genomes that represent taxonomic and ecological diversity across monocots and dicots including plants that have changed their lifestyle from terrestrial to aquatic in both clades (Fig. 2A). Interestingly, we identified multiple

independent reductions in the number of NLRs in eelgrass (*Zostera marina*), duckweed (*Spirodela polyrhiza*), orchid (*Phalaenopsis equestris)*, resurrection grass (*Oropetium thomaeum*), humped bladderwort (*Utricularia gibba)* and corkscrew plant (*Genlisea aurea*). Compared to the average 300-400 NLRs in monocots, the genomes of duckweed and eelgrass encode only 62 and 10 NLRs (Fig. 2B). Likewise, the aquatic dicot *G. aurea* had only 8 NLRs while humped bladderwort had no NLRs that retained all motifs associated with this protein family (Fig. 2B). However, not all aquatic species show NLR reduction. An example for this is chinese lotus (*Nelumbo nucifera*), whose genome encodes 237 NLRs.

To test whether the low number of NLRs found in each species could be attributed to poor assembly or annotation, we assessed presence of single-copy orthologs using the Benchmarking Universal Single-Copy Orthologs (BUSCO) pipeline (42) on the 19 species used in the study (S1 Table). The number of complete single-copy orthologs for duckweed and eelgrass was comparable to other monocots including; orchid, *A. hypochondriachus* and pineapple. Humped bladderwort and corkscrew flower have a lower number of complete single-copy orthologs than other species. This can be explained as genes from pathways known to be lost in these two aquatic species are present in the BUSCO land plant database. For example, the stomatal development genes (EOG0936017K (*AT5G03730*) and EOG09360A7W (*AT1G80080*)) are known to be lost in eelgrass which lacks stomata (41), however these genes are part of the BUSCO database (S2 Table). Presence of these genes in the BUSCO database reduces the number of complete BUSCOs in these aquatic species but due to absence of a conserved pathway rather than poor annotation. Subsequently, we investigated if genome size reduction could explain the NLR reduction in aquatic species. The 99Mb genome of humped bladderwort (38) and 63.5Mb genome of corkscrew plant (43) are two of the smallest among all plant species. Duckweed and eelgrass have genomes of 202Mb (40) and 142Mb (41) in size respectively and thus have a larger genome size than the *A. thaliana* 135Mb genome (44) which

has a higher number of NLRs in both TIR-1 NLR and CNL clades. While NLRs constitute 0.6% of Arabidopsis protein encoding genes, among aquatic species the percentage of NLR protein coding genes is between 0.003- 0.43% (S3 Table). The low percentage of NLR proteins in the total proteomes thus are not be explained by genome reduction alone.

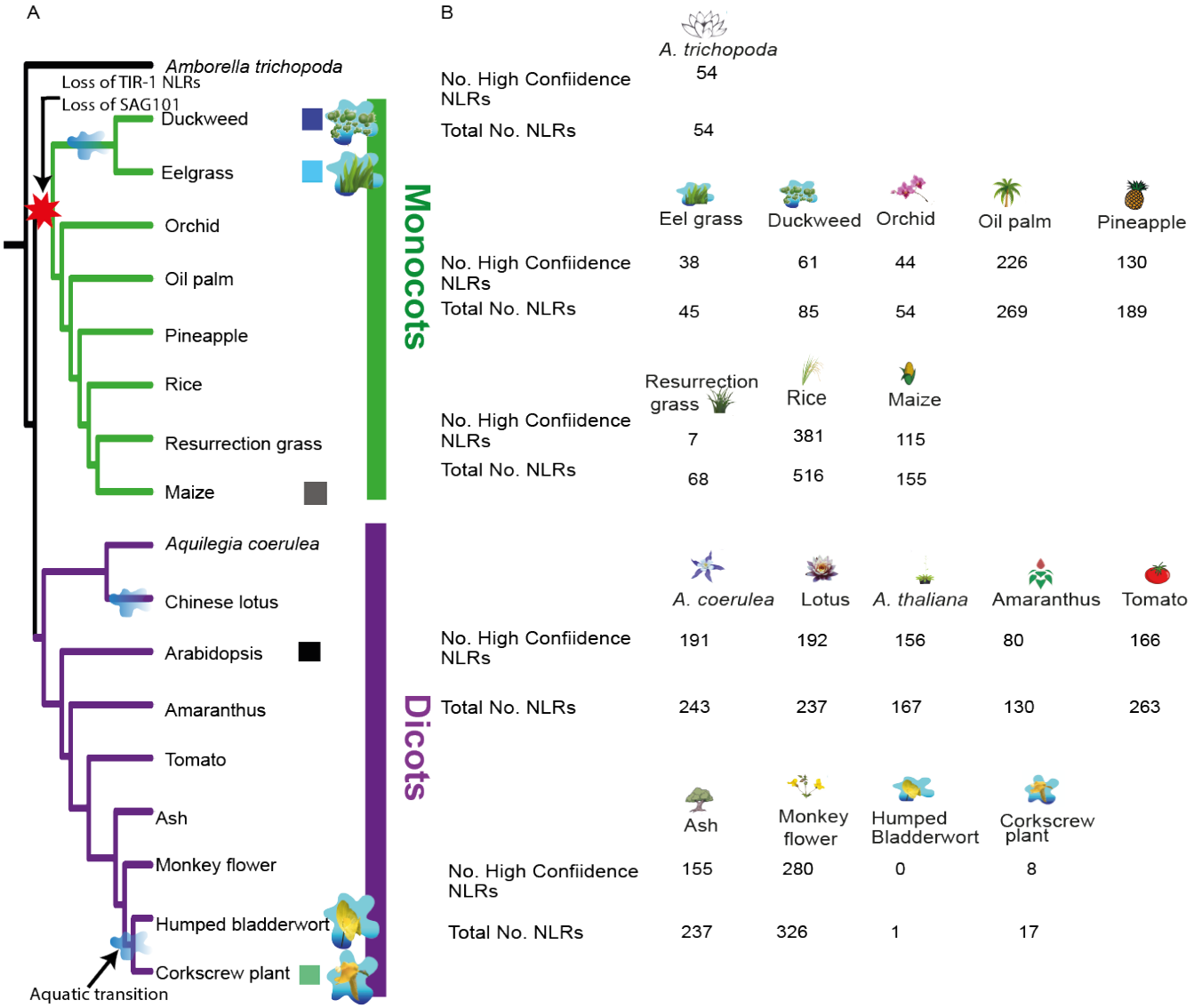


Figure 2 – **Phylogenetic relationship, lifestyles and NLR repertoires of the plant species used in this study** (A) A species tree of monocot and dicot genomes of interest. A red star indicates loss of *SAG101* and *TIR1-NLRs*, blue splash represents estimated time of transition of common ancestor from land to water. Squares refer to branch colours used to designate that species in the phylogenetic tree in Fig 3. Caricatures provide a reference of species considered aquatic (plant surrounded by splash) and are used for subsequent analysis. (B) Number of high confidence NLRs annotated in each species and total number of NLRs identified by PfamScan and plant_rgenes pipeline.

We further tested which NLR clades were lost and which retained among the four low NLR aquatic species (Fig. 3). In terrestrial plants, TIR-1 NLRs have been lost in monocots soon after the divergence from dicots (29,45,46). Interestingly, TIR-1 NLRs, which are a major component of the dicot immune system (Fig. 3), are absent from aquatic dicot species. The RPW8 like coiled-coil clade C1 found in all terrestrial monocots and dicots, is absent in all four of the aquatic species (Fig. 3). This is pertinent as CC_{RPW8} NLR NRG1 is required for TIR-1 NLR signalling (16,17). The absence of this clade in the dicot aquatic species supports its function as a helper for TIR-1 NLR signalling. Overall, 13 NLR clades were lost in multiple aquatic species and only two clades (C2 and C3) are retained in more than one aquatic species. The C2 clade includes the *Arabidopsis NLR* gene called *Resistance to Pseudomonas Syringae 2 (RPS2),* whose protein product acts as a guard detecting pathogen effector triggered modification to RPM1-INTERACTING PROTEIN 4 (RIN4) (47,48). The C3 clade includes Arabidopsis *LRRAC1 (49).* The C3 clade is characterised by a large expansion of monocot NLRs including 7 NLRs with known defence functions (S4 Table). Two distinct clades (C6 and C10) are present only in one of the aquatic species and have recently undergone expansions. Corkscrew flower has retained a single clade of NLRs, whilst eelgrass and duckweed have retained NLR members in 3 clades, both retaining C2 and C3 membership in addition to C4 and C10, respectively. Duckweed additionally has 4 NLRs which did not fall into any defined clades based on our criteria. Phylogenetic analysis not only highlighted the loss of NLRs among aquatic species, but also demonstrated recent expansion of some of the clades implying a much greater initial reduction of the NLR repertoire than what is currently observed. The predominant example of NLR clade expansion is in the C2 NLR family which contains a sub-group of 16 duckweed NLRs (0.5 branch length, 100 bootstrap).

To test if aquatic plants underwent a general reduction in large protein families, we annotated receptors belonging to the Receptor-like kinase family (RLKs) and actin proteins. We focused

on RLKs as they represent another family of immune receptors conserved across plants and would allow us to observe if PTI was affected in addition to ETI. The actin gene family does not play a major role in disease perception or resistance and thus represents how the number of housekeeping genes is affected in the different plant species. Our analyses showed that the percentage of RLKs is similar between monocot aquatic species and monocot terrestrial species falling in a range of 0.84-1.0% (S5 Table). However, dicot aquatic species showed a reduction in the RLKs. Orchid has the lowest percentage of proteins encoding RLKs among terrestrial dicots at 0.76%, far greater than the 0.3% and 0.006% we found in corkscrew plant and humped bladderwort, respectively. Across all aquatic species the reduction of RLKs compared to sister terrestrial lineages was not as pronounced as with NLRs, with the exception of corkscrew plant (S3,5 Table). For the actin gene family, the percentage of actin encoding genes in the proteome were the 2nd (0.13%), 3rd (0.11%), 4th (0.11%) and 8th (0.076%) highest among the 12 species sampled. This is consistent with the conservation of core genes despite the genome reduction among aquatic species (S5 Table).

Both monocot and dicot aquatic plants independently lost immune signalling pathway components *EDS1/PAD4/SAG101*

To identify if other immune signalling components had been lost in aquatic plant species we performed reciprocal BLASTP search for 15 known immune components (S1 Fig and S6 Table) (Fig. 4A). Of the 15 immune genes characterized in Arabidopsis, 5 were found to be conserved across monocots and dicots, including aquatic lineages (Fig. 4B). Among the genes that we identified as lost was *SAG101,* which has previously been shown to be absent in monocots and a few dicots including *A. coerulea* and *M. guttatus (32)*. The absence of a reciprocal orthologue for *PBS1-Like-1* (*PBL1) (50)* can be explained by the recent duplications at this locus (S2 Fig).

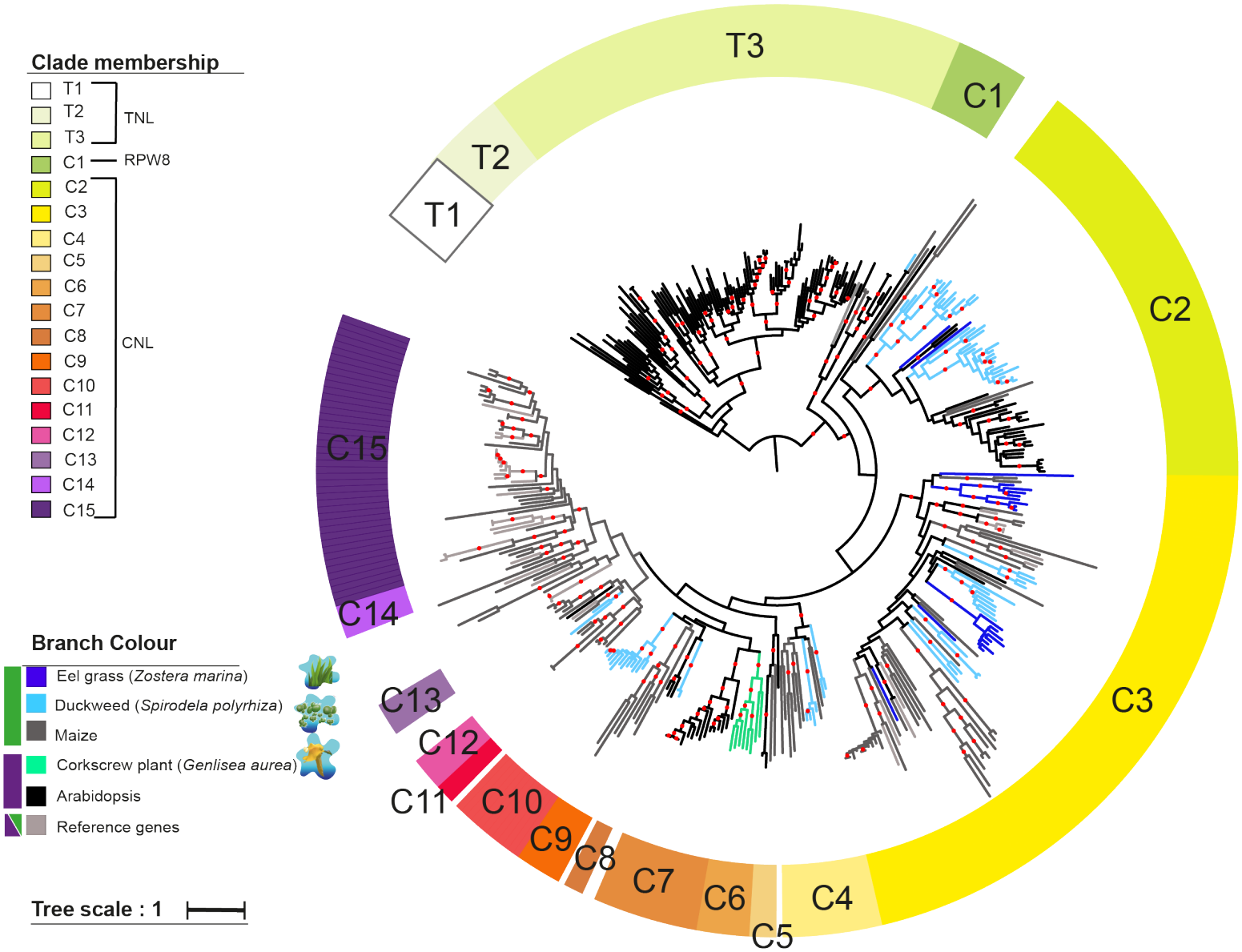


Figure 3 - **Maximum likelihood phylogeny of NLRs in monocot and dicot aquatic species, rice, Arabidopsis and selected reference NLRs.** The maximum likelihood tree is based on the alignment of NB-ARC domain. Bootstraps >80 are indicated by a red dot, branch colours denote species. Clades as defined by bootstrap >80 and more than 2 genes are denoted by coloured arc. Domain structure associated with clades is TIR-NLRs (T1-3), CC_{RPW8}-NLRs (C1) and CC-NLRs (C2-C15). Green and purple bars on branch colour figure legend denote monocot and dicot species, respectively.

In addition to *SAG101*, the aquatic monocot and dicot species with low *NLR* number appeared to have convergently lost *EDS1* and *PAD4* (Fig. 4B). To confirm that the inferred absence of *EDS1, PAD4* and *SAG101* is not an annotation artifact, we scanned all the genomes using TBLASTN and HMMER motif searches for the indicative lipase 3 motif. These two analyses supported absence of all three of these proteins, which have been shown to assemble into a complex (S3 Fig, https://github.com/krasileva-group/Aquatic_NLR/tree/master/).

Another key signalling component downstream of NLRs, *NON RACE SPECIFIC DISEASE RESISTANCE 1 (NDR1),* is known for its role in activating both PTI and ETI branches of immunity (51) and has a demonstrated role in maintaining adhesion between the plasma membrane and cell wall (52). Similar to *SAG101*, *NDR1* appears to be absent in some terrestrial monocot species and in both sets of aquatic plants (Fig. 4B). The phylogeny of the identified *NDR1* orthologs suggests three independent loss events of *NDR1*, one at the branch point of each of the aquatic clades and a further loss in the *Poales* order after the divergence from oil palm (S4 Fig). RIN4, which is known to interact with NDR1 in *Arabidopsis,* is markedly present in all species (53).

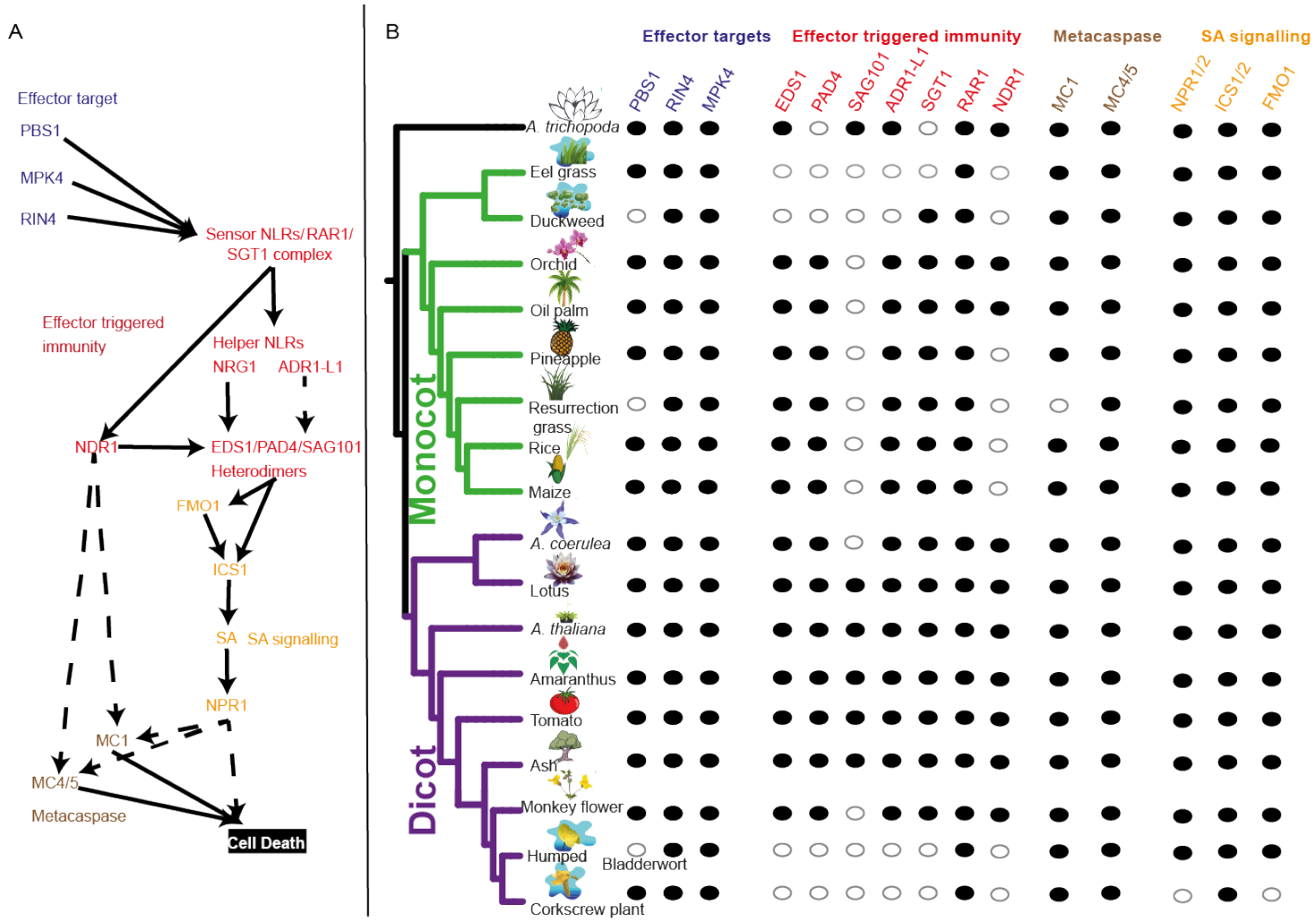


Figure 4 – **Presence/absence analysis of the known plant immunity components** (A) Rows denote species, which are arranged as per phylogenetic relationship, with green branches indicating monocots and purple branches for dicots. Columns denote the presence or absence of known components of the NLR signalling pathway. Gene names are listed on top and arranged by their functional group. Filled circles indicate presence and clear circles represent absence, as defined by reciprocal BLAST and confirmed by TBLASTN (-max_hsps_per_subject =1, evalue 1e-10) and where available by Ensembl gene trees.

(B) - Schematic model of interactions between components of Fig 3A. Arrows denote direct or indirect interaction with dotted arrows for interactions which have been suggested but lack clear functional validation in the literature.

Orthogroup analysis of protein families provides a global view of the genes convergently lost in aquatic plants, while conserved in terrestrial lineages

The convergent loss of *EDS1* and *EDS1*-dependent *NLRs* led us to hypothesize that other, as yet unknown, components of the *EDS1*-dependent signalling cascade would also have been convergently lost in aquatic species. To uncover novel proteins which can potentially function in conjunction with the EDS1 mediated NLR signalling cascade, we performed two analyses to identify orthogroups; OrthoMCL and GeneSeqToFamily. These methods were applied to the 18 plant proteomes including; monocots (*Z. marina, S. polyrhiza, P. equestris, Dioscorea rotundata*, *Eleais guineesis, O. thomaeum, Oryza sativa* and *Zea mays*) and dicots (*Aquilegia coerulea, Nelumbo nucifera, Arabidopsis thaliana, Amaranthus hypochondriachus, Solanum lycopersicum, Fraxinus excelsior, Mimulus guttatus, G. aurea* and *U. gibba*) both with the outgroups *Selaginella moellendorffii* and *Amborella trichopoda* (Fig 5A).

We identified 17 genes lost in only aquatic species from combining the monocot and dicot OrthoMCL runs and 31 genes from 10 orthogroups lost in only aquatic species from GeneSeqToFamily (Fig. 5). Four genes were identified by both pipelines including *EDS1, PAD4, ACTIVATED DISEASE RESISTANCE-LIKE1 (ADR1-L1) and REGULATOR OF CHROMOSOME CONDENSATION 1- LIKE (RCC1-like)*. The former three are known to be involved in plant defence, whilst *RCC1-like* has not previously been implicated in defence response. We further focused on the 44 genes identified by either OrthoMCL or GeneSeqToFamily methods (Fig 5, S7 Table). We designated these genes as *AngioSperm Terrestrial-Retained, Aquatic-Lost (ASTRAL).*

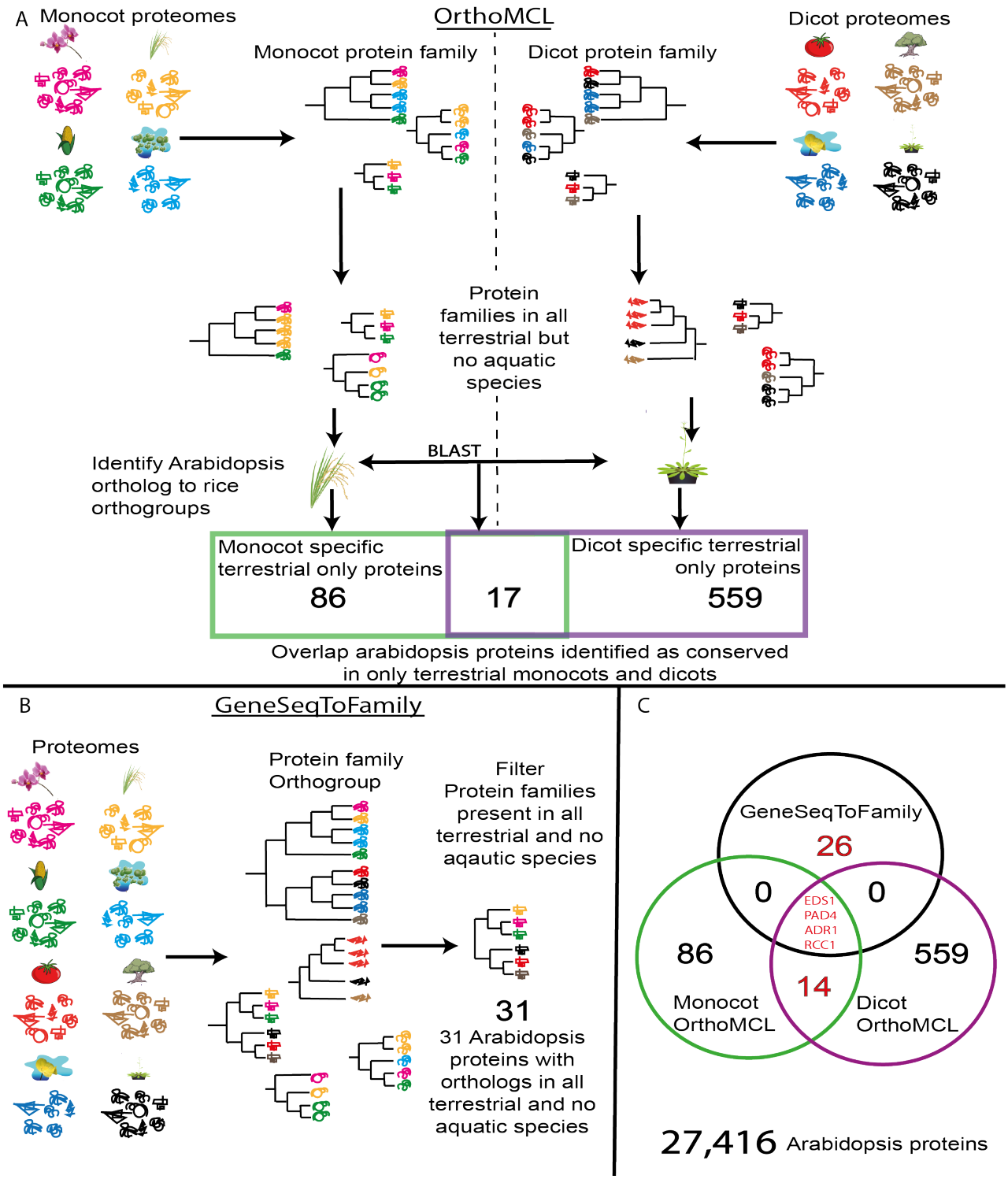


Figure 5 – **Protein family analysis to identify *AngioSperm Terrestrial-Retained, Aquatic-Lost (ASTRAL)* genes** (A) Schematic of the OrthoMCL approach to cluster protein families separately among monocot and dicot species and then filtering for protein families present in terrestrial and absent in aquatic species. Proteins are denoted by different line drawings, colours of proteins represent species origin and phylogenetic trees represent gene trees for each protein. The Venn diagram at bottom provides the number of *Arabidopsis* proteins, which are absent in just monocot, dicot or both species. (B) Illustration of the GeneSeqToFamily method which uses the monocot and dicot proteomes together to establish gene trees across the angiosperms. (C) A Venn diagram summarizing the results of two methods. Genes and numbers marked in red are those subsequently referred to as *ASTRAL* genes.

Arabidopsis and rice homologs of genes lost in aquatic plants are differentially regulated upon drought response and disease resistance

We first took the 44 *ASTRAL* gene candidates from *Arabidopsis* and identified only two broad condition perturbations, drought and pathogen infection, were identified as causing differential expression (up or down) (fold change > 3, p-value = < 0.001) of more than 20 of our candidate genes from the 697 microarray experiments available at https://www.ncbi.nlm.nih.gov/geo/. To further investigate the effects of two of these conditions on the *ASTRAL* genes, we selected representative gene expression sets for these conditions (drought - AT-00626 (54) / AT-00419 (55) (micro-array), pathogen-AT-00744 (56) / AT-00736 (57) (RNAseq)). We extracted the differentially expressed genes in control vs treated conditions (FDR = 0.05, S1 data) and identified the number of *ASTRAL* genes that were differentially expressed (Fig. 6).

A significant association was present for differentially expressed *ASTRAL* genes upon drought compared to untreated in the Col-0 and in a mutant of a major regulator of drought, *srk2cf* background (AT-00419, One sided fisher-exact test, p-value = 0.03612) and at both 1 hour and 4 hours after drought stress of Col-0 compared to control (AT-00626, One sided fisher-exact test, p-value = 0.03678) (Fig. 6A). Across the two drought experiments tested, the same 11 *ASTRAL* genes were differentially expressed in response to drought (Fig. 6A). Among the upregulated *ASTRAL* genes was *EID-1 Like 3 (EDL3) (58),* a known drought response regulator. The *ASTRAL* genes also appear to show greater differential expression after longer incubation with the hormone abscisic acid ABA, in an *SNF1-RELATED PROTEIN KINASE* 2 (SnRK2) subclass II (SRK2C and SRK2F) independent manner (Fig. 6A).

The *ASTRAL* gene set was associated with differential expression upon treatment of Col-0 rosette leaves with *Pseudomonas syringae* (One sided fisher-exact test, p-value = 0.00433).

Upon treatment of necrotrophic pathogen *Botrytis cineria* there was no association between differentially expressed genes and ASTRAL candidates in a wild-type genetic background (One sided fisher-exact test, p-value = 0.076) (Fig. 6B). Interestingly, in the *wrky33* mutant background there was a significant association between differentially expressed and ASTRAL genes (One sided fisher-exact test, p-value = 0.0147) (Fig. 6B). WRKY33 has been shown to be crucial for providing immunity to *B. cineria* in *A. thaliana* through negative regulation of ABA (57). 15 *ASTRAL* genes were differentially expressed in the *B. cineria wrky33* mutant background whilst 20 *ASTRAL* genes were differentially expressed upon *P. syringae* treatment of Col-0. Of these genes 11 *ASTRAL* genes were differentially expressed in both experiments. Since *EDS1* is known to be induced upon pathogen infection, we were interested to see that upon hierarchical clustering of expression pattern 11 *A. thaliana* genes fell into a cluster with EDS1 , whilst 8 of these genes have previously been implicated in effector triggered immunity, 3 genes (*RPP13-like Protein 1 (RPPL1)*, *AT1G55790* and *AT5G66890*) have not previously been experimentally shown to play a role in immunity (Fig. 6B). Two other clades of differentially expressed genes were of particular interest: the 5 gene clade characterised by another NLR helper gene *NRG1* and the 7 gene clade containing *CTC-INTERACTING DOMAIN 9 (CID9)*, *MICROTUBULE-ASSOCIATED PROTEIN 65-8 (MAP65-8)* and *ASPARTIC PROTEASE IN GUARD CELL 2 (ASPG2)* all of which were down-regulated upon pathogen infection (Fig. 6B). All three clades of up and down-regulated *ASTRAL* genes showed minimal changes in expression in *SALICYLIC ACID INDUCTION DEFICIENT 2* (*sid2-1)* knockouts compared to wild type. This matches our expectation since SA pathway components were conserved in several aquatic species. We observed much larger effect in the *AGD-2 DEFENCE LIKE PROTEIN 1* (*ald-1-2)* background. *ALD1* encodes an aminotransferase that works synergistically with PAD4, the two genes have been shown to affect expression of one another (59,60).

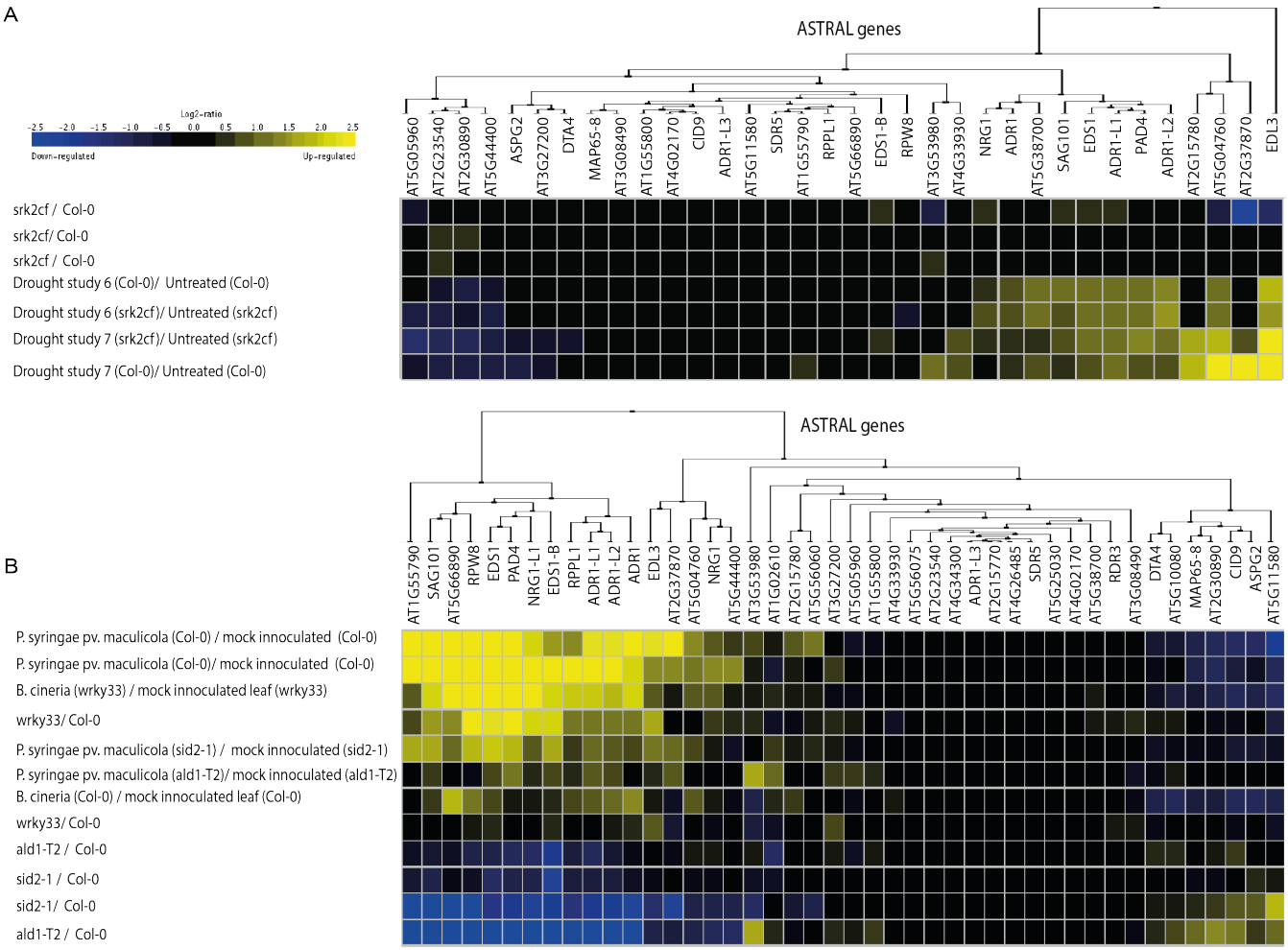


Figure 6 **Hierarchical clustering of differential expression of *ASTRAL* genes in *A. thaliana* upon biotic and abiotic stress** (A) Microarray (AT-00419) differential gene expression of ASTRAL genes upon drought stimulus in Col-0 and *srk2cf* genetic background. (B) RNAseq (AT-00744, AT-00736) of differential expression of *A. thaliana ASTRAL* genes upon necrotrophic (*B. cineria*) and biotrophic (*P. syringae*) treatment.

Four of the 11 drought differentially expressed *ASTRAL* genes were also differentially expressed upon *P. syringae* and *B. cineria wrky33* treatments. An additional 4 *ASTRAL* genes that changed expression upon drought where differentially expressed in one of the two pathogen treatments.

To identify if the same stresses disproportionately affect differential expression of *ASTRAL* genes in a monocot species we used the 34 rice genes which make up the *ASTRAL* gene set (S8 Table). We looked for patterns of high differential expression of rice *ASTRAL* genes across 25 RNAseq and 142 microarray datasets available at https://www.ncbi.nlm.nih.gov/geo/. Drought conditions in several studies appeared to cause an increase in the number of differentially expressed *ASTRAL* genes. We tested the association between *ASTRAL* and differentially expressed genes for two rice drought studies (OS-00140 (61) and OS-00143 (62) using FDR = 0.05). A significant association was found between differentially expressed and *ASTRAL* genes on the 9th day of sampling rainfed rice compared to irrigated gene expression (One sided fisher-exact test, p-value = 0.00089) but not on day 14 or 15 respectively (One sided fisher-exact test, p-value = 0.81, 0.245). A significant association was also found between *ASTRAL* and differentially expressed genes upon drought experiments simulating 10% or 15% available water content across three rice cultivars (One sided fisher-exact test, p-value = 0.0046) (Fig. 7A). A total of 11 genes were found to be differentially expressed upon drought. Six of these *ASTRAL* genes were found differentially expressed in both conditions. The rice ortholog of *EDL3* (*LOC_Os01g58850*) was differentially expressed upon conditions of 15% available water capacity (AWC) but not at 10% AWC.

We then tested if there was a significant association between rice *ASTRAL* genes and differential expression upon pathogen infection. For this we looked at three infection datasets: bacterial blight *Xanthomonas oryzae,* rice blast *Magnaporthe oryzae* and rice gall midge *Orseolia oryzae* (OS-00139 (63), OS-00045 (64), OS-00082 (65)). We decided to include a dataset of the insect O. oryzae as it has been shown that insects secrete effectors which activate plant immunity in a similar manner to other pathogens. Furthermore, insects commonly act as vectors for plant infection (65-67). There was no significant association between differential gene expression and rice *ASTRAL* genes upon *X. oryzae* infection with only 5 of the

34 ASTRAL genes differentially expressed upon infection when comparing all isolates of *X. oryzae* to control treatments (One sided fisher-exact test, p-value = 0.3054) (Fig. 7B). Conversely, upon infection with *O. oryzae (GMB1, GMB4M)* and *M. oryzae* (2dpi) there was a significant association between *ASTRAL* and differentially expressed genes (One sided fisher-exact test, p-value =0.0001, 0.0026, 0.0428) (Fig. 7B). All 4 genes differentially expressed upon rice gall midge GMB1 infection were also differentially expressed in GMB4M along with one additional gene. However, a different set of 5 other non-overlapping *ASTRAL* genes were differentially expressed in response to *M. oryzae* (Fig. 7B). Among the genes differentially expressed upon *M. oryzae* infection was rice ortholog of *PAD4* (*LOC_Os11g09010*), although there was no significant differentially expression of *EDS1* (*LOC_Os09g22450*) between treated and untreated conditions.

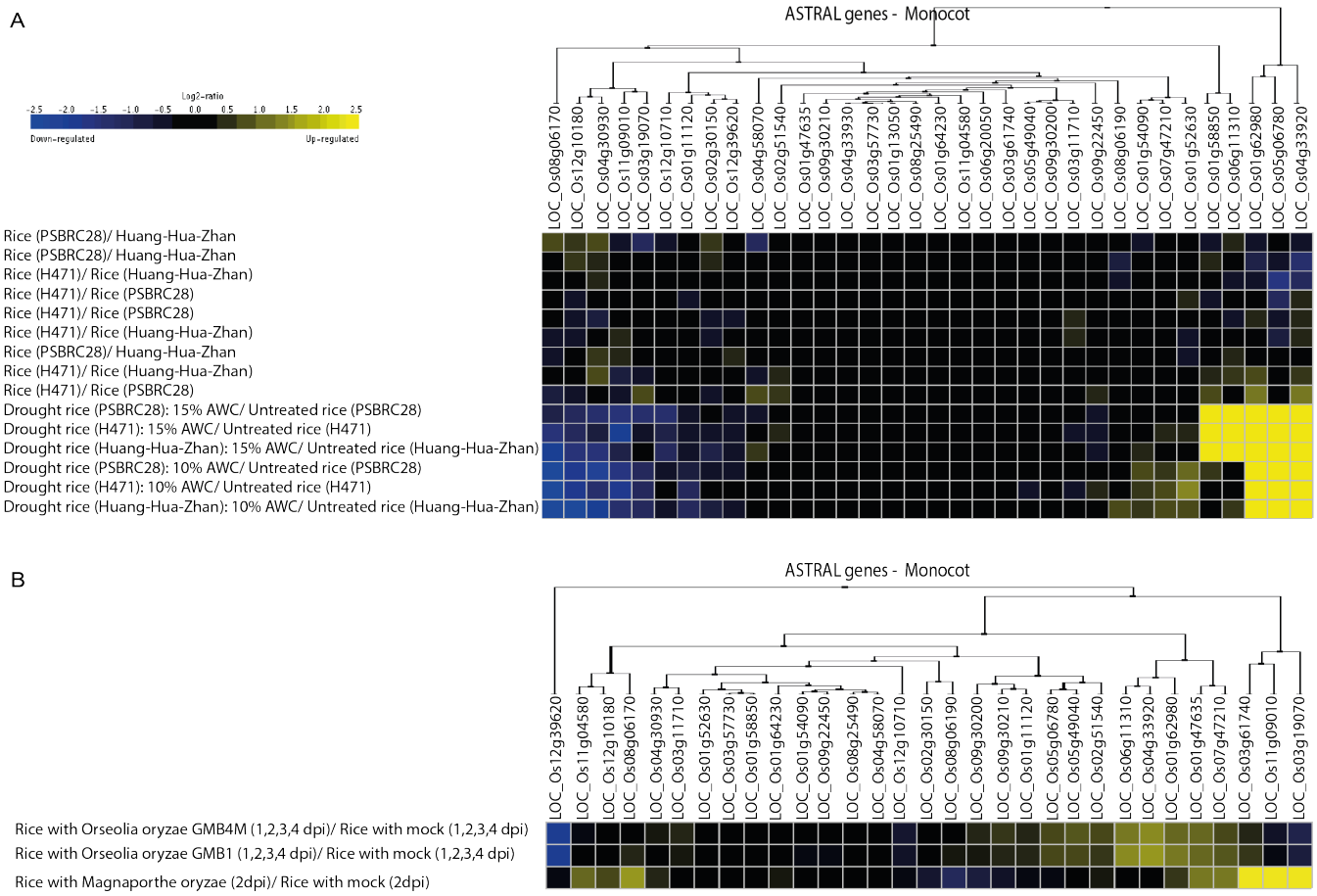


Figure 7 **Differentially expression analysis of rice *ASTRAL* genes upon biotic and abiotic stress** (A) RNAseq (OS-00143) heatmap of differential expression of rice ASTRAL genes upon drought treatment regimes. (B) Microarray (OS-00045, OS-00082) differential gene expression heatmap of rice ASTRAL genes upon pathogen and pest treatment.

A model placing *ASTRAL* genes in immunity and drought response pathways

We assembled a working model of known plant immunity and drought pathways from literature (Fig 8A, 8B, S9, S10 Table). *ASTRAL* genes experimentally validated in the plant immunity (15-17,35,68) pathway were underlined in purple and those validated in the drought pathway (69,70) in brown, interactions consistent with current literature were depicted with solid arrows. For the remaining *ASTRAL* genes identified as differentially expressed upon pathogen treatment we investigated there expression further. In plant immunity, we identified whether the genes expression was *EDS1* dependent, utilising microarray from eds1 mutant plants (71), this revealed 7 *EDS1* dependent genes which had not previously been implicated with *EDS1* yet have an inverted expression pattern in the eds1 mutant (S9 Table). *ASTRAL* genes genetically downstream of *EDS1* based on expression data were connected by a dashed arrow. To construct how the *EDS1* downstream *ASTRAL* genes contribute to known immunity processes a further literature search was used to identify known roles of the genes or conditions in which they were differentially expressed (57,71,72). For the 7 genes whose expression was *EDS1* independent we conducted a literature search to see if they had been implicated previously in plant immunity and incorporated this data into the model (Fig 8A, S9 Table) (72,73).

For the drought response pathway, all *ASTRAL* genes which were known to be differentially expressed upon ABA were included (Fig 6, 7). To dissect where in the ABA drought response pathway the genes were involved we carried out a literature search (S10 Table) using information from literature and associated gene expression datasets (74,75) we were able to

identify putative genetic positions of *ASTRAL* genes in ABA response, these were indicated by dashed arrows (Fig 8B). Interestingly, we were also able to identify some literature linking the pathogen responsive *ASTRAL* genes to ABA signaling (74-76) and the reciprocal (77,78) (79,80) even though a gene was often only identified as differentially expressed in one of the two conditions. Furthermore, the models highlighted the convergence of drought and pathogen response on similar phenotypes such as senescence and stomatal changes.

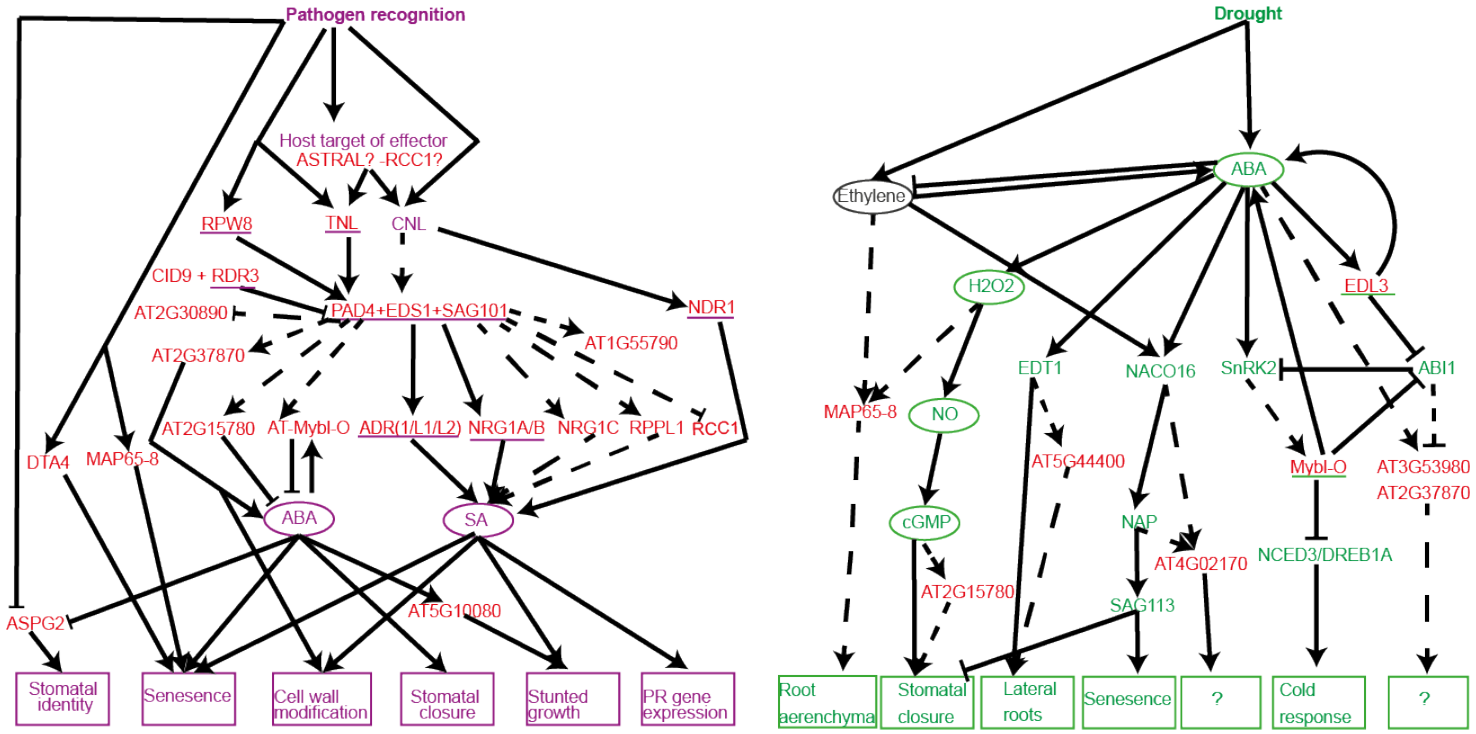


Figure 8 **Schematic model of relationship between *ASTRAL* genes and known biotic and abiotic stress pathways in *A. thaliana*** (A) Model based on available RNAseq of potential position of ASTRAL genes within known *Arabidopsis thaliana* biotrophic pathogen disease resistance genetic pathway. (B) Model based on available RNAseq of potential position of ASTRAL genes within known *Arabidopsis thaliana* drought response genetic pathway

Discussion

Previous studies have investigated the genome content of one or two aquatic plant species and highlighted gene loss linked to embryogenesis and root development in humped bladderwort (43), cell wall processes and ABA in duckweed (39,40), and defence response, stomata, terpenoid and hormone pathways in eelgrass (41). In this study, we investigated the convergent adaptation to an aquatic environment across four aquatic plants and identified gene losses linked to the ETI immune signalling pathway. We observed a drastic reduction of *NLR* immune genes together with the absence of the known signalling components. Using comparative genomics, we identified additional genes that were convergently lost among aquatic lineages and mapped them to defence and drought response pathways using differential expression analyses.

While aquatic species have low numbers of NLRs, phylogenetic analysis shows that many of the retained NLRs come from independent recent expansions of species-specific clades. Furthermore, we were able to identify the parallel loss of the EDS1/PAD4 immune complex in these species as well as other genes such as *SAG101 (32,34), NRG1 (16,17), ADR1 (26)* and *NDR1 (51,53)* which have all previously been shown to be genetically required for signalling of certain *NLRs*. The loss of these components is supported by the absence of *TIR1 NLRs* in dicot aquatic species, which require *EDS1/PAD4* and often *NRG1* to cause plant cell death. We subsequently identified orthogroups of genes convergently lost in monocots and dicots with recent aquatic life history. We show that some of these genes have differential expression upon pathogen infection. Finally, we highlighted several orthogroups with similar differential expression patterns upon specific stimuli, suggesting these convergently lost genes in aquatic lineages may function in distinct pathways. It is tempting to speculate that the loss of the majority of *NLR* clades and the *EDS1* complex in aquatic species is due to selection and re-

wiring of plant immunity toward a single common mechanism that facilitates increased fitness in the aquatic environment.

Duckweed, eelgrass, corkscrew plant, humped bladderwort and lotus all have the ability to survive with a substantial part of the organism submerged underwater. However, lotus does not have a reduced number of *NLRs* and retains both *EDS1* and *PAD4*. One possible explanation for this is that lotus has roots in the soil despite its aquatic stem. Additionally, lotus has a substantial surface area of leaves and flowers on the water surface which may affect the pathogen load. Hence, for this study we have not considered lotus to truly be an aquatic species.

Interestingly, though the aquatic species typically have a low number of *NLRs*, the retained *NLRs* have undergone a recent rapid expansion. This observation is supported by the short branch length and highly branched nature of the *NLR* clades present in aquatic species. It suggests that these *NLRs* have a function in the aquatic plants. Nevertheless, whether or not they play a role in pathogen defence requires further work. Previous studies have shown NLRs including *VARIATION IN COMPOUND TRIGGERED ROOT GROWTH RESPONSE* (81) and *CHILLING SENSITIVE 1 (82)* can act as receptors of abiotic stress.

Upon pathogen stimulus there are some *ASTRAL* genes which are not differentially expressed, however, we cannot rule out the possibility that those genes may be involved in disease resistance. We have only looked at differential expression in a small subset of the possible combinations of conditions, tissues and pathogens that can result in an immune response. In addition, a gene doesn’t necessarily need to be differentially expressed in order for the protein or RNA it encodes to function in mediating a defence response.

The findings of the recurrent loss of *EDS1/PAD4* and *SAG101* together is consistent with genetic, biochemical and structural studies of EDS1/PAD4/SAG101, which show the three proteins function in heterodimeric complexes (32). Previous studies have also shown the absence of *TIR-1 NLR*s in monocots despite the presence of hundreds of *TIR-1 NLR*s in some dicot species (29). Despite absence of *TIR-1 NLR*s in terrestrial monocots *EDS1* known to function in *TIR-1 NLR* signalling is retained (33). Recently, evidence was provided supporting the requirement of CC_{RPW8} NLR *NRG1* in the signalling of *NLR*s (16,17). We propose that all members of the CC_{RPW8} clade play a role in NLR signalling that involves EDS1 as we find the CC_{RPW8} clade is absent in both independent groups of aquatic monocots and dicots despite its retention along with EDS1 in all terrestrial monocots and dicots. Another signalling component, NDR1, was absent in all aquatic species and some terrestrial species despite the retention of CNLs which commonly signal through NDR1. Interestingly, NDR1 is thought to mediate resistance by controlling fluid loss in the cell (52). This property highlights a possible intersection between the drought responsive *ASTRAL* genes and immunity. Both NDR1 and CC_{RPW8} mediated signalling converge in triggering increase in SA, which appears to be maintained in the aquatic species in our study (83).

Until now, discoveries of crucial components of the plant immune system have relied heavily on mutant screens and differential expression analysis. Here we have shown a complementary approach to identify potential actors in the plant immune system, which can circumvent issues of genetic redundancy by harnessing conservation and independent transitions in distantly related plant lineages. The study also has practical implications in providing new candidates for roles in disease resistance and shedding light on the important question of the downstream genetic reliance of *NLR*s used in agriculture crops. To slow the rate of breakdown of resistance by fast evolving pathogen effectoromes, *NLR*s are often stacked. The downstream signalling components required for NLRs within a stack are rarely considered but if they converge on a

single helper or signaller, this creates a strong selection for effectors that would compromise the downstream component and subsequently break the defence conferred by several *NLR*s at once. In addition, it is crucial to understand the conservation of downstream signalling components to facilitate the successful interspecies transfer of *NLR*s. This study also provides fundamental understanding towards a minimum plant immune system and in doing so reveals potential new model systems such as duckweed, a rapid growing small plant whose reduced ETI immune system could provide a reduced-complexity background for investigating plant immunity. Unexpectedly, this study has begun to further elucidate the complex cross-talk between the plant immune system and drought tolerance. Future studies could use candidates identified to further query the interconnection of the two pathways.

Materials and methods

Genomic datasets used in this study

Genomic assemblies and annotations were obtained from: Phytozome V12 (https://phytozome.jgi.doe.gov/pz/portal.html) for *A. coerulea*, *A. comosus* (v3), *A. hypochondriacus* (325_v1.0), *A. thaliana* (167_TAIR9), *A. trichopoda* (291_v1.0), *M. guttatus* (256_v2.0), *O. sativa* (v7), *O. thomaeum*, *S. lycopersicum* (390_v2.5), *S. moellendorffii* (91_v1), *S. polyrhiza* (v2) and *Z.marina* (v2.2); from COGE (https://genomevolution.org/coge/) for *U. gibba (29027)*; from KEGG for *N. nucifera* (4432), *E. guineesis* (TO3921); from NCBI *P. equestrius* (PRJNA382149), from Ash Tree Genomes (http://www.ashgenome.org/data) for *F. excelsior* (BATG-0.5), from Ensembl for *Dioscorea rotundata* (TDr96_F1_Pseudo_Chromosome_v1.0) and Maize genome database (https://www.maizegdb.org/) for *Z. mays* (AGPv4). The BUSCO scores were calculated using v1.22 version of BUSCO software to compare proteomes to embryophyta_odb9 BUSCO lineage (42).

Annotation, alignment and phylogenetic analysis of NLRs

To annotate NLRs in plant proteomes, the MEME suite (84) based tool NLR-parser (85) was used in addition to the updated version of the NLR-ID pipeline (29) available at https://github.com/krasileva-group/plant_rgenes. Annotations were combined into a non-redundant list of putative NLRs. Where multiple transcripts were present, these were filtered to retain the longest transcript. In addition, a series of characterised NLRs were added. These are available at https://github.com/krasileva-group/Aquatic_NLR/tree/master/Reference%20NLRs. The HMMALIGN programme from the HMMER3.0 suite (86) was used to align proteins to the NB-ARC1_ARC2_prank_aln_domain_ONLY HMM of the NB-ARC domain (22). The alignment was trimmed to the NB-ARC domain region using Belvu (87) and columns and sequences with over 80% gaps where removed. The NB-ARC domain of the remaining NLRs was then manually curated in Jalview (88) allowing no more than 2 consecutive characteristic NLR motifs (Walker A, RNBS-A, WALKER-B, RNBS-C, GLPL, RNBS-D) to be absent. A maximum likelihood phylogenetic tree was constructed using RAXML-MPI (v.8.2.9) (89) with parameters set as: -f a -x 1123 -p 2341 -# 100 -m PROTCATJTT. Trees were visualised and annotated using iTOL and are available at: [**http://itol.embl.de/shared/erin_baggs**].

Ortholog identification

To identify orthologs of specific genes of interest, we used reciprocal blast searches using BLASTP (-max_target_seqs 1 -evalue 1e-10) (BLAST+ 2.2.28.mt) (90). If upon reciprocal BLAST, a homologous gene was not identified, we used ensemble gene trees to check for recent duplication events in the *Arabidopsis* lineage and to confirm our results. Results were manually inspected and filtered using a script available at project github. The presence of *EDS1* in *O. thomaeum* was validated using RNAseq data (BioProject SRS957807) mapped onto the

Oropetium V1 Bio_nano genome assembly (http://www.oropetium.org/resources) (91,92) using HISAT2 (93). Bam files were processed using SAMtools-1.7 (94) and results visualised using IGB (95). The absence of EDS1 was validated by running HMMSEARCH was run on the proteomes of *A. thaliana*, *S. polyrhiza* and *Z. marina* using the Lipase 3 domain which is characteristic of EDS1. Proteins containing the domain were then aligned against the domain using HMMALIGN and the Pfam Lipase 3 HMM. The alignment was manually curated in Belvu before submission to RAXML as above.

To identify orthologous gene groups, OrthoMCL (v2.0.9) (96) was used as described previously (97). Due to large computational requirements, we ran OrthoMCL separately for monocots (*Z. marina, S. polyrhiza, P. equestris, Dioscorea rotundata*, *Eleais guineesis, O. thomaeum, Oryza sativa* and *Zea mays*) and dicots (*Aquilegia coerulea, Nelumbo nucifera, Arabidopsis thaliana, Amaranthus hypochondriachus, Solanum lycopersicum, Fraxinus excelsior, Mimulus guttatus, G. aurea* and *U. gibba*) both with the outgroups *Selaginella moellendorffii* and *Amborella trichopoda.* We overlaid monocot analyses with dicot orthogroups by mapping the monocot gene IDs to the Arabidopsis homologs using reciprocal BLASTP. Additionally, we applied the GeneSeqToFamily pipeline (98) on all 19 genomes. We then filtered orthogroups to identify orthogroups lost in aquatic species (*S. polyrhiza*, *Z. marina, U. gibba*, *G. aurea)* but retained in all terrestrial species of the same monocot/dicot clade. We cannot preclude the possibility that some gene families that have been convergently lost in aquatic species have not also been lost independently in some of the terrestrial lineages. After grouping and manual curation, gene families for which pan-species evolutionary history had been previously established were compared to gene families in our orthogroups. This curation led to the decision to mask *S. moellendorffii*, *A. trichopoda*, *A. coerulea* and *D. rotundata* from later analysis with the former two species rarely having homologs due to large phylogenetic distance and the latter two species having many erroneous gene fusion annotations. The analyses were filtered using

scripts available at https://github.com/krasileva-group/Aquatic_NLR. Grep on cigar string output of parse_output.pl was used to identify gene families present in all terrestrial species diverging after *A. trichopoda* but absent in aquatic species. For comparison of OrthoMCL gene families retained between monocots and dicots, we used *Arabidopsis* and rice proteins as representative genome members and cross-referenced them using blastp reciprocal search (e-value cutoff 1e-10). The validity of this approach was checked on random protein families using Plant Ensembl trees (http://plants.ensembl.org/index.html). The rice gene ids were converted between Phytozome and Plant Ensembl using Plant Ensembl conversion tool (http://rapdb.dna.affrc.go.jp/tools/converter/run).

Expression profiling

The expression analysis for this study was performed and visualised using the 706 rice mRNA samples and 2,836 rice Affymetrix rice genome array samples available on Genevestigator v7.0.3 (https://genevestigator.com). For Arabidopsis the 1,031 Arabidopsis RNA samples and 10,615 Affymetrix Arabidopsis array sample available on Genevestigator v7.0.3 (https://genevestigator.com) were used. The datasets used to dissect drought and pathogen infection gene expression across the 44 *A. thaliana ASTRAL*, monocot and dicot overlapping genes were as follows: RNAseq - Pathogen - AT-00744 (56), AT-00736 (57), microarray - drought - AT-00419 (55), AT-00626 (54). The RNAseq dataset OS-00143 (62) and OS-00140 (61) was used to assay drought induced differential expression among *ASTRAL* genes. Gene expression experiments OS-00139 (63) (RNAseq), OS-0045 (64) (microarray), OS-0082 (65) (microarray) were used to analyse the differential expression of rice *ASTRAL* genes under pathogen infection.

Lists of differentially expressed genes were produced in Genevestigator using parameters FDR = 0.05 (S1 Data). A Fisher-exact test was performed in R (data <- matrix(c(DE ASTRAL, Non-

ASTRAL DE, non-DE ASTRAL genes, remaining genes), nrow =2); fisher.test(data, alternative = "greater"). Hierarchical clustering was generated considering both genes and conditions with parameters of Euclidean distance and optimal leaf-ordering.

Acknowledgements

The authors are grateful to all members of the Krasileva group and their many colleagues, for thoughtful discussion on the presented material. We thank Daniil Prigozhin and Janina Tamborski for suggestions on the manuscript. The high-performance computing resources and services used in this work were supported by the EI Scientific Computing group alongside the NBIP Computing infrastructure for Science (CiS) group.

Contributions

ELB, WH and KVK designed the study. ELB performed NLR, phylogenetic, analysis. AST performed GeneSeqToFamily pipeline. ELB analysed orthogroups from OrthoMCL and GeneSeqToFamily to identify ASTRAL candidates. RNAseq analysis of ASTRAL genes performed by ELB and statistical tests by ELB and KVK. Model of gene interaction by ELB. All authors contributed to final manuscript. All authors read and approved final manuscript.

Supporting information captions

Supplementary table 1 - BUSCO scores for proteomes of 19 species considered for NLR analysis.

Supplementary table 2 - *Arabidopsis* gene identifiers which represent BUSCOs which are validated to be part of pathways lost in aquatic species.

Supplementary table 3 - Table of genome assembly size and number of encoded high and low confidence NLR genes.

Supplementary table 4 - List of characterised NLRs present in clade C3.

Supplementary table 5 - Table of number of actin and kinase HMM domain encoding genes present in genome of species used in OrthoMCL and GeneSeqToFamily.

Supplementary table 6 - Table of genes and gene identifiers of characterised immune genes.

Supplementary table 7 - Table of *Arabidopsis ASTRAL* gene identifiers.

Supplementary table 8 - Table of *rice ASTRAL* gene identifiers.

Supplementary figure 1 - Extended table of presence and absence of known immunity genes from reciprocal BLAST searches.

Supplementary figure 2 - ENSEMBL plant gene tree for PBL1.

Supplementary figure 3 - ENSEMBL plant gene tree for VIP1.

Supplementary figure 4 - Maximum-likelihood gene tree from proteins containing lipase-3 HMM domain.

Supplementary figure 5 - Maximum-likelihood gene tree from proteins in the NPR orthogroup as defined by GeneSeqToFamily.

References

(1) Sanderson MJ, Thorne JL, Wikström N, Bremer K. Molecular evidence on plant divergence times. Am J Bot 2004 Oct;91(10):1656-1665.
(2) Zhong B, Sun L, Penny D. The Origin of Land Plants: A Phylogenomic Perspective. Evol Bioinform Online 2015;11:EBO.S29089.
(3) Liping Zeng, Qiang Zhang, Renran Sun, Hongzhi Kong, Ning Zhang, Hong Ma. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. Nature Communications 2014 Sep 1,;5(1):4956.
(4) FAO Statistics. Available at: http://www.fao.org/faostat/. Accessed 7.1., 2019.
(5) Ramegowda V, Senthil-Kumar M. The interactive effects of simultaneous biotic and abiotic stresses on plants: Mechanistic understanding from drought and pathogen combination. J Plant Physiol 2015;176:47-54.
(6) Michael Krings, Thomas N. Taylor, Hagen Hass, Hans Kerp, Nora Dotzler, Elizabeth J. Hermsen. Fungal Endophytes in a 400-million-yr-old Land Plant: Infection Pathways, Spatial Distribution, and Host Responses. The New Phytologist 2007 Jan 1,;174(3):648-657.
(7) Villarreal A JC, Crandall-Stotler BJ, Hart ML, Long DG, Forrest LL. Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. New Phytologist 2016 Mar;209(4):1734-1746.
(8) Urbach JM, Ausubel FM. Data from: The NBS-LRR architectures of plant R-proteins and

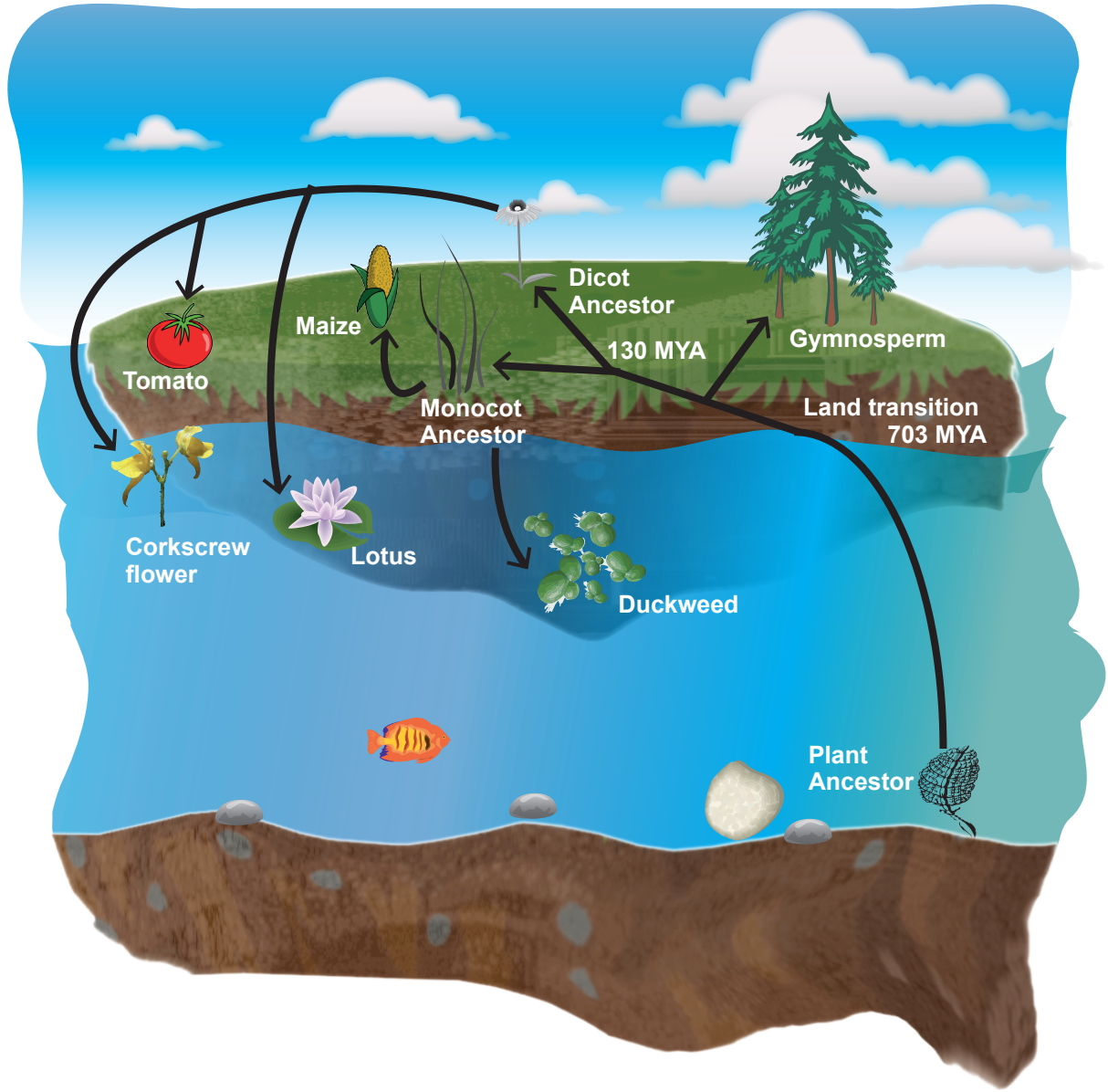
metazoan NLRs evolved in independent events. 2017;114(5):1063-1068.
(9) Barba-Montoya J, Reis M, Schneider H, Donoghue PCJ, Yang Z. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. New Phytologist 2018 Apr;218(2):819-834.
(10) Jones JDG, Vance RE, Dangl JL. Intracellular innate immune surveillance devices in plants and animals. Science 2016;354(6316).
(11) Catanzariti A, Dodds PN, Ve T, Kobe B, Ellis JG, Staskawicz BJ. The AvrM effector from flax rust has a structured C-terminal domain and interacts directly with the M resistance protein. Mol Plant-Microbe Interact 2010;23(1):49-57.
(12) Dodds PN, Lawrence GJ, Ellis JG. Six amino acid changes confined to the leucine-rich repeat β-strand/β-turn motif determine the difference between the P and P2 rust resistance specificities in flax. Plant Cell 2001;13(1):163-178.
(13) Krasileva KV, Dahlbeck D, Staskawicz BJ. Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. Plant Cell 2010;22(7):2444-2458.
(14) Bernoux M, Ve T, Williams S, Warren C, Hatters D, Valkov E, et al. Structural and Functional Analysis of a Plant Resistance Protein TIR Domain Reveals Interfaces for Self-Association, Signaling, and Autoregulation. Cell host & microbe 2011;9(3):200-211.
(15) Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL. Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. Proceedings of the National Academy of Sciences 2011;108(39):16463-16468.
(16) Castel B, Ngou P, Cevik V, Redkar A, Kim D, Yang Y, et al. Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. New Phytol 2018;0(0).
(17) Qi T, Seong K, Thomazella DPT, Kim JR, Pham J, Seo E, et al. NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. Proc Natl Acad Sci USA 2018;115(46):E10979.
(18) Le Roux C, Huet G, Jauneau A, Camborde L, Trémousaygue D, Kraut A, et al. A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. Cell 2015;161(5):1074-1088.
(19) Ortiz D, De Guillen K, Cesari S, Chalvon V, Gracy J, Padilla A, et al. Recognition of the Magnaporthe oryzae effector AVR-Pia by the decoy domain of the rice NLR immune receptor RGA5. Plant Cell 2017;29(1):156-168.
(20) Sarris PF, Duxbury Z, Huh SU, Ma Y, Segonzac C, Sklenar J, et al. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. Cell 2015;161(5):1089-1100.
(21) Saucet SB, Ma Y, Sarris PF, Furzer OJ, Sohn KH, Jones JD. Two linked pairs of Arabidopsis TNL resistance genes independently confer recognition of bacterial effector AvrRps4. Nature communications 2015;6:6338.
(22) Bailey PC, Schudoma C, Jackson W, Baggs E, Dagdas G, Haerty W, et al. Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions. Genome Biol 2018;19(1):23.
(23) Cesari S, Bernoux M, Moncuquet P, Kroj T, Dodds PN. A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. Frontiers in plant science 2014;5:606.
(24) Wu C, Derevnina L, Kamoun S. Receptor networks underpin plant immunity. Science 2018;360(6395):1300.
(25) Wu C, Abd-El-Haliem A, Bozkurt TO, Belhaj K, Terauchi R, Vossen JH, et al. NLR network mediates immunity to diverse plant pathogens. Proc Natl Acad Sci USA 2017;114(30):8113.
(26) Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL. Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. Proc Natl Acad Sci USA 2011;108(39):16463.

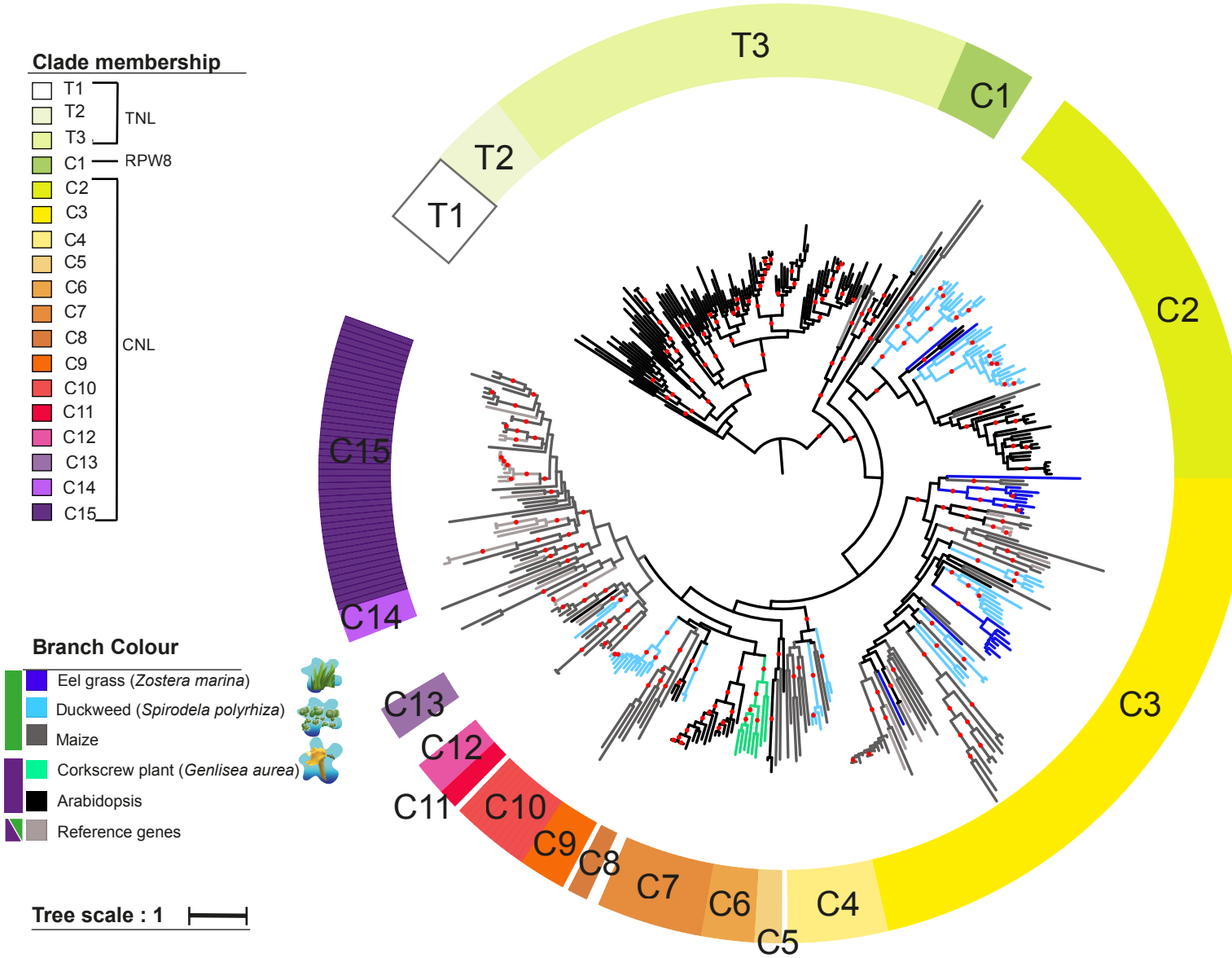
(27) Baggs E, Dagdas G, Krasileva KV. NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. Curr Opin Plant Biol 2017;38:59-67.
(28) Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. Genome research 2017 May;27(5):885-896.
(29) Sarris PF, Cevik V, Dagdas G, Jones JDG, Krasileva KV. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. BMC Biology 2016;14(1):8.
(30) Williams SJ, Sohn KH, Wan L, Bernoux M, Sarris PF, Segonzac C, et al. Structural basis for assembly and function of a heterodimeric plant immune receptor. Science 2014;344(6181):299-303.
(31) Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. Plant Cell 2003;15(4):809-834.
(32) Wagner S, Stuttmann J, Rietz S, Guerois R, Brunstein E, Bautor J, et al. Structural Basis for Signaling by Exclusive EDS1 Heteromeric Complexes with SAG101 or PAD4 in Plant Innate Immunity. Cell Host & Microbe 2013;14(6):619-630.
(33) Chen G, Wei B, Li G, Gong C, Fan R, Zhang X. TaEDS1 genes positively regulate resistance to powdery mildew in wheat. Plant Mol Biol 2018;96(6):607-625.
(34) Zhu S, Jeong R, Venugopal SC, Lapchyk L, Navarre D, Kachroo A, et al. SAG101 Forms a Ternary Complex with EDS1 and PAD4 and Is Required for Resistance Signaling against Turnip Crinkle Virus. PLOS Pathogens 2011;7(11):e1002318.
(35) Cui H, Qiu J, Zhou Y, Bhandari DD, Zhao C, Bautor J, et al. Antagonism of Transcription Factor MYC2 by EDS1/PAD4 Complexes Bolsters Salicylic Acid Defense in Arabidopsis Effector-Triggered Immunity. Mol Plant 2018 Aug 6;11(8):1053-1066.
(36) Maekawa T, Kracher B, Vernaldi S, Ver Loren van Themaat, Emiel, Schulze-Lefert P. Conservation of NLR-triggered immunity across plant lineages. PNAS 2012 -12-04 00:00:00;109(49):20119-20123.
(37) Ross TG, Barrett CF, Soto Gomez M, Lam VKY, Henriquez CL, Les DH, et al. Plastid phylogenomics and molecular evolution of Alismatales. Cladistics 2016;32(2):160-178.
(38) Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang T, et al. Architecture and evolution of a minute plant genome. Nature 2013;498:94.
(39) Michael TP, Bryant D, Gutierrez R, Borisjuk N, Chu P, Zhang H, et al. Comprehensive definition of genome features in Spirodela polyrhiza by high-depth physical mapping and short-read DNA sequencing strategies. Plant J 2017;89(3):617-635.
(40) Wang W, Haberer G, Gundlach H, Gläßer C, Nussbaumer T, Luo MC, et al. The Spirodela polyrhiza genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nature Communications 2014;5:3311.
(41) Olsen JL, Rouzé P, Verhelst B, Lin Y, Bayer T, Collen J, et al. The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. Nature 2016;530:331.
(42) Kriventseva EV, Zdobnov EM, Simão FA, Ioannidis P, Waterhouse RM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015;31(19):3210-3212.
(43) Leushkin EV, Sutormin RA, Nabieva ER, Penin AA, Kondrashov AS, Logacheva MD. The miniature genome of a carnivorous plant Genlisea aurea contains a low number of genes and short non-coding sequences. BMC Genomics 2013 Jul 15;14:476.
(44) Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. Genesis 2015;53(8):474-485.
(45) Nandety RS, Caplan JL, Cavanaugh K, Perroud B, Wroblewski T, Michelmore RW, et al. The role of TIR-NBS and TIR-X proteins in plant basal defense responses. Plant Physiol

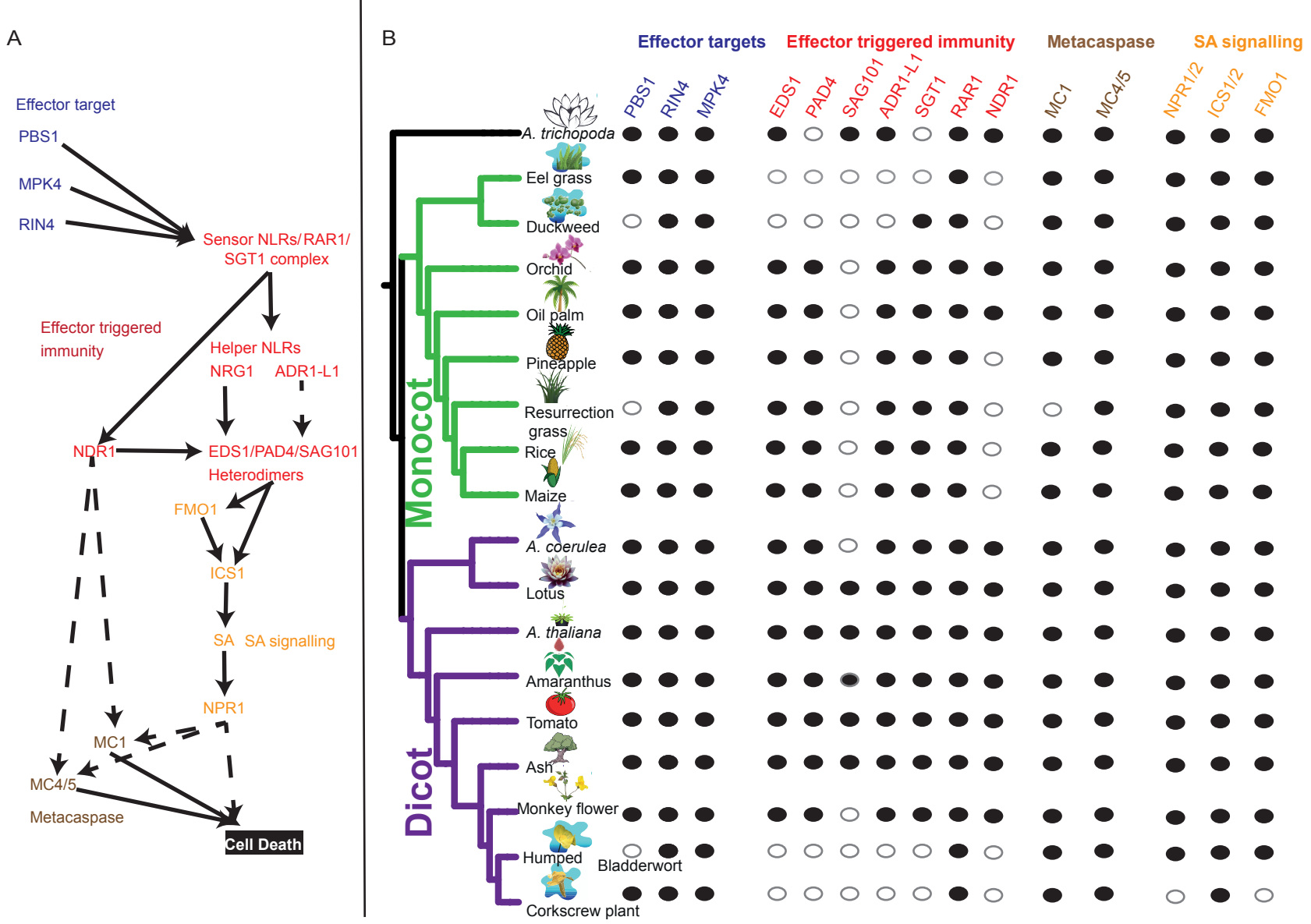
2013;162(3):1459-1472.
(46) Tarr E, M Alexander H. TIR-NBS-LRR genes are rare in monocots: Evidence from diverse monocot orders. ; 2009.
(47) Axtell MJ, Staskawicz BJ. Initiation of RPS2-Specified Disease Resistance in Arabidopsis Is Coupled to the AvrRpt2-Directed Elimination of RIN4. Cell 2003;112(3):369-377.
(48) Mackey D, Belkhadir Y, Alonso JM, Ecker JR, Dangl JL. Arabidopsis RIN4 Is a Target of the Type III Virulence Effector AvrRpt2 and Modulates RPS2-Mediated Resistance. Cell 2003;112(3):379-389.
(49) Bianchet C, Wong A, Quaglia M, Alqurashi M, Gehring C, Ntoukakis V, et al. An Arabidopsis thaliana leucine-rich repeat protein harbors an adenylyl cyclase catalytic center and affects responses to pathogens. J Plant Physiol 2019;232:12-22.
(50) Zhang J, Li W, Xiang T, Liu Z, Laluk K, Ding X, et al. Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a Pseudomonas syringae effector. Cell Host Microbe 2010 Apr 22;7(4):290-301.
(51) Knepper C, Savory EA, Day B. The role of NDR1 in pathogen perception and plant defense signaling. Plant signaling & behavior 2011;6(8):1114-1116.
(52) Knepper C, Savory EA, Day B. Arabidopsis NDR1 Is an Integrin-Like Protein with a Role in Fluid Loss and Plasma Membrane-Cell Wall Adhesion. Plant Physiol 2011;156(1):286-300.
(53) Day B, Dahlbeck D, Staskawicz BJ. NDR1 interaction with RIN4 mediates the differential activation of multiple disease resistance pathways in Arabidopsis. Plant Cell 2006;18(10):2782-2791.
(54) Pandey N, Ranjan A, Pant P, Tripathi RK, Ateek F, Pandey HP, et al. CAMTA 1 regulates drought responses in Arabidopsis thaliana. BMC Genomics 2013;14(1):216.
(55) Umezawa T, Shinozaki K, Mizoguchi M, Takasaki H, Yamaguchi-Shinozaki K, Kidokoro S, et al. Two Closely Related Subclass II SnRK2 Protein Kinases Cooperatively Regulate Drought-Inducible Gene Expression. pcp 2010;51(5):842-847.
(56) Bernsdorff F, Doring AC, Gruner K, Schuck S, Brautigam A, Zeier J. Pipecolic Acid Orchestrates Plant Systemic Acquired Resistance and Defense Priming via Salicylic Acid-Dependent and -Independent Pathways. Plant Cell 2016 Jan;28(1):102-129.
(57) Liu S, Kracher B, Ziegler J, Birkenbihl RP, Somssich IE. Negative regulation of ABA signaling by WRKY33 is critical for Arabidopsis immunity towards Botrytis cinerea 2100. Elife 2015 Jun 15;4:e07295.
(58) Koops P, Pelser S, Ignatz M, Klose C, Marrocco-Selden K, Kretsch T. EDL3 is an F-box protein involved in the regulation of abscisic acid signalling in Arabidopsis thaliana. J Exp Bot 2011 Nov;62(15):5547-5560.
(59) Song JT, Lu H, McDowell JM, Greenberg JT. A key role for ALD1 in activation of local and systemic defenses in Arabidopsis. Plant J 2004 Oct;40(2):200-212.
(60) Song JT, Lu H, Greenberg JT. Divergent Roles in *Arabidopsis thaliana* Development and Defense of Two Homologous Genes, *ABERRANT GROWTH AND DEATH2* and *AGD2-LIKE DEFENSE RESPONSE PROTEIN1*, Encoding Novel Aminotransferases. Plant Cell 2004;16(2):353-366.
(61) Plessis A, Hafemeister C, Wilkins O, Gonzaga ZJ, Meyer RS, Pires I, et al. Multiple abiotic stimuli are integrated in the regulation of rice gene expression under field conditions. eLife 2015;4:e08411.
(62) Huang L, Zhang F, Zhang F, Wang W, Zhou Y, Fu B, et al. Comparative transcriptome sequencing of tolerant rice introgression line and its parents in response to drought stress. BMC Genomics 2014 Nov 26;15:1026.
(63) Wilkins K, Booher N, Wang L, Bogdanove A. TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen Xanthomonas oryzae pv. oryzicola while strict conservation suggests universal importance of five TAL effectors. Frontiers in Plant Science 2015;6:536.

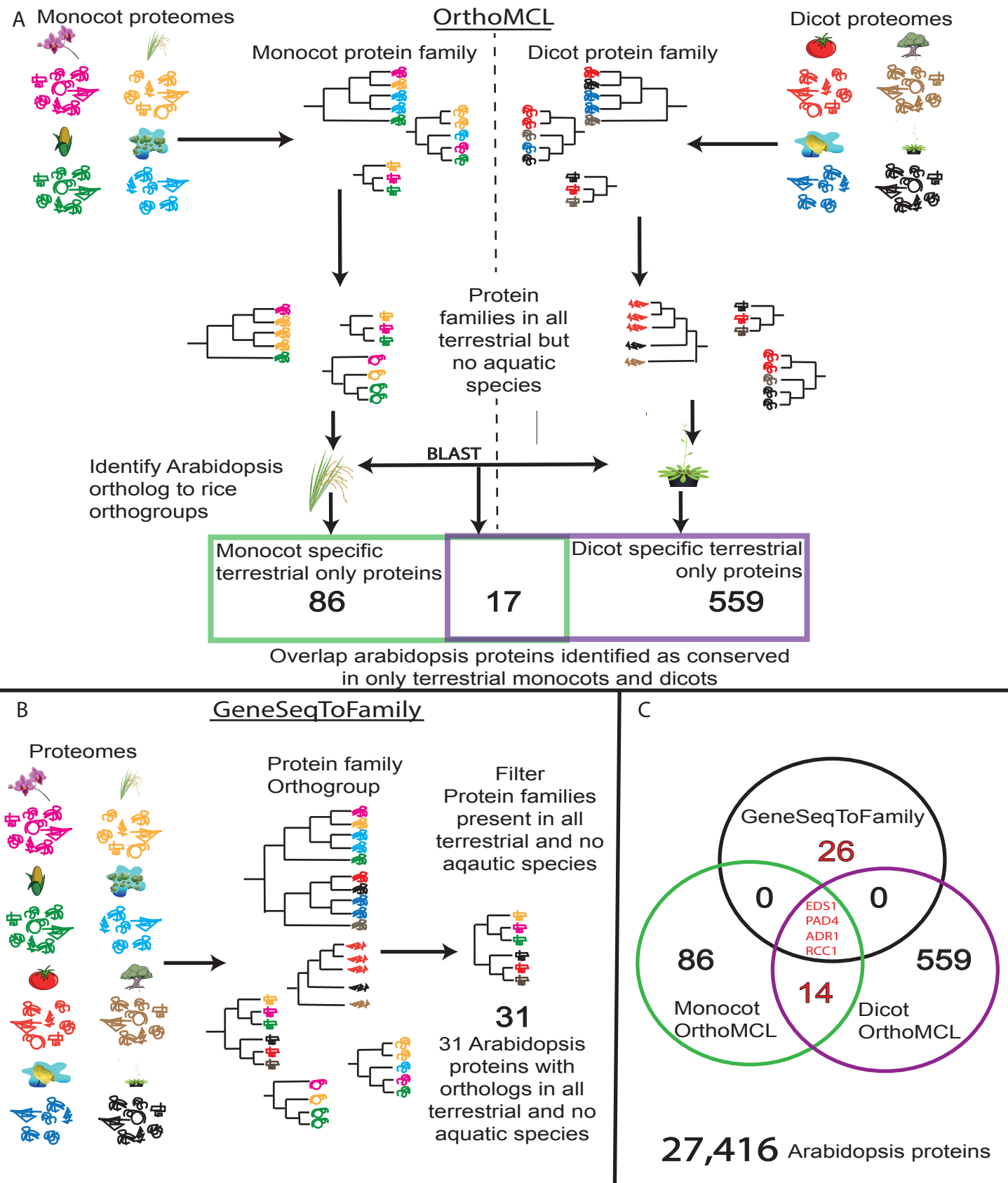
(64) Marcel S, Sawers R, Oakeley E, Angliker H, Paszkowski U. Tissue-adapted invasion strategies of the rice blast fungus Magnaporthe oryzae. Plant Cell 2010 Sep;22(9):3177-3187.
(65) Rawat N, Chiruvuri Naga N, Raman Meenakshi S, Nair S, Bentur JS. A novel mechanism of gall midge resistance in the rice variety Kavya revealed by microarray analysis. Funct Integr Genomics 2012 Jun;12(2):249-264.
(66) Hogenhout SA, Bos JI. Effector proteins that modulate plant–insect interactions. Current Opinion in Plant Biology 2011;14(4):422-428.
(67) Stuart J. Insect effectors and gene-for-gene interactions with host plants. Current Opinion in Insect Science 2015;9:56-61.
(68) Zhang X, Shi Y, Powers JJ, Gowda NB, Zhang C, Ibrahim HM, et al. Transcriptome analyses reveal SR45 to be a neutral splicing regulator and a suppressor of innate immunity in Arabidopsis thaliana. BMC Genomics 2017;18(1):772.
(69) Asselbergh B, De Vleesschauwer D, Höfte M. Global switches and fine-tuning—ABA modulates plant pathogen defense. Mol Plant-Microbe Interact 2008;21(6):709-719.
(70) Cao FY, Yoshioka K, Desveaux D. The roles of ABA in plant–pathogen interactions. J Plant Res 2011;124(4):489-499.
(71) Bartsch M, Gobbato E, Bednarek P, Debey S, Schultze JL, Bautor J, et al. Salicylic Acid–Independent ENHANCED DISEASE SUSCEPTIBILITY1 Signaling in *Arabidopsis* Immunity and Cell Death Is Regulated by the Monooxygenase *FMO1* and the Nudix Hydrolase *NUDT7*. Plant Cell 2006;18(4):1038.
(72) Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, et al. The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. Plant J 2008 Aug;55(3):526-542.
(73) Tehseen M, Cairns N, Sherson S, Cobbett CS. Metallochaperone-like genes in Arabidopsis thaliana. Metallomics 2010;2(8):556-564.
(74) Kim TH, Hauser F, Ha T, Xue S, Bohmer M, Nishimura N, et al. Chemical genetics reveals negative regulation of abscisic acid signaling by a plant immune response pathway. Curr Biol 2011 Jun 7;21(11):990-997.
(75) Jeong CY, Lee SA, Kang C, Cheong Y, Lee H. Overexpression of abiotic stress-induced AtMYBL-O results in negative modulation of abscisic acid signaling through the downregulation of abscisic acid-responsive genes in Arabidopsis thaliana. Plant Growth Regulation 2018;84(1):25-36.
(76) Hachez C, Ohashi-Ito K, Dong J, Bergmann DC. Differentiation of Arabidopsis Guard Cells: Analysis of the Networks Incorporating the Basic Helix-Loop-Helix Transcription Factor, FAMA. Plant Physiol 2011;155(3):1458-1472.
(77) Ma C, Guo J, Kang Y, Doman K, Bryan AC, Tax FE, et al. AtPEPTIDE RECEPTOR2 mediates the AtPEPTIDE1-induced cytosolic Ca2+ rise, which is required for the suppression of Glutamine Dumper gene expression in Arabidopsis roots. J Integr Plant Biol 2014;56(7):684-694.
(78) Jeon H, Kim M, Le P, Ko J, Park E, Lee H, et al. Gain-of-function mutation of AtDICE1, encoding a putative endoplasmic reticulum-localized membrane protein, causes defects in anisotropic cell elongation by disturbing cell wall integrity in Arabidopsis. aob 2018;122(1):151-164.
(79) Brotman Y, Landau U, Cuadros-Inostroza Á, Takayuki T, Fernie AR, Chet I, et al. Trichoderma-Plant Root Colonization: Escaping Early Plant Defense Responses and Activation of the Antioxidant Machinery for Saline Stress Tolerance. PLOS Pathogens 2013;9(3):e1003221.
(80) Dittrich H, Kutchan TM. Molecular cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of benzophenanthridine alkaloids in the response of plants to pathogenic attack. Proc Natl Acad Sci USA 1991;88(22):9969-9973.
(81) Kim T, Kunz H, Bhattacharjee S, Hauser F, Park J, Engineer C, et al. Natural Variation in

Small Molecule–Induced TIR-NB-LRR Signaling Induces Root Growth Arrest via EDS1- and PAD4-Complexed R Protein VICTR in *Arabidopsis*. Plant Cell 2012;24(12):5177.
(82) Zbierzak AM, Porfirova S, Griebel T, Melzer M, Parker JE, Dörmann P. A TIR–NBS protein encoded by Arabidopsis Chilling Sensitive 1 (CHS1) limits chloroplast damage and cell death at low temperature. Plant J 2013;75(4):539-552.
(83) Cui H, Gobbato E, Kracher B, Qiu J, Bautor J, Parker JE. A core function of EDS1 with PAD4 is to protect the salicylic acid defense sector in Arabidopsis immunity. New Phytol 2017 Mar;213(4):1802-1817.
(84) Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 2006 /07/01;34(suppl_2):W373.
(85) Steuernagel B, Jupe F, Witek K, Jones JDG, Wulff BBH. NLR-parser: rapid annotation of plant NLR complements. Bioinformatics 2015 /05/15;31(10):1665-1667.
(86) Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. Bioinformatics 2013 Oct 01,;29(19):2487-2489.
(87) Barson G, Griffiths E. SeqTools: visual tools for manual analysis of sequence alignments. BMC Research Notes 2016;9(1):39.
(88) Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009 /05/01;25(9):1189-1191.
(89) Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014 May 1,;30(9):1312-1313.
(90) Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009 Dec 15;10:421.
(91) VanBuren R, Bryant D, Edger PP, Tang H, Burgess D, Challabathula D, et al. Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum. Nature 2015;527:508.
(92) VanBuren R, Wai CM, Keilwagen J, Pardo J. A chromosome scale assembly of the model desiccation tolerant grass *Oropetium thomaeum* bioRxiv 2018:378943.
(93) Daehwan Kim, Ben Langmead, Steven L Salzberg. HISAT: a fast spliced aligner with low memory requirements. Nature Methods 2015 Apr 1,;12(4):357-360.
(94) Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009 Aug 15,;25(16):2078-2079.
(95) Nicol JW, Helt GA, Blanchard SG, Raja A, Loraine AE. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 2009 /10/15;25(20):2730-2731.
(96) Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 2003 Sep;13(9):2178-2189.
(97) Johnson RN, O'Meally D, Chen Z, Etherington GJ, Ho SYW, Nash WJ, et al. Adaptation and conservation insights from the koala genome. Nature genetics 2018 Aug;50(8):1102-1111.
(98) Thanki AS, Soranzo N, Haerty W, Davey RP. GeneSeqToFamily: a Galaxy workflow to find gene families based on the Ensembl Compara GeneTrees pipeline. Gigascience 2018 03 01,;7(3):1-10.

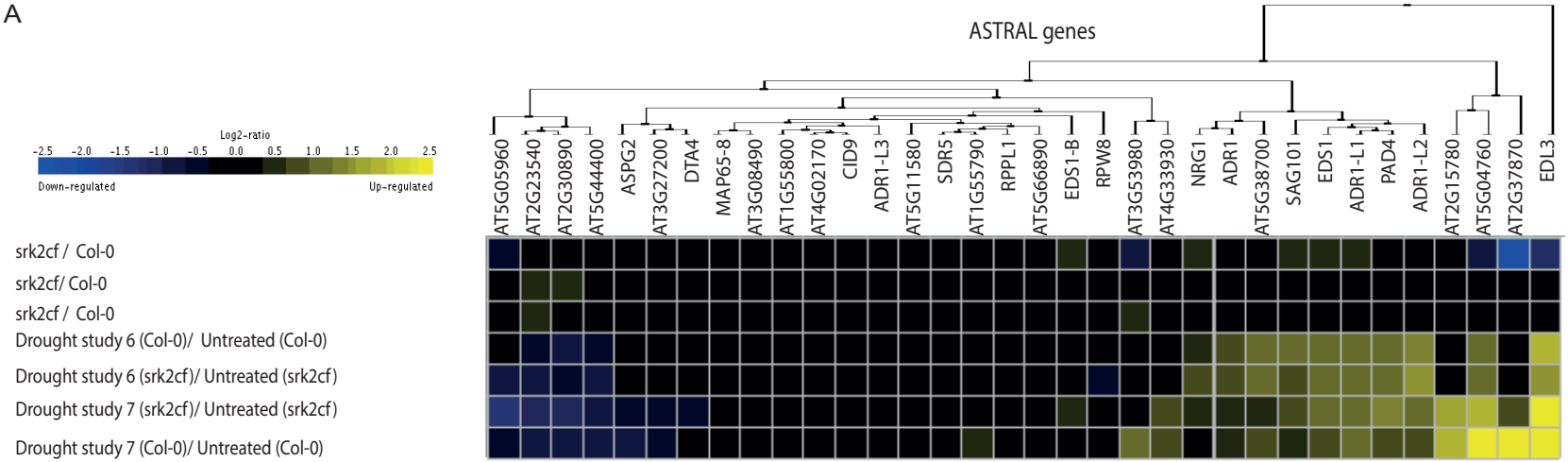

Dicot
Ancestor
Maize
130 MYA
Gymnosperm
Tomato
Monocot
Ancestor
Land transition
703 MYA
Corkscrew
flower
Lotus
Duckweed
Plant
Ancestor


Clade membership
T1
T2
T3
TNL
C1
RPW8
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
CNL
Branch Colour
Eel grass (Zostera marina)
Duckweed (Spirodela polyrhiza)
Maize
Corkscrew plant (Genlisea aurea)
Arabidopsis
Reference genes
Tree scale : 1

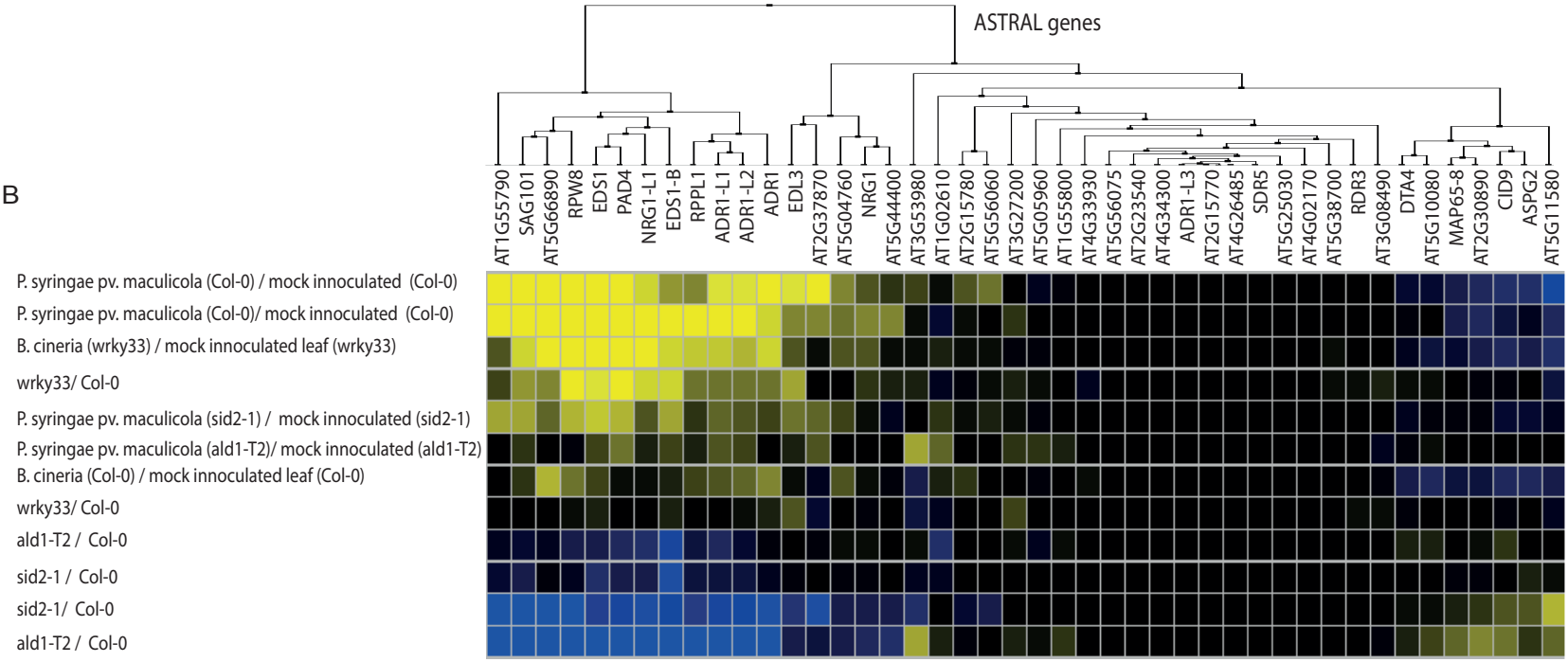

A
Effector target
PBS1
MPK4
RIN4
Sensor NLRs/ RAR1/ SGT1 complex
Effector triggered immunity
Helper NLRs
NRG1
ADR1-L1
NDR1
EDS1/PAD4/SAG101 Heterodimers
FMO1
ICS1
SA
SA signalling
NPR1
MC1
MC4/5
Metacaspase
Cell Death
B
Effector targets
Effector triggered immunity
Metacaspase
SA signalling
PBS1
RIN4
MPK4
EDS1
PAD4
SAG101
ADR1-L1
SGT1
RAR1
NDR1
MC1
MC4/5
NPR1/2
ICS1/2
FMO1
Monocot
Dicot
A. trichopoda
Eel grass
Duckweed
Orchid
Oil palm
Pineapple
Resurrection grass
Rice
Maize
A. coerulea
Lotus
A. thaliana
Amaranthus
Tomato
Ash
Monkey flower
Humped Bladderwort
Corkscrew plant


A
Monocot proteomes
OrthoMCL
Dicot proteomes
Monocot protein family
Dicot protein family
Protein families in all terrestrial but no aquatic species
BLAST
Identify Arabidopsis ortholog to rice orthogroups
Monocot specific terrestrial only proteins
86
17
Dicot specific terrestrial only proteins
559
Overlap arabidopsis proteins identified as conserved in only terrestrial monocots and dicots
B
GeneSeqToFamily
Proteomes
Protein family Orthogroup
Filter Protein families present in all terrestrial and no aqautic species
31
31 Arabidopsis proteins with orthologs in all terrestrial and no aquatic species
C
GeneSeqToFamily
26
0
0
EDS1
PAD4
ADR1
RCC1
86
559
14
Monocot OrthoMCL
Dicot OrthoMCL
27,416 Arabidopsis proteins

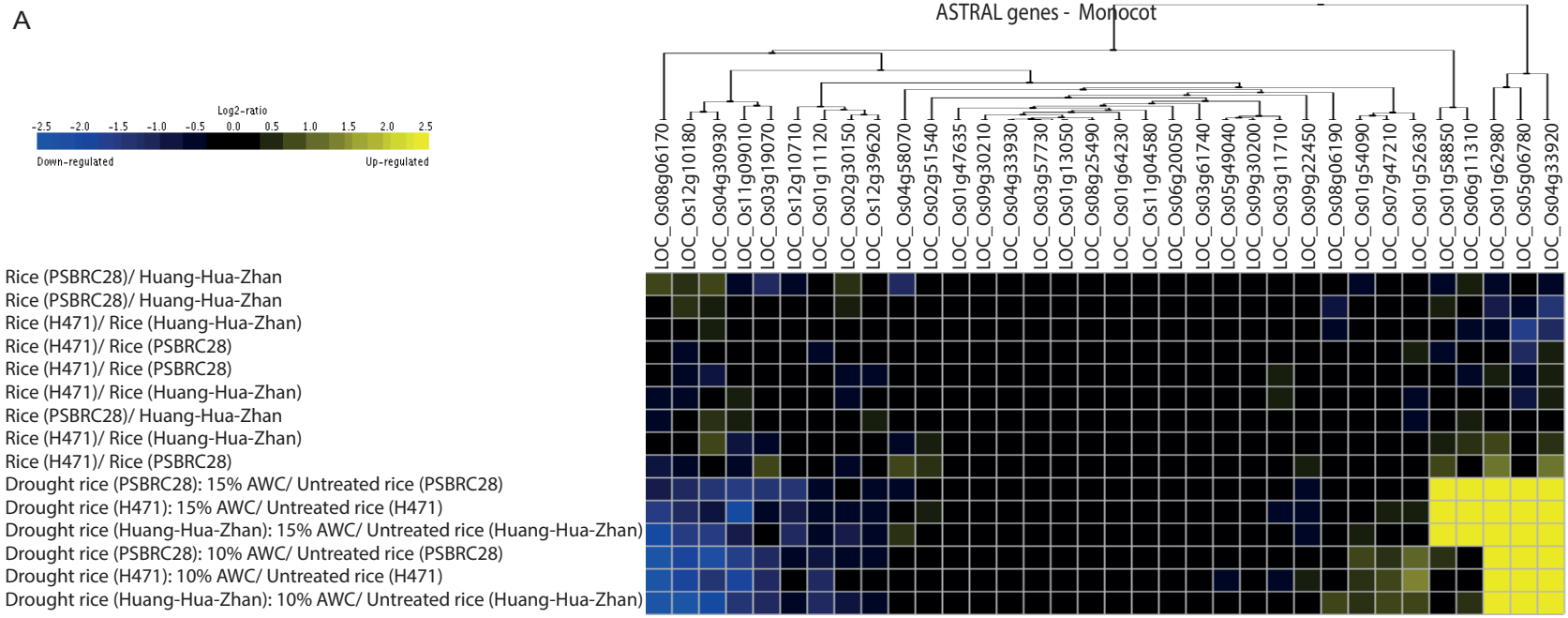
A


ASTRAL genes
Log2-ratio
-2.5 -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 2.5
Down-regulated
Up-regulated
AT5G05960
AT2G23540
AT2G30890
AT5G44400
ASPG2
AT3G27200
DTA4
MAP65-8
AT3G08490
AT1G55800
AT4G02170
CID9
ADR1-L3
AT5G11580
SDR5
AT1G55790
RPPL1
AT5G66890
EDS1-B
RPW8
AT3G53980
AT4G33930
NRG1
ADR1
AT5G38700
SAG101
EDS1
ADR1-L1
PAD4
ADR1-L2
AT2G15780
AT5G04760
AT2G37870
EDL3
srk2cf / Col-0
srk2cf/ Col-0
srk2cf / Col-0
Drought study 6 (Col-0)/ Untreated (Col-0)
Drought study 6 (srk2cf)/ Untreated (srk2cf)
Drought study 7 (srk2cf)/ Untreated (srk2cf)
Drought study 7 (Col-0)/ Untreated (Col-0)

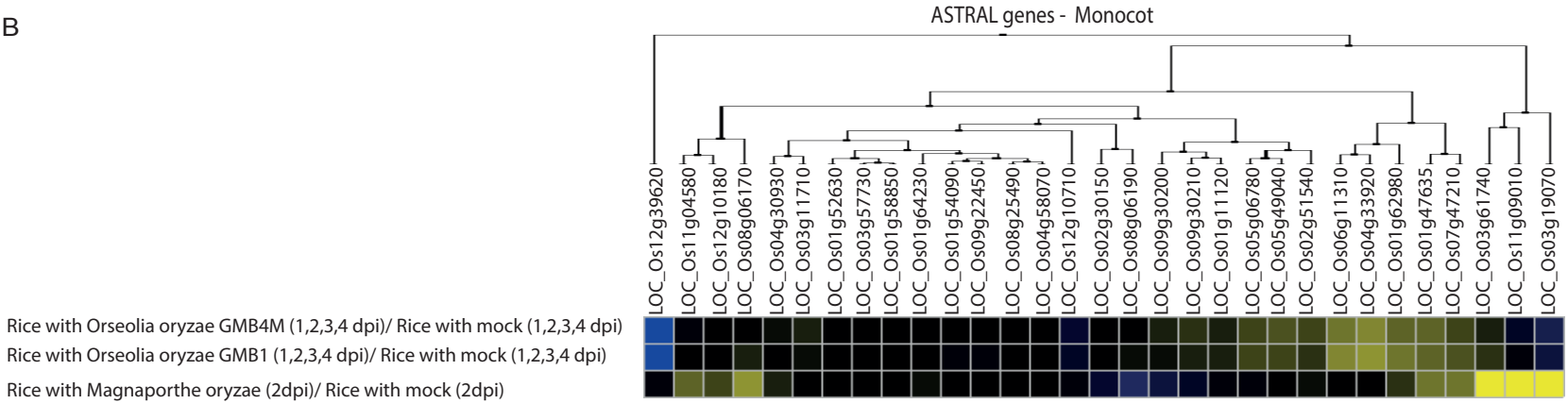
B

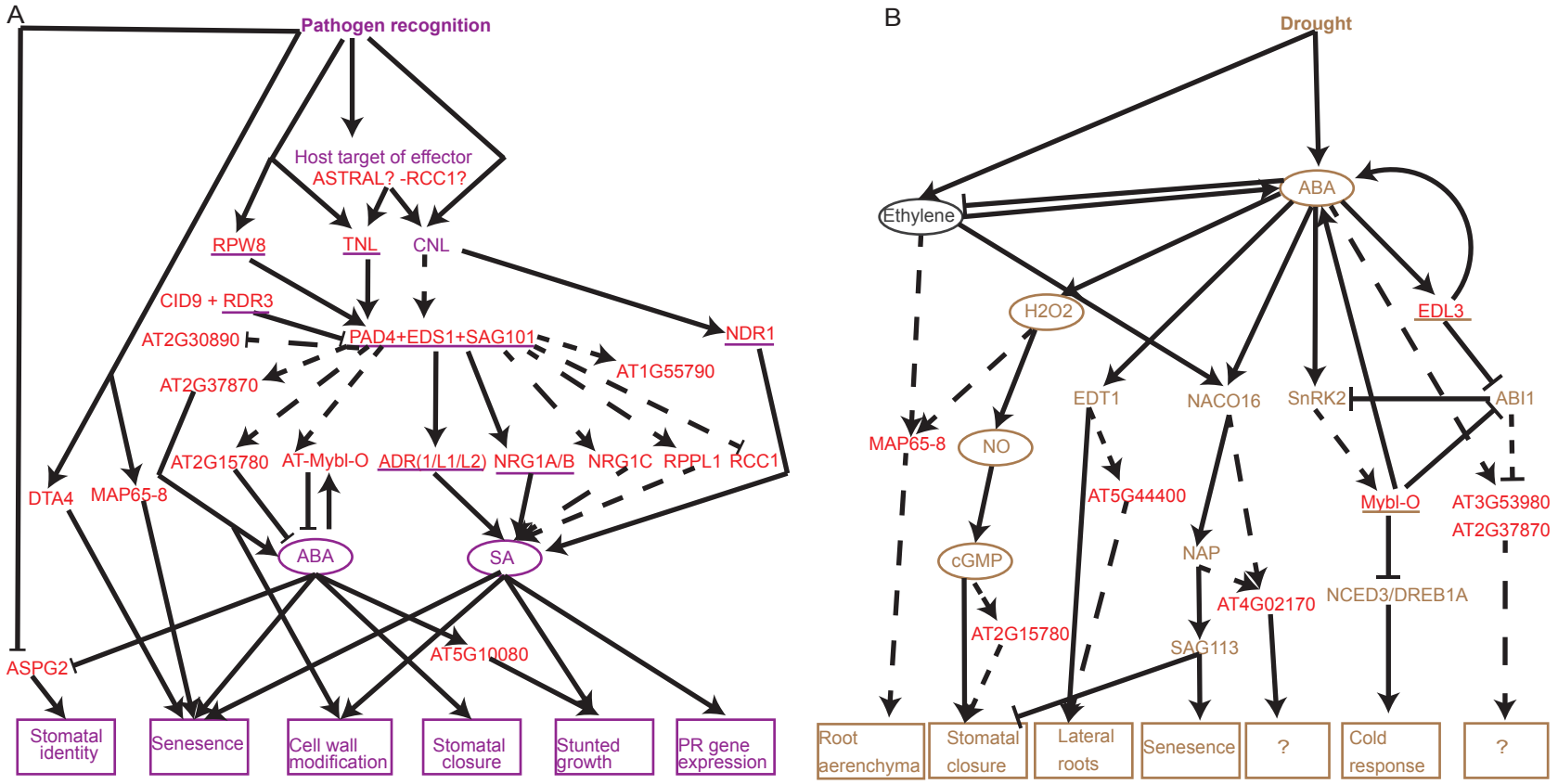

ASTRAL genes
AT1G55790
SAG101
AT5G66890
RPW8
EDS1
PAD4
NRG1-L1
EDS1-B
RPPL1
ADR1-L1
ADR1-L2
ADR1
EDL3
AT2G37870
AT5G04760
NRG1
AT5G44400
AT3G53980
AT1G02610
AT2G15780
AT5G56060
AT3G27200
AT5G05960
AT1G55800
AT4G33930
AT5G56075
AT2G23540
AT4G34300
ADR1-L3
AT2G15770
AT4G26485
SDR5
AT5G25030
AT4G02170
AT5G38700
RDR3
AT3G08490
DTA4
AT5G10080
MAP65-8
AT2G30890
CID9
ASPG2
AT5G11580
P. syringae pv. maculicola (Col-0) / mock innoculated (Col-0)
P. syringae pv. maculicola (Col-0)/ mock innoculated (Col-0)
B. cineria (wrky33) / mock innoculated leaf (wrky33)
wrky33/ Col-0
P. syringae pv. maculicola (sid2-1) / mock innoculated (sid2-1)
P. syringae pv. maculicola (ald1-T2)/ mock innoculated (ald1-T2)
B. cineria (Col-0) / mock innoculated leaf (Col-0)
wrky33/ Col-0
ald1-T2 / Col-0
sid2-1 / Col-0
sid2-1/ Col-0
ald1-T2 / Col-0

A

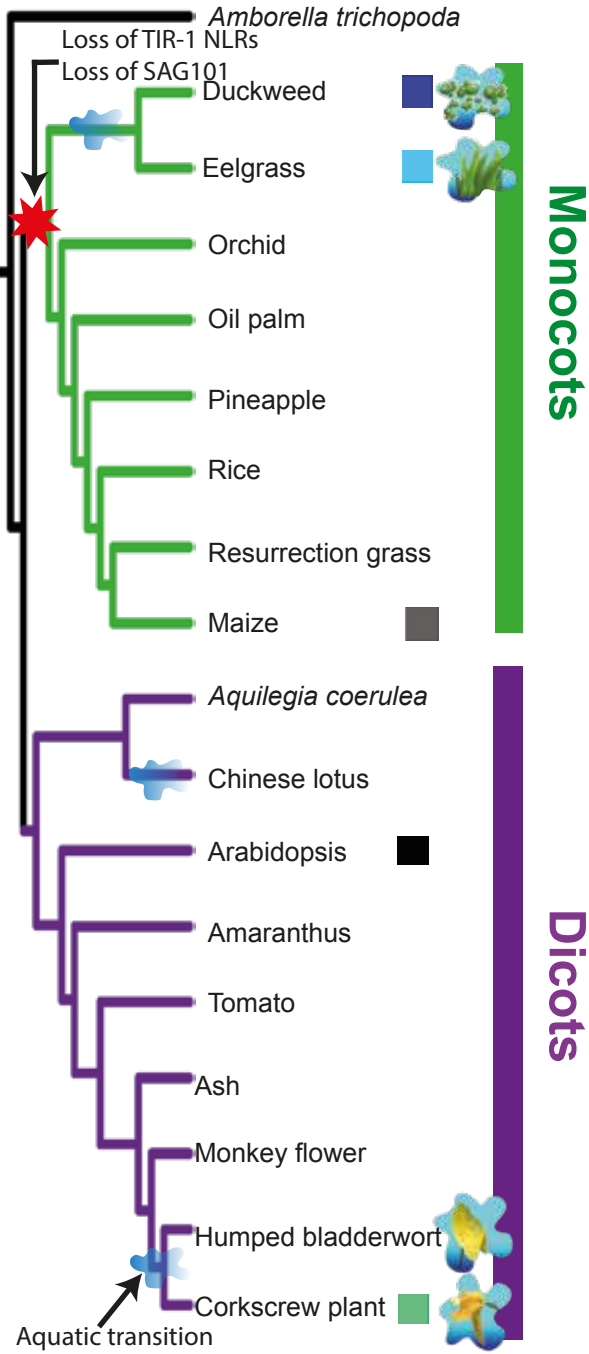

ASTRAL genes - Monocot
Log2-ratio
-2.5 -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 2.5
Down-regulated
Up-regulated
LOC_Os08g06170
LOC_Os12g10180
LOC_Os04g30930
LOC_Os11g09010
LOC_Os03g19070
LOC_Os12g10710
LOC_Os01g11120
LOC_Os02g30150
LOC_Os12g39620
LOC_Os04g58070
LOC_Os02g51540
LOC_Os01g47635
LOC_Os09g30210
LOC_Os04g33930
LOC_Os03g57730
LOC_Os01g13050
LOC_Os08g25490
LOC_Os01g64230
LOC_Os11g04580
LOC_Os06g20050
LOC_Os03g61740
LOC_Os05g49040
LOC_Os09g30200
LOC_Os03g11710
LOC_Os09g22450
LOC_Os08g06190
LOC_Os01g54090
LOC_Os07g47210
LOC_Os01g52630
LOC_Os01g58850
LOC_Os06g11310
LOC_Os01g62980
LOC_Os05g06780
LOC_Os04g33920
Rice (PSBRC28)/ Huang-Hua-Zhan
Rice (PSBRC28)/ Huang-Hua-Zhan
Rice (H471)/ Rice (Huang-Hua-Zhan)
Rice (H471)/ Rice (PSBRC28)
Rice (H471)/ Rice (PSBRC28)
Rice (H471)/ Rice (Huang-Hua-Zhan)
Rice (PSBRC28)/ Huang-Hua-Zhan
Rice (H471)/ Rice (Huang-Hua-Zhan)
Rice (H471)/ Rice (PSBRC28)
Drought rice (PSBRC28): 15% AWC/ Untreated rice (PSBRC28)
Drought rice (H471): 15% AWC/ Untreated rice (H471)
Drought rice (Huang-Hua-Zhan): 15% AWC/ Untreated rice (Huang-Hua-Zhan)
Drought rice (PSBRC28): 10% AWC/ Untreated rice (PSBRC28)
Drought rice (H471): 10% AWC/ Untreated rice (H471)
Drought rice (Huang-Hua-Zhan): 10% AWC/ Untreated rice (Huang-Hua-Zhan)

B


ASTRAL genes - Monocot
LOC_Os12g39620
LOC_Os11g04580
LOC_Os12g10180
LOC_Os08g06170
LOC_Os04g30930
LOC_Os03g11710
LOC_Os01g52630
LOC_Os03g57730
LOC_Os01g58850
LOC_Os01g64230
LOC_Os01g54090
LOC_Os09g22450
LOC_Os08g25490
LOC_Os04g58070
LOC_Os12g10710
LOC_Os02g30150
LOC_Os08g06190
LOC_Os09g30200
LOC_Os09g30210
LOC_Os01g11120
LOC_Os05g06780
LOC_Os05g49040
LOC_Os02g51540
LOC_Os06g11310
LOC_Os04g33920
LOC_Os01g62980
LOC_Os01g47635
LOC_Os07g47210
LOC_Os03g61740
LOC_Os11g09010
LOC_Os03g19070
Rice with Orseolia oryzae GMB4M (1,2,3,4 dpi)/ Rice with mock (1,2,3,4 dpi)
Rice with Orseolia oryzae GMB1 (1,2,3,4 dpi)/ Rice with mock (1,2,3,4 dpi)
Rice with Magnaporthe oryzae (2dpi)/ Rice with mock (2dpi)


A
Pathogen recognition
Host target of effector
ASTRAL? -RCC1?
RPW8
TNL
CNL
CID9 + RDR3
AT2G30890
PAD4+EDS1+SAG101
NDR1
AT2G37870
AT1G55790
AT2G15780
AT-Mybl-O
ADR(1/L1/L2)
NRG1A/B
NRG1C
RPPL1
RCC1
DTA4
MAP65-8
ABA
SA
ASPG2
AT5G10080
Stomatal identity
Senesence
Cell wall modification
Stomatal closure
Stunted growth
PR gene expression
B
Drought
ABA
Ethylene
H2O2
EDL3
EDT1
NACO16
SnRK2
ABI1
MAP65-8
NO
AT5G44400
Mybl-O
AT3G53980
AT2G37870
cGMP
NAP
NCED3/DREB1A
AT4G02170
AT2G15780
SAG113
Root aerenchyma
Stomatal closure
Lateral roots
Senesence
?
Cold response
?

A



B

	A. trichopoda
No. High Confiidence NLRs	54
Total No. NLRs	54

	Eel grass	Duckweed	Orchid	Oil palm	Pineapple
No. High Confiidence NLRs	38	61	44	226	130
Total No. NLRs	45	85	54	269	189

	Resurrection grass	Rice	Maize
No. High Confiidence NLRs	7	381	115
Total No. NLRs	68	516	155

	A. coerulea	Lotus	*A. thaliana*	Amaranthus	Tomato
No. High Confiidence NLRs	191	192	156	80	166
Total No. NLRs	243	237	167	130	263

	Ash	Monkey flower	Humped Bladderwort	Corkscrew plant
No. High Confiidence NLRs	155	280	0	8
Total No. NLRs	237	326	1	17